How are genetic factors influencing cancer susceptibility and mortality patterns in the St. Lawrence beluga?

By: Suchinta Arif

A Thesis Submitted to
Saint Mary's University, Halifax, Nova Scotia
in Partial Fulfillment of the Requirements for
the Degree of Master of Science in Applied Science.

August, 2016, Halifax, Nova Scotia

Approved: Dr. Timothy Frasier
Supervisor

Approved: Dr. Linda Campbell
Supervisory Committee

Approved: Dr. David Dansereau
Supervisory Committee

Approved: Dr. Don Stewart
External Committee

Date: August 2016

Abstract

How are genetic factors influencing cancer susceptibility and mortality patterns in the St. Lawrence beluga?

by Suchinta Arif

The purpose of this project was to understand how genetic factors may be influencing cancer susceptibility and mortality patterns in the St. Lawrence beluga. To assess if variation along candidate genes are influencing cancer susceptibility, I designed primers for the amplification of the p53 gene, and microsatellite regions along seven additional genes: p16, Rb1, BRCA1, BRCA2, PTEN, HPSE2, and FOXO3. Comparing cancer individuals to a control group showed no strong differences with respect to variation along the p53 gene (sequencing analysis) or other candidate loci (microsatellite association analysis), though some patterns warrant further investigation. The degree of inbreeding also showed no relationship with cancer susceptibility. However, bacteria and parasite-induced mortalities were associated with relatively high and low levels of inbreeding, respectively, highlighting the different ways in which inbreeding may be influencing mortality patterns in this population.

Date: August 2016

Acknowledgements

I would first like to thank my supervisor Tim Frasier for allowing me the opportunity to work on this project. I am grateful for all the guidance he has provided me through the last two years, which included: broadening my knowledge of conservation genetics, thinking more deeply about how to conduct scientific research, and introducing me to Bayesian stats. Most importantly, I would like to thank him for allowing me to work autonomously and in my own way, even when it clashed with more conventional ways of doing things. I would also like to thank my committee members David Dansereau and Linda Campbell for providing me valuable advice throughout my thesis, and a special thanks to Anne Dalziel for being an awesome Prof in general.

Thanks to GREMM and DFO for mortality data on the St. Lawrence beluga population, as well as for collecting skin samples used for this study.

I would like to acknowledge all my wonderful Halifax friends, especially Meredith Perich, Natasha Daze-Query and Christine Konrad for helping me both in and out of the lab. Thanks for keeping me relatively sane!

Table of Contents

List of Figures

List of Tables

Chapter 1:

General Introduction

Cetaceans (whales, dolphins, and porpoises) include a diverse range of marine mammals that have been inhabiting the oceans for over 50 million years (Gingerich et al. 1983). Perhaps the most destructive event to take place throughout the history of cetaceans began with the advent of commercial whaling, which subsequently resulted in the depletion of many populations worldwide (Rocha et al., 2015). Populations of both baleen (mysticete) and toothed (odontocete) whales were systematically hunted to near extinction during the past two hundred years, with recent estimates suggesting that approximately 3 million individuals were slaughtered during the 1900s alone (Rocha et al., 2015). In 1986, the International Whaling Commission (IWC) banned commercial whaling, hoping to lead to a steady recovery of affected populations. However, despite receiving decades of protection, some populations, such as the North Pacific right whale, western Pacific gray whale, and the southern hemisphere blue whale are showing no signs of recovery (Clapham et al. 1999).

One example of a population that has still not recovered from the past effects of whaling is the St. Lawrence Beluga (SLB) (*Delphinapterus leucas*) (DFO, 2012). The SLB is one of eight populations of beluga occurring in Canadian waters, and one of 29 populations worldwide (DFO, 2012). It occurs in the southern-most limit of the species' range, occupying over 8000 km^2 in the St. Lawrence Estuary, the Gulf of St. Lawrence, and the Saguenay River (DFO, 2012). Due to their genetic (Brennin et al., 1997; Brown Gladden et al., 1997; Murray et al., 1999; de March and Postma, 2003) and physical isolation from other populations, it is considered a distinct 'Designatable Unit' by the Committee on the

Status of Endangered Wildlife in Canada (COSEWIC) (COSEWIC, 2004), and therefore requires specific conservation and management consideration.

Commercial whaling of the SLB began in the 1600s and continued until the 1950s (DFO, 2012). Reeves and Mitchell (1984) postulate that up to 15000 individuals were killed during this time. Belugas were also hunted in an attempt to protect fisheries, as many fishermen saw them as competitors, and in the 1920s the government of Quebec offered a $15 bounty for each beluga killed and allowed for the use of bombs to drive belugas out of fishing areas (Scharrer, 1983). Hunting also occurred for recreational purposes and sport hunting continued into the 1970s (DFO, 2012). In 1979, after a drastic and noticeable decline in population size, hunting was finally banned under the Canadian Fisheries Act (Lesage and Kingsley, 1998). Unfortunately, despite receiving protection for the last 35 years, the population has yet to show signs of recovery (DFO, 2012). Whereas historical estimates of population size range from 7,800 to 10,100 individuals (DFO, 2005; Hammill et al., 2007), estimates for the past ~20 years hovered around 1,000 individuals (Hammill et al., 2007). More recently, it appears that the population size is decreasing, with a recent estimate of 889 individuals in 2012 (Mosnier et al., 2015).

To add to the problem, the SLB have had one of the highest known cancer rates noted in any wildlife population (Martineau et al., 2002). In the past, cancer has been suggested as the second leading cause of death, accounting for 18% of mortalities (Martineau et al., 2002, but see Hammill et al. 2003). The cancer rate in the SLB population is presumed to

be much higher than that of other Arctic populations. For example, no tumors were found in 50 beluga whales examined in the Canadian Arctic (De Guise et al., 1994). Although these data may not be representative because these were randomly selected live animals (of unknown age) and not stranded animals, it does suggest that the cancer rate for other beluga populations are comparatively low. Because the St. Lawrence belugas live downstream of a highly industrialized Great Lakes Region, many scientists presume that their high cancer rate is caused by the contaminants found in their environment (De Guise et al., 1994; Martineau et al., 2002; Newman and Smith, 2006). Indeed, past toxicological studies have shown that the tissues of St. Lawrence belugas have high levels of organochlorines (OCs) and polycyclic aromatic hydrocarbons (PAHs) (Martineau et al. 1988). The correlation of contaminants with the presence of neoplasm and other lesions have led some scientists to suggest that the contaminants are either directly acting as carcinogens, or indirectly leading to cancer by weakening the immune systems of belugas (DFO, 2012). However, although all belugas in this region are subjected to similar amounts of contaminants in their environment, not all individuals are dying of cancer. This differential response, despite similar exposure, suggests that intrinsic factors, such as genetic characteristics, may play an important role in cancer susceptibility.

The factors limiting recovery in this population are currently unknown. Specific hypothesis testing, rather than relying on presumed cause and effect scenarios, is needed to help ensure that the correct limiting factors are identified, and thus increase the likelihood of successful conservation initiatives. Past research on the SLB has ascribed their failure to recover to anthropogenic activities (Kingsley, 2002), including polluting

the belugas' environment (Beland et al., 1993; Lebeuf et al., 2014). Other factors potentially limiting recovery include additional extrinsic factors, such as habitat disturbance and degradation, as well as intrinsic factors such as a low level of genetic variability and inbreeding. However, no clear link has been made between the lack of recovery and any of these hypothesized factors.

From a conservation genetics perspective, a small population size may make the SLB genetically vulnerable due to reduced genetic variability, which can subsequently reduce a population's overall fitness, as well as lower a population's ability to adapt to novel changes (Amos and Balmford, 2001). Compared to other Canadian beluga populations, past studies have shown that the SLB have reduced genetic diversity, suggesting that genetic drift, inbreeding, or both can influence the genetic characteristics of this population (Patenaude et al., 1994; Murray et al., 1999; de March and Postma, 2003). Reduced genetic variability is detrimental as it can impact survival, reproductive success, and a population's ability to persist over time (e.g. Keller et al., 1994; Hedrick and Kalenowski, 2000; Moyle et al., 2003). Although low genetic variability is listed as a key threat to recovery (COSEWIC, 2004), it is currently unknown if, and to what extent, genetic factors are influencing mortality rates and disease susceptibility in this population.

The purpose of this project was to determine if genetic characteristics are influencing cancer susceptibility and mortality patterns in the SLB. As such, my thesis will focus on answering the following questions:

1. Does variation (with respect to polymorphisms and cancer-linked microsatellite alleles) in key cancer-related genes influence cancer susceptibility in the SLB population?
2. Does the degree of inbreeding influence mortality patterns in the SLB population?

Cancer in Wildlife:

Cancer has only recently been recognized as a conservation threat for several wildlife populations. Although difficult to detect, increased research and monitoring efforts have identified a high prevalence of benign and malignant tumors across a wide range of taxa (McAloose and Newton 2009). A few prominent examples include: the Tasmanian devil (*Sarcophilus harrisii*) facial tumor disease (McCallum et al., 2007; Pyecroft et al., 2007), California sea lion (*Zalophus californianus*) genital carcinoma (Gulland et al., 1996; Browning et al., 2014), as well as carcinoma (mostly gastrointestinal adenocarcinoma) found in the St. Lawrence Beluga population (Martineau et al., 2002). Other cetacean species with noted tumors include the sperm whale (*Physeter macrocephalus*) (Lambertsen et al. 1987), Burmeister's porpoise (*Phocoena spinipinnis*) (Van Bressem et al. 2007), dusky dolphin (*Lagenorhynchus obscurus*) (Van Bressem et al. 2000), bottlenose dolphin (*Tursiops truncatus*) (Bossartt et al. 2005), Atlantic white-sided dolphin (*Lagenorhynchus acutus*) (Geraci et al. 1987), harbor porpoise (*Phocoena phocoena*) (Van Bressem et al. 1999), killer whale (*Orcinus orca*) (Bossart et al. 1996), and narwhal (*Monodon monoceros*) (Geraci et al. 1987).

Unfortunately, several factors make it challenging to study cancer in wildlife. First, the identification of cancer in wild animals is difficult as it relies on access to carcasses that are fresh enough to adequately examine (McAllose and Newton, 2009). Access to appropriate samples can be hindered by environmental obstacles, such as the wide dispersal of animals across oceans, tissue loss through environmental decomposition, and predation and/or scavenging. Another issue is that advanced cancer diagnostics, such as computed tomography and magnetic resonance imaging, are often not available (or beyond the budget) for wildlife research. Therefore, cancer in wildlife can go largely undetected. However, even with these limitations several cases of populations that are plagued with a high incidence of cancer have emerged (McAllose and Newton, 2009), and efforts continue to investigate the etiology and spread of this disease.

Understanding the factors influencing cancer development in a given population is necessary to help mitigate the problem. In the past, research centering around the etiology of cancer has focused on environmental factors. This makes sense given that many of the wildlife populations suffering from a high incidence of cancer live in environments that are heavily contaminated with pollutants. For example, fresh water, marine, and estuarine fishes inhabiting industrialized areas with high levels of polycyclic aromatic hydrocarbons (PAHs) have been associated with the presence of epizootic tumors (Malins et al., 1985; Smith et al., 1989; Baumann et al., 1996). In addition to observational studies, experimental studies have demonstrated that exposure to PAH can lead to tumour development in fishes (Hendricks et al., 1985; Black et al., 1984), and the

subsequent decrease of PAH in their environment can lead to the decline of cancer rate as well (Baumann and Harshbarger, 1995).

High levels of PAHs are also hypothesized to cause cancer development in the St. Lawrence beluga. In particular, benzo[a]pyrene (a type of PAH) ingestion is thought to result in the high incidence of intestinal neoplasia observed in this population (McAllose and Newton, 2009). Benzo[a]pyrene (BaP) metabolites are mutagenic and highly carcinogenic, with various studies documenting links between BaP and cancer (Kielbohmer, 2001). Carcinogenesis with respect to BaP exposure depends on its enzymatic metabolism to BaP diol epoxide, which intercalates in DNA (specifically by covalently bonding to the guanine bases) in order to distort it (Volk et al., 2003). This in turn disrupts DNA replication and induces mutations leading to cancer development (Volk et al., 2003). Past research suggests that BaP diol epoxide may specifically target the p53 gene by inducing G (guanine) to T (thymidine) transversions within the p53 gene, which in turn can inactivate this tumor suppressor (Pfeifer et al., 2002).

It is possible that the high levels of BaP in the St. Lawrence estuary may be predominantly responsible for cancer development in the SLB population. Concentration of BaP in sediment and invertebrates in the St. Lawrence estuary are relatively high compared to other areas. For example, the BaP concentrations of blue mussels were 200 times higher after they were transplanted into the Saguenay River (Cross et al., 1983). Because belugas dig into sediments and feed on significant amounts of bottom invertebrates (Dalcourt et al., 1992), BaP can bioaccumulate (McAllose and Newton,

2009), which may help explain why BaP DNA adducts have been detected in stranded SLB tissue, but have not been detected in Arctic belugas (which live in presumably less contaminated environments) (Shugart et al., 1990).

Because the etiology of cancer is often multifactorial, and can involve a complex interplay between environmental and genetic factors, it is important to understand the genetic basis of cancer development as well. A few studies have highlighted the role of genetics in cancer susceptibility in wildlife. One example comes from the California sea lion population, which has a high occurrence of urogenital cancer, with studies revealing metastatic carcinoma in 26% of individuals admitted to a rehabilitation centre between 1998 and 2012 (Browning et al., 2014). Cancer susceptibility in this population has recently been linked to genetic differences occurring along the heparanase 2 (HPSE2) gene (Browning et al., 2014), as well as increased levels of inbreeding (Acevedo-Whitehouse et al., 2003). Genetic factors also influence cancer development in the Tasmanian devil, in which low diversity at the MHC class I genes have influenced the spread of their facial tumor disease (Siddle et al., 2007). As a third example, nephroblastomas in Japanese eels (*Anguilla japonica*) result from a mutation in the Wilm's tumour suppressor (WTI) gene (Harshbarger and Slatick, 2001).

Determining the influence of genetic factors on cancer susceptibility is needed to assess if, and to what degree, genetic factors are influencing health and limiting recovery in populations threatened with a high incidence of cancer. For example, possessing point mutations can make some individuals more prone than others to developing neoplasms

when exposed to environmental carcinogens. Such an effect may help explain the intrapopulation variation in cancer susceptibility, despite individuals living in the same environment, observed in the SLB. Disentangling the role of genetic factors in cancer development can also help guide conservation decisions for specific populations, providing information on the relative impacts of genetic and environmental factors on individual health, and thus indicating where conservation actions can most effectively be directed.

Conservation Genetics:

Conservation genetics is an interdisciplinary science that draws from ecology, population genetics, molecular biology, and other fields to understand how genetic characteristics are influencing the persistence of a population or species. Conservation genetics became prominent in the early 1980s (Schonewald-Cox et al., 1983) and while there was initial debate concerning the relative importance of genetics vs. demography in extinction probabilities (e.g., Lande 1988; Avise 1989), it is becoming increasingly apparent that genetic factors can play an important role in determining population persistence over time (e.g., Vucetich and Waite, 1998; Westemeier et al. 1998; Spielman et al. 2004).

Conservation genetics studies tend to focus on small and/or declining populations, as they are more susceptible to genetic problems. One major concern for populations is the loss of genetic variability (O'Brien 1994). When abundance has drastically declined, a large proportion of genetic variability can become lost due to the chance loss of alleles.

Reduced genetic variability at the population level can result in offspring having lower heterozygosity, which in turn can reduce their overall fitness. Genome-wide heterozygosity is advantageous because, in many cases, a heterozygote genotype has a higher fitness than homozygote genotypes. This phenomenon has been observed across a variety of traits for many different species, and past studies linking heterozygosity at a set of molecular markers to variation in fitness-associated traits have shown significant multilocus heterozygosity-fitness correlations (e.g. Rikjs et al. 2008; Forstmeier et al., 2012). Although reduced heterozygosity directly impacts the fitness of individuals, the accumulation of these individual impacts can result in reducing the overall fitness of the population, and subsequently impact a species' viability. Reduced variability is also detrimental because it can lower a population's ability to react to novel changes in the environment. For example, when four populations of *Drosophila melanogaster*, varying in their level of genetic diversity, were exposed to a novel environment of high salinity, the populations with higher diversity were better able to adapt over time (Frankham et al., 1999). An inability to adapt to changes can be harmful because many species are now being faced with added anthropogenic stresses such as habitat alteration and/or increased contaminant loads in their environment, and populations with low variability may be less able to cope with these changes.

One of the ways genetic diversity can be further reduced in a small population is through increased inbreeding, which decreases heterozygosity within inbred individuals. When a population is comprised of few individuals, the chance of matings between related individuals becomes more likely than in a larger population. One of the ways increased

inbreeding (and hence, reduced heterozygosity) can reduce fitness is through the expression of deleterious recessive alleles (Crow, 1948). Because many genes in a diploid organism can operate with one single functional copy, loss of function mutations are generally recessive, and only show their effect in the homozygous state (Crow, 1948). Mating between close relatives increases the likelihood of homozygosity, including for deleterious alleles, which in turn can lower fitness, having further negative consequences for population size. Inbreeding can also lead to an overall decline in fitness, as many traits show a pattern of heterozygote advantage. For example, previous studies have linked reduced genome-wide heterozygosity to decreased parasite resistance (Coltman et al., 1999), increased cancer susceptibility (Acevedo-Whitehouse et al., 2003), juvenile mortality (Rijks et al., 2008) and reduced reproductive performance (Wildt et al., 1982) in several wildlife populations.

The recognition that inbreeding can have negative effects on fitness has made inbreeding depression a concern for small populations. There is growing evidence that inbreeding depression can impact not just individuals, but the population as a whole (e.g., Amos and Balmford, 2001). One of the negative impacts of increased inbreeding is that it may contribute to an extinction vortex (Lacy and Lindenmayer, 1995). This occurs when inbreeding depression leads to a further reduction in population size, which in turn can increase both inbreeding depression and vulnerability to stochastic events. In some cases, inbreeding depression can even lead to extinction. For example, in Glanville fritillary butterflies (*Melitaea cinxia*), relatively inbred populations have a higher extinction probability than outbred populations (Saccheri et al., 1998). As a result, inbreeding

depression is often considered one of the most important genetic threats facing small populations (Amos and Balmford, 2001).

Conservation Genetics of the St. Lawrence Beluga:

Despite receiving protection over the last 35 years, the SLB is showing no signs of recovery and has a recent population estimate of only 889 individuals (Mosnier et al., 2015). Although the factors limiting recovery are not explicitly known, genetic factors may be influencing mortality patterns in this population. For example, cancer susceptibility may, in part, be driven by genetic characteristics. Variation in cancer-related genes may be influencing cancer susceptibility, which can help explain why some individuals are dying from cancer while others are not, despite the entire population being subjected to contaminants. Increased inbreeding may also help explain some of the observed pathologies seen in this population. Along with a high rate of cancer, infectious diseases caused by parasites or bacteria are also common causes of death in the population, accounting for 20% and 18% of mortalities, respectively (DFO, 2012). Inbred offspring may be more prone to developing these diseases than offspring from more distantly related individuals. Further, inbreeding depression may also explain the high rate of juvenile deaths seen in recent years (Monsnier et al., 2015). To address these issues, this thesis will focus on addressing the following questions: 1) Does variation in key cancer-related genes influence cancer susceptibility in the SLB population? and 2) Does degree of inbreeding influence cancer susceptibility and mortality patterns in the SLB population?

To determine whether polymorphisms in key cancer-related genes are influencing cancer susceptibility, I compared the DNA of individuals that died of cancer with those that died of other causes. I first focused on directly sequencing the p53 gene, as this is arguably one of the most important genes involved in cancer in humans (Kang, 2009). The p53 protein plays a critical role in DNA repair and cell-cycle control and regulates the transcription of many genes in response to a variety of stress signals. Following DNA damage, it is responsible for regulating key processes that suppress the development of cancer, which include: DNA repair, cell-cycle arrest, senescence, and apoptosis. While there are numerous cancer-related genes, the p53 pathway seems to be the most important with respect to tumor prevention (Kang, 2009). In some cases, disruption of normal p53 function is a prerequisite for the development or progression of tumors (Kang, 2009). It is also the most frequently mutated gene in the presence of human tumors, with over 50% of tumors containing mutations within this region (Bennett, 1999).

In addition to sequencing the p53 gene, microsatellite loci were identified within, or in close proximity to, several other genes that influence cancer susceptibility. Specifically, these genes were: p16, Rb1, BRCA1, BRCA2, PTEN, HPSE2, and FOXO3. Individuals were then genotyped at these loci to test if genetic characteristics at specific genes are associated with cancer risk. This approach is based on the assumption that microsatellite alleles associated with the presence of cancer are linked to a nearby genomic region in which genetic variation influences cancer development. Microsatellite markers are ideal for linkage studies as they are highly polymorphic, and relatively inexpensive and easy to

analyze (Schlötterer et al. 1991, Dawson et al. 2000). Because sequencing can be time consuming and expensive, microsatellite linkage analysis provides a more economical alternative when screening many candidate loci.

To assess the degree to which inbreeding is influencing cancer susceptibility and mortality patterns, I first had to estimate how inbred each individual is with respect to the population as a whole. This was done through analyses of neutral microsatellite loci, which were used as a proxy for measuring genome-wide heterozygosity. Individuals that have higher homozygosity are assumed to be more inbred than those that are more heterozygous. By comparing between groups that died of different causes (e.g. cancer, parasites, etc.), I could test the hypotheses that inbreeding is influencing mortality patterns.

Formally testing hypotheses regarding the influence of genetic characteristics on cancer susceptibility and mortality patterns can help identify what role, if any, genetic factors have in limiting population recovery in the St. Lawrence Beluga. This information is necessary for identifying limiting factors, understanding population trends, identifying what recovery rates are possible, and for identifying where conservation actions can best be focused.

Literature Cited:

Acevedo-Whitehouse, K. F., D. Gulland, D. Greig, and W. Amos. 2003. Inbreeding: Disease susceptibility in California sea lions. Nature 422: 35.

Amos, W. and A. Balmford. 2001. When does conservation genetics matter? Heredity 87: 257-265.

Avise, J. C. 1989. A role for molecular genetics in the recognition and conservation of endangered species. TREE 4: 279-281.

Baumann, P. C., I.R. Smith, and C.D. Metcalfe. 1996. Linkages between chemical contaminants and tumors in benthic Great lakes fish. J. Great Lakes Resources 22:131–152.

Baumann, P. C. and J.C. Harshbarger. 1995. Decline in liver neoplasms in wild brown bullhead catfish after coking plant closes and environmental PAHs plummet. Environmental Health Perspective 103:168–170.

Beland, P., S. Guise, C. Girard, A. Lagace, D. Martineau, R. Michaud, *et al*. 1993. Toxic compounds and health and reproduction effects in St. Lawrence beluga whales. Journal of Great Lakes Research 19: 766-775.

Belchis, D., C. Meece, F. Benko, P. Rogan, R. Williams and C. Gocke. 1996. Loss of heterozygosity and microsatellite instability at the retinoblastoma locus in osteosarcomas. Diagnostic. Molecular Pathology 5(3): 213-219.

Black, J. J. et al. in Water Chlorination Chemistry, Environmental Impact and Health Effects (eds Jolley, R. L et al.) 415–427 (Plenum, New York, 1984).

Brennin, R., B. W. Murray, M. K. Friesen, L. D. Maiers, J. W. Clayton and B. N. White. 1997. Population genetic structure of beluga whales (Delphinapterus leucas): Mitochondrial DNA sequence variation within and among North American populations. Canadian Journal of Zoology 75(5): 795-802.

Brown Gladden, J. G. and J. W. Clayton. 1993. Variable nucleotide sequences of control region mitochondrial DNA and population structure in North American beluga whales. Tenth biannual conference on the biology of marine mammals. November 11-15, 1993. Galveston, Texas.

Brown Gladden, J. G., M. M. Ferguson and J. W. Clayton. 1997. Matriarchal genetic population structure of North America beluga whales (*Delphinapterus leucas*) (Cetacea: Monodontidae). Molecular Ecology 6: 1033-1046.

Browning, H., K. Acevedo-Whitehouse, F. Gulland, A. Hall, J. Finlayson, M. Dagleish, *et al*. 2014. Evidence for a genetic basis of urogenital carcinoma in the wild California sea lion. Proceedings of the Royal Society B 281: 20140240.

Caughley, G. and A. Gunn. 1996. Conservation Biology in Theory and Practice. Blackwell Science, Cambridge, MA.

Chen, X., W. Zhang, X. Lin, and J. Wang. 2006. Genetic instability of BRCA1 gene at locus D17S855 is related to clinicopathological behaviours of gastric cancer from Chinese population. World J Gastroenterol 12: 4246-4249.

Clapham, P.J., B.Y. Sharon and L.B. Robert. 1999. Baleen whales: conservation issues and status of the most endangered populations. Mammal Review 29: 35-60.

Crow, M. 1948. Alternative hypotheses of hybrid vigor. Genetics 33: 477-487.

COSEWIC. 2004. COSEWIC assessment and update status report on the beluga whale Delphinapterus leucas in Canada. Committee on the Status of Endangered Wildlife in Canada. Ottawa. 70p.

Culp, S., D. Gaylor, W. Sheldon, L. Goldstein and F. Beland. 1998. A comparison of the tumors induced by coal tar and benzo[a]pyrene in a 2-year bioassay. Carcinogenesis 19: 117-124.

Dalcourt M.F., P. Beland, E. Pelletier and Y. Vigneault Y. 1992. Caractarisation des communautacs benthiques et actude des contaminants dans des aires fracquentaces par le bacluga du Saint-Laurent. Department of Fisheries and Oceans Canada. Rapp Tech Can Sci Halieut Aquat 1845: 1-86.

Dawson, D.A., O. Hanotte, C. Greig, I.R.K. Stewart, and T. Burke. 2000. Polymorphic microsatellites in the blue tit Parus caeruleus and their cross-species utility in 20 songbird families. Molecular Ecology 9: 1941-1944.

De Guise, S., A. Lagace and P. Beland. 1994. Tumors in St. Lawrence beluga whale (*Delphinapterus leucas*). Vet Pathology 31: 444-449.

de March, B. G. E. and L. D. Postma. 2003. Molecular genetic stock discrimination of belugas (*Delphinapterus leucas*) hunted in eastern Hudson bay, northern Québec, Hudson strait, and Sanikiluaq (Belcher island), Canada, and comparisons to adjacent populations. Arctic 56: 111-124.

de March, B. G. E., L. D. Maiers and M. K. Friesen. 2002. An overview of genetic relationships of Canadian and adjacent populations of belugas (*Delphinapterus leucas*) with emphasis on Baffin Bay and Canadian eastern Arctic populations. NAMMCO Scientific Publications 4:17-38.

DFO. 2005. Recovery Assessment Report on Cumberland Sound, Ungava Bay, Eastern Hudson Bay, and St. Lawrence Beluga Populations (*Delphinapterus leucas*). Canadian Science Advisory Secretariat. Science Advisory Report 2005/036. 14p.

DFO. 2012. Recovery strategy for the beluga whales (*Delphinapterus leucas*) St. Lawrence estuary population in Canada. Species at Risk Act Recovery Strategy Series. Fisheries and Oceans Canada, Ottawa.

Dommering, C., T. Marees, A. van der Hout, S. Imhof, H. Meijers-Heijboer, P. Ringens *et al*. 2012. RB1 mutations and second primary malignancies after hereditary retinoblastoma. Familial Cancer 11(2): 225-233.

Driscoll, E. E., J.I. Hoffman, L.E. Green, G.F. Medley and W. Amos. 2011. A preliminary study of genetic factors that influence susceptibility to bovine tuberculosis in the British cattle herd. PLoS One 6: e18806.

Dumont, P., J. Leu, A. Pietra, D. George and M. Murphy. 2003. The codon 72 polymorphic variants of p53 have markedly different apoptotic potential. Nature Genetics 33: 357-364.

Forstmeier, W., H. Schielzeth, J. Mueller, H. Ellegren and B. Kempenaers. 2012. Heterozygosity-fitness correlations in zebra finches: microsatellite markers can be better than their reputation. Molecular Ecology 21: 3237-3249.

Frankham, R., K. Lees and M. Montgomery. 1999. Do population size bottlenecks reduce evolutionary potential? Animal Conservation 2: 255-260.

Frankham, R. 2003. Genetics and conservation biology. Comptes Rendus Biologies 326: S22-S29.

Gemignani, F., V. Moreno, S. Landi, N. Moullan, A. Chabrier, and S. Gutierrez-Enriques. 2004. A TP53 polymorphism is associated with increased risk of colorectal cancer and with reduced levels of TP53 mRNA. Oncogene 23: 1954-1956.

Hammill, M. O., L. N. Measures, J.-F. Gosselin and V. Lesage. 2007. Absence de rétablissement du béluga de l'estuaire du Saint-Laurent. Secrétariat canadien de consultation scientifique. Document de recherche 2007/026. 23p.

Harshbarger, J. C. and M.S. Slatick. 2001. Lesser known aquarium fish tumor models. Marine Biotechnology 3: S115–S129.

Hendrick, W. and T. Kalinowski. 2000. Inbreeding depression in conservation biology. Annual Review of Ecology and Systematics 31: 139-162.

Hendricks, J. D., T.R. Meyers, D.W. Shelton, J.L. Casteel and G.S. Bailey. 1985. Hepatocarcinogenicity of benzo[a]pyrene to rainbow trout by dietary exposure and intraperitoneal injection. Journal of National Cancer Institute 74: 839–851.

Kang, H., Z. Feng, Y. Sun, G. Atwal, M. Murphy, T. Rebbeck, T., et al. 2009. Single-nucleotide polymorphisms in the p53 pathway regulate fertility in humans. PNAS 106(24): 9761-9766.

Keane, M., Semeiks, J., Webb, A., Li, Y., Quesada, V., Craig, T. *et al*. 2015. Insight into the evolution of longevity from the bowhead whale genome. Cell Reports 10(1): 112-122.

Keller, L. F., P. Aecese, J.N.M. Smith, W.M. Hochachka and S.C. Stearns. 1994. Selection against inbred song sparrows during a natural population bottleneck. Nature 372: 356–357.

Kielbohmer, W. 2001. Polycyclic Aromatic Hydrocarbon (PAH) Metabolites. Environmental Analysis (Volume 3 of Handbook of Analytical Separations). Elsevier. Pp. 99-122.

Knapp, L., J. Ha, amd G. Sackett. 1996. Parental MHC antigen sharing and pregnancy wastage in captive pigtailed macaques. Journal of Reproductive Immunology 32: 73-88.

Lacy, R. C. and D.B. Lindenmayer. 1995. A simulation study of the impacts of population subdivision on the mountain bushtail possum Trichosurus caninus Ogilby (*Phalangeridae, Marsupialia*) in south-eastern Australia. 2. Loss of genetic variability within and between subpopulations. Biological Conservation 73: 131-142.

Lande, R. 1988. Genetics and demography in biological conservation. Science 241: 1455-1460.

Lebeuf, M., L. Measures, M. Noël, M. Raach and S. Trottier. 2014. A twenty-one year temporal trend of persistent organic pollutants in St. Lawrence Estuary beluga Canada. Science and the Total Environment 485: 377–386.

Lesage, V. and M. C. S. Kingsley. 1998. Updated status of the St. Lawrence river population of the beluga, Delphinapterus leucas. Canadian Field Naturalist 112: 98-114.

Lesueur, F., H. Song, S. Ahmed, C. Luccarini, C. Jordon, R. Luben, *et al*. 2006. Single-nucleotide polymorphisms in the Rb1 gene and associated with breast cancer in the British population. British Journal of Cancer 94: 1921-1926.

Lohmann, D., B. Brandt, W. Hopping, E. Passarge and B. Horsthemke. 1996. The spectrum of RB1 germ-line mutations in hereditary retinoblastoma. American Journal of Human Genetics 58: 940-949.

Malins, D. C., M.M. Krahn, D.W. Brown, L.D. Rhodes, M.S. Myers, B.B. McCain and S.L. Chan. 1985. Toxic chemicals in marine sediment and biota from Mukilteo, Washington: relationships with hepatic neoplasms and other hepatic lesions in English sole (*Parophrys vetulus*). Journal of National Cancer Institute 74: 487–494.

Mancuso, S. 1995. Population genetics of Hudson Bay beluga 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.

Martineau, D., A. Lagace, P. Beland, D. Higgins, L. Armstrong, and R. Shugart. 1988. Pathology of stranded beluga whales (*Delphinapterus leucas*) from the St. Lawrence estuary, Quebec, Canada. Journal of Comparitive Pathology 98: 287-311.

Martineau, D., K. Lemberger, A. Dallaire, P. Labelle, T. Lipscomb, P. Michel and I. Mikaelian. 2002. Cancer in wildlife, a case study: Beluga from the St. Lawrence estuary, Quebec, Canada. Environmental Health Perspectives 110(3): 285-292.

Moyle, L.C., J.R. Stinchcomeb, B.R. Hudgens and W.F. Morris. 2003. Conservation genetics in the recovery of endangered animal species: a review of US endangered species recovery plans (1977-1998). Animal Biodiversity and Conservation 26(2): 85-95.

Murray, B. W., R. Michaud and B. N. White. 1999. Allelic and haplotype variation of major histocompatibility complex class II DRB1 and DQB loci in St. Lawrence beluga, (*Delphinapterus leucas*). Molecular Ecology 8: 1127-1139.

Newman, S. and S. Smith. 2006. Marine mammal neoplasia: a review. Vet pathology 43: 865-880.

Osorio, A., M. Hoya, R. Rofrigues-Lopez, J. Granizo, O. Diex, A. Vega, *et al*. 2003. Over-representation of two specific haplotypes among chromosomes harbouring BRCA1 mutations. European Journal of Human Genetics 11: 489-492.

Patenaude, N., J. Quinn, P. Beland, M. Kingsley and B. White. 1994. Genetic variation of the St. Lawrence beluga whale population assessed by DNA fingerprinting. Molecular Ecology 3: 375-381.

Paoloni, M. and C. Khanna. 2008. Translation of new cancer treatments from pet dogs to humans. Nature Review Cancer 8: 147-156.

Pfeifer, G., M. Denissenko, M. Olivier, N. Tretyakova, S. Hecht, and P. Hainaut. 2002. Tobacco smoke carcinogens, DNA damage and p53 mutations in smoking-associated cancers. Oncogene 21(48): 7435-7451.

Ralls, K., J. Ballou, and A. Templeton. 1988. Estimates of lethal equivalents and the cost of inbreeding in mammals. Conservation Biology 2: 185-193.

Rijks, J., J. Hoffman, T. Kuiken, A. Osterhaus, and W. Amos. 2008. Heterozygosity and lungworm burden in seals (*Phoca vitulina*). Heridity 100: 587-593.

Reeves, R. R. and E. Mitchell. 1984. Catch history and initial population of white whales (Delphinapterus leucas) in the river and Gulf of the St.Lawrence, eastern Canada. Naturaliste Canadien 111: 63-121.

Rocha, R., P. Clapham and Y. Ivashchenko. 2015. Emptying the oceans: a summary of industrial whaling catches in the 20th century. Marine Fisheries Review 76(4): 37-47.

Shugart LR, Martineau D, Beland P. Detection and quantitation of benzo[a]pyrene adducts in brain and liver tissue of beluga whales (Delphinapterus leucas). 1990. In: Proceedings of the International Forum for the Future of the Beluga. Tadoussac, Quebec, Canada (Prescott J, Gauquelin M, eds). Sillery:Presses de l'Universite du Quebec, 219-223.

Thakur, N., S. Hussain, V. Nasare, B. Das, S. Basir, and M. Bharadwaj. 2011. Association analysis of p16 (CDKN2A) and RB1 polymorphisms with susceptibility to cervical cancer in an Indian population. Molecular Biology Reports 39(1): 407-414.

Saccheri, I., M., M. Kuussaari, P. Kankare, W. Vikman and I. Hanski. 1998. Inbreeding and extinction in a butterfly metapopulation. Nature 392(2): 491-494.

Scharrer, E. 1983. A small whale hunt in Canada. Canadian Translation for Fisheries and Aquatic Sciences 4973: 16.

Schlötterer, C., B. Amos, and D. Tautz. 1991. Conservation of polymorphic simple sequence loci in cetacean species. Nature 354: 63-65.

Smith, I., H. Ferguson and M. Hayes. 1989. Histopathology and prevalence of epidermal papillomas epidemic in brown bullhead, *Ictalurus nebulosus* (Lesueur) and white sucker, *Catostomus commersoni* (Laceped) populations from Ontario, Canada. Journal of Fish Diseases 12: 373–388.

Schonewald-Cox, C.M., S.M. Chambers, B. MacBryde, and L. Thomas (eds.). 1983. Genetics and Conservation. Benjamin/Cummings, Menlo Park, California.

Spielman, D., B. W. Brook, and R. Frankham. 2004. Most species are not driven to extinction before genetic factors impact them. Proceedings of the National Academy of Sciences USA 101: 15261-15264.

Valverde, J., J. Alonso, I. Palacios and A. Pestana. 2005. RB1 gene mutation up-date, a meta-analysis based on 932 reported mutations available in a searchable database. BMC Genetics 6:53.

Volk, D.E., V. Thiviyanathan., J.S. Rice, B.A. Luxon, J.H. Shah, H. Yagi, et al. 2003. Solution structure of a cis-opened (10R)-N6-deoxyadenosine adduct of (9S,10R)-9,10-epoxy-7,8,9,10-tetrahydrobenzo[a]pyrene in a DNA duplex. Biochemistry 42(6): 1410-1420.

Vucetich, J.A., and T.A. Waite. 1998. Erosion of heterozygosity in fluctuating populations. Conservation Biology 13: 860-868.

Westemeier, R., J.D. Brawn, S.A. Simpson, T.L. Esker, R.W. Jansen, J.W. Walk, *et al*. 1998. Tracking the long-term decline and recovery of an isolated population. Science 282: 1695-1698.

Whibley, C., P. Pharoah and M. Hollstein. p53 polymorphisms: cancer implications. 2009. Nature Reviews 9: 95-107.

Gingerich P, N.A. Wells, D.E. Russell and S.M.I. Shah. 1983. Origin of whales in epicontinental remnant seas: new evidence from the early Eocene of Pakistan. Science 220: 403-406.

Wildt, D.E., E.J. Baas, P.K. Chakraborty, T.L. Wolfle and A.P. Stewart. 1982. Influence of inbreeding on reproductive performance, ejaculate quality and testicular volume in the dog. Theriogenology 17(4): 445-452.

Chapter 2

Molecular markers for the study of cancer in cetaceans

(Formatted as a Note for Marine Mammal Science)

Introduction:

In recent years, cancer has emerged as a conservation concern for many wildlife populations. Although cancer in wildlife is difficult to detect, increased research and monitoring efforts have identified a high prevalence of benign and malignant tumors across a wide range of taxa (McAloose and Newton 2009). A few examples include the Tasmanian devil facial tumor disease (McCallum et al. 2007, Pyecroft et al. 2007), sea lion genital carcinoma (Gulland et al. 1996, Browning et al. 2014), sea turtle fibropapillomatosis (Herbst 1994), and beluga carcinoma (Martineau et al. 2002). Cancer can be detrimental for the conservation of a species by directly increasing mortality rates, as well as by indirectly altering population dynamics and reducing reproductive success (McAloose and Newton 2009). In some cases (e.g., Tasmanian devil facial tumor disease) it may even threaten a species with extinction (McCallum et al. 2007).

Despite the difficulties associated with cancer detection in marine mammals, neoplastic conditions have been reported in several species. The St. Lawrence beluga (*Delphinapterus leucas*) is a prominent example, where malignant tumors have accounted for 18% of mortalities in the past (Martineau et al. 2002). Tumors have also been noted in other marine mammal species, including the sperm whale (*Physeter macrocephalus*) (Lambertsen et al. 1987), Burmeister's porpoise (*Phocoena spinipinnis*) (Van Bressem et al. 2007), dusky dolphin (*Lagenorhynchus obscurus*) (Van Bressem et al. 2000), bottlenose dolphin (*Tursiops truncatus*) (Bossartt et al. 2005), Atlantic white-sided dolphin (*Lagenorhynchus acutus*) (Geraci et al. 1987), harbor porpoise (*Phocoena phocoena*) (Van Bressem et al. 1999), killer whale (Orcinus orca) (Bossart et al. 1996),

and narwhal (*Monodon monoceros*) (Geraci et al. 1987). One study found that up to 66.7% of Dusky dolphins and 48.5% of Burmeister's porpoises had genital papillomas (Van Bressem et al. 1996), which were invasive enough to interfere with copulation (Van Bressem et al. 1996, 1999).

The etiology of cancer is multifactorial and often involves a complex interplay between environmental, genetic, and epigenetic factors. Many researchers hypothesize that environmental factors, such as increased pollution and contaminants, are at least partially responsible for the high rate of cancer seen in some wildlife populations (e.g. Martineau et al. 2002). Although the focus of cancer research in most wildlife studies has centered around environmental factors, a few studies have highlighted the role of genetic factors as well. For example, low diversity at major histocompatibility complex (MHC) class I genes is influential in the spread of Tasmanian devil facial tumor disease (Siddle et al. 2007). Sea lion genital carcinoma has also been linked to genetic differences occurring along the heparanase 2 (HPSE2) gene (Browning et al. 2014), as well as increased levels of inbreeding (Acevedo-Whitehouse et al. 2003).

Understanding the influence of genetic factors on cancer susceptibility is needed to assess if, and to what degree, genetic factors are influencing health and limiting recovery in populations threatened with a high incidence of cancer. For example, variation in genetic characteristics can make some individuals more prone than others to developing neoplasms when exposed to environmental carcinogens. Such an effect may help explain the intrapopulation variation in cancer susceptibility, despite individuals living in the

same environment. Disentangling the role of genetic factors in cancer development can also help guide conservation decisions for specific populations, providing information on the relative impacts of genetic and environmental factors on individual health, and thus indicating where conservation actions can most effectively be directed.

One way to determine whether polymorphisms within a gene are linked to cancer development is to directly sequence the gene of interest. However, this can become very costly and time consuming, particularly if the relevant gene(s) has not yet been identified, resulting in many potentially relevant genes being sequenced. Alternatively, an effective way to first narrow down the gene(s) that may be involved is through linkage analyses of candidate microsatellite loci. Microsatellite markers are ideal for linkage studies as they are highly polymorphic, relatively cheap and easy to analyze, and can exhibit cross-species utility (Schlötterer et al. 1991, Dawson et al. 2000). As a result, many studies have successfully used microsatellite markers to identify genes influencing common genetic diseases (e.g., Driscoll et al. 2011).

Here, we applied this approach and developed primers to amplify microsatellite loci within, or in close proximity to, cancer susceptibility genes, with the goal of providing a toolbox for studying the genetic basis of cancer in cetacean populations. Specifically, we describe primers for the amplification of microsatellite markers linked to key tumor suppressor genes known to play a critical role in cancer resistance in humans (p16, BRCA1, PTEN, BRCA2, Rb1, FOXO3), as well as the HPSE2 gene, which has recently been linked to the development of cancer in the California sea lion (Browning et al.

2014). Polymorphisms occurring along these genes may influence cancer susceptibility because they play an important role in cancer prevention (Table 2.1). In addition to these microsatellite markers, we also developed primers for amplifying a large section of the p53 gene (spanning exons 2-9). The p53 gene is arguably the most important gene involved in cancer (Levine et al. 1991), and sequencing this gene may be useful for understanding cancer development. Taken together, the markers described here should have a broad utility for studying genetic aspects of cancer in cetaceans, and may provide a strong starting point for elucidating the genetic etiology of cancer in wildlife.

Methods:

Candidate microsatellite markers were identified by first locating candidate genes within the minke whale (*Balaenoptera acutorostrata*) genome on GenBank (whole genome accession number ATDI00000000.1, Yim et al. 2013). We did this by aligning the nucleotide sequence of cetacean and human mRNA of a chosen gene with the minke whale genome using BLAST (Basic Local Alignment Search Tool, Altschul et al. 1990). Comparing mRNA to a genomic sequence allowed us to determine where the candidate genes were located. Subsequently, we searched appropriate sections of the minke whale genome for microsatellites, including the introns of the candidate genes and adjacent regions located no more than 100,000 base pairs away from the gene. This distance was chosen because previous studies in humans have shown that the size of linkage disequilibrium can extend to 100,000 base pairs or more (Taillon-Miller et al. 2000, Abecasis et al. 2001). We also limited our search to microsatellites containing the core sequence repeated from 6 to 30 times. This criterion was chosen because microsatellites

with too few repeats are likely to have limited variability (Wren et al. 2000), and those with very long repeats may have increased mutation rates and incidences of homoplasy (Wierdl et al. 1997), which would limit their informativeness in linkage studies. Once suitable microsatellites were detected, we cross-referenced each region from the minke whale genome with up to four genomes from other cetacean species available on GenBank (Table 2.2). We ensured that chosen microsatellites with fewer than 8 tandem repeats in the minke reference

genome had higher numbers of repeats in other cetaceans. For the four candidate markers identified, we found at least one genomic sequence for which the number of tandem repeats exceeded 17 repeats, suggesting that these microsatellites may be ideal for linkage analysis despite their low number of tandem repeats in the reference minke genome. For all markers, we made sure to compare the genomic sequences of at least one mysticete and one odontocete species to capture the variation between the major lineages of cetaceans. We noted all positions that had variation between species, and created primers in conserved flanking regions. For the p53 gene, we used the mRNA of beluga (accession number AF475081) and compared it to the mRNA (accession number X02469) and genomic sequence (accession number NC_000017) of human to determine the start and end of each exon. We then designed primers in appropriate exonic regions based on the sequence in beluga. Table 2.2 lists the accession codes of all mRNA and genomic sequences used for this process, and Figure 2.1 shows the location of the 13 selected microsatellite loci.

For microsatellite markers, each primer pair was tested for amplification in beluga whales, long-finned pilot whales (*Globicephala melas*), humpback whales (*Megaptera novaeangliae*), and gray whales (*Eschrichtius robustus*), using 10 samples for each species. To test if these primers successfully amplify in more distant members of cetartiodacyla, we also tested each primer set on 10 samples of horse (*Equus ferus caballus*), and one sample from cow (*Bos taurus*). For p53 markers, we tested each primer pair on 10 samples of beluga. Each 15µl PCR cocktail included 2µL (10ng) of template DNA, 1× PCR Buffer, 0.2 mM of each dNTP, 1.5 mM $MgCl_2$, 0.3 mg/mL bovine serum albumin, and 0.05 U/µL Taq DNA polymerase (Promega Corporation). The PCR amplification conditions were as follows: an initial denaturing step of 95°C for 5 min; followed by 30 cycles of 95°C for 30 sec, an annealing temperature of 55°C for 1 min, 72°C for 1 min; with a final extension period of 60°C for 45 min. PCR products were size-separated on 2% agarose gels, and visualized by staining with ethidium bromide and UV illumination. Primers that did not amplify under these initial conditions were subsequently tested at 50°C and 60°C annealing temperatures, with all other conditions kept constant.

Results and Discussion:

Thirteen microsatellite loci amplified successfully for all cetaceans tested (Table 2.3). Because these primers amplified across all tested mysticetes and odontocetes, they will likely amplify most, if not all, cetacean species. Approximately half of the primers successfully amplified in cow and horse (Table 2.3), suggesting that they may also be useful for the study of more distant species in cetartiodactyla and perissodactyla.

To determine which loci were polymorphic, 100 beluga whale samples were amplified (90 from the St. Lawrence Estuary and 10 from the Arctic). Nine loci were polymorphic in the St. Lawrence population, and 10 were polymorphic in the Arctic samples (Table 2.3). It is important to note that the St. Lawrence population has fairly low genetic variability (Patenaude et al. 1994) and therefore this test likely represents an underestimate of how polymorphic these markers are for this species as a whole. Further, because only 10 individuals from the Arctic population were amplified, it is likely that not all alleles present in this population were detected. For these reasons, some of the loci found to be monomorphic in our sample set may be polymorphic in other populations and species. In addition to being used for linkage analysis, these loci may also be useful as general markers in studies regarding inbreeding, population structure, and population biology in general.

Our p53 primers amplified successfully for beluga (Table 2.3). These primers aligned with several cetacean species on GenBank, suggesting that they may amplify successfully for other species as well. The regions amplified by our primers include both the DNA binding domain (exon 5-8) as well as regions where previously studied cancer-linked polymorphisms reside (e.g. codon 72 in exon 4; Whibley et al. 2009). We sequenced 50 beluga individuals to determine whether selected regions were variable, and found two different C/T polymorphisms occurring along intron 7 and a G/A polymorphism in exon 8 (Table 2.3).

Although all of our primers were designed for the study of cetaceans, we hope that they will have a broad utility for the study of cancer across a wide range of mammalian species. For example, cetaceans can be used as an ideal model to study cancer resistance. Due to the long lifespan of many cetaceans (e.g., bowhead whales, *Balaena mysticetus*, are estimated to live over 200 years, George et al. 1999), researchers hypothesize that these animals may have molecular adaptations that protect them against age-related diseases such as cancer (Keane et al. 2015). Identifying genetic characteristics involved in cancer resistance may provide a broader understanding of the evolution of mammalian longevity and disease resistance. Further, as many of our primers worked on horse and cow (Table 2.3), these markers may also be a useful tool for studying other mammalian populations. A high incidence of cancer has previously been noted in a cattle population (Jarrett et al. 1978), and increased monitoring may reveal the presence of cancer in subsequent ungulate populations as well.

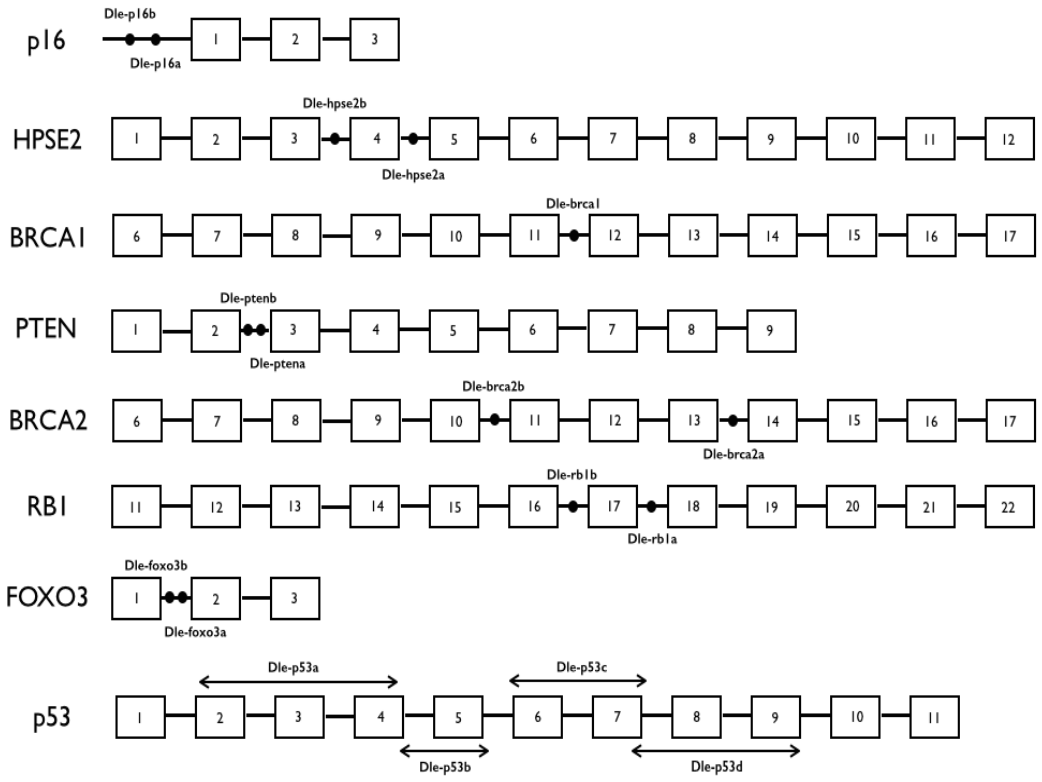

Figure 2.1. For each gene, exons are represented by boxes and introns by a line. Note that BRCA1, BRCA2, and RB1, have 23, 27, and 27 exons, respectively, but only some are depicted in this figure. Dle-p16b and Dle-p16a are located ~20,000 and ~17,000 base pairs away from the p16 gene. The lengths of introns and exons are not drawn to scale.

Table 2.1. Examples of various cancers associated with polymorphisms occurring within 8 genes of interest.

Gene	Cancer
p16	Melanoma (Mantelli et al. 2002); pancreatic cancer (Bartsch et al. 2002), breast cancer (Borg et al. 2000); neurofibroma (Petronzelli et al. 2001); glioma (Tachibana et al. 2000)
HPSE2	Urogenital carcinoma in sea lion (Browning et al. 2014)
BRCA1	Breast cancer and ovarian cancer (Easton et al. 1995); pancreatic cancer (Thompson et al. 2002)
PTEN	Breast cancer (Mills et al. 2001); gastric cancer (Canbay et al. 2013); colon cancer (Canbay et al. 2013)
BRCA2	Breast cancer and ovarian cancer (Antoniou et al. 2003); prostate cancer, gallbladder and bile duct cancer, stomach cancer, and melanoma (Breast Cancer Linkage Consortium 1999)
RB1	Retinoblastoma (Lohmann et al. 1996); breast cancer (Lesuer et al. 2006); cervical cancer (Thakur et al. 2012); sarcoma, melanoma, and epithelial cancers (Dommering et al. 2012)
FOXO3	Thyroid cancer (Rohlen et al. 2014); acute lymphoblastic leukemia (Wang et al. 2014)
p53	Ovarian cancer (Wang-Gohrke et al. 1999); breast cancer (Costa et al. 2008); lung cancer (Boldrini et al. 2008); colorectal cancer (Gemignani et al. 2004); cervical cancer (Jee et al. 2004)

Table 2.2. Accession codes for mRNA and genomic sequences used to locate microsatellites in or near genes of interest.

Gene	mRNA	Genomic Sequence
p16	DQ318021.1 (*Homo sapiens*) XM_012534921.1 (*Orcinus orca*)	NW_006726353 (*Balaenoptera acutorostrata*) NW_006782029 (*Lipotes vexillifer*) NW_004438482 (*Orcinus orca*) NW_004202595 (*Tursiops truncatus*)
HPSE2	NM_021828.4 (*Homo sapiens*) XM_004268434 (*Orcinus orca*)	NW_006728682 (*Balaenoptera acutorostrata*) NW_006783561 (*Lipotes vexillifer*) NW_004438441 (*Orcinus orca*)
BRCA1	NM_007294.3 (*Homo sapiens*) XM_007102091 (*Physeter macrocephalus*)	NW_006726432.1 (*Balaenoptera acutorostrata*) NW_006775310 (*Lipotes vexillifer*) NW_004438583 (*Orcinus orca*) NW_006712931 (*Physeter macrocephalus*) NW_004197870 (*Tursiops truncatus*)
PTEN	NM_001309477.1 (*Orcinus orca*)	NW_006726643 (*Balaenoptera acutorostrata*) NW_006798376 (*Lipotes vexillifer*) NW_004438524 (*Orcinus orca*) NW_006724123 (*Physeter macrocephalus*)

Gene	mRNA	Genomic Sequence
BRCA2	NM_000059.3 (*Homo sapiens*) XM_007105456 (*Physeter macrocephalus*)	NW_006732567 (*Balaenoptera acutorostrata*) NW_006791240 (*Lipotes vexillifer*) NW_004438576 (*Orcinus orca*) NW_006713280 (*Physeter macrocephalus*)
RB1	NM_000321.2 (*Homo sapiens*) XM_004274596.2 (*Orcinus orca*)	NW_006728570 (*Balaenoptera acutorostrata*) NW_006793498 (*Lipotes vexillifer*) NW_004438477 (*Orcinus orca*) NW_006714382 (*Physeter macrocephalus*) NW_004200515 (*Tursiops truncatus*)
FOXO3	NM_001455.3 (*Homo sapiens*) XM_004264747 (*Orcinus orca*)	NW_006730789 (*Balaenoptera acutorostrata*) NW_006772017 (*Lipotes vexillifer*) NW_004438423 (*Orcinus orca*) NW_006712784 (*Physeter macrocephalus*)
p53	AF475081 (*Delphinapterus leucas*) X02469 (*Homo sapiens*)	NC_000017 (*Homo sapiens*)

Table 2.3. Characteristics of 13 microsatellite markers, and four p53 primer sets developed for the study of cancer in cetaceans. Each microsatellite primer set was tested on beluga whales (bw), long-finned pilot whales (pw), humpback whales (hw), gray whale (gw), horse (h), and cow (c); each p53 primer set was tested on beluga. The number of microsatellite alleles found in 100 genotyped beluga samples is shown for both the St. Lawrence (SLB) and Arctic (A) population. The number of p53 alleles identified from sequencing 50 SLB individuals is also shown.

Name	Repeat	Sequence	Amplified for	Number of alleles
Dle-p16b	tg(n)	F: aga agg gtt tcc tgg acc cag gca g R: ttg cag act tca gac ttg cca gcc c	bw, pw, gw, hw, c	4-SLB, 3-A

Name	Repeat	Sequence	Amplified for	Number of alleles
Dle-p16a	tg(n)	F: tcc act act gga agg agg gag gg R: ccc act tta gaa gtc agt cac tcc tg	bw, pw, gw, hw, h, c	2-SLB, 2-A
Dle-hpse2b	ca(n)	F: tcc tct tca gaa ccc tga cag ag R: ctc agg tac tat cac aat gca aag tg	bw, pw, gw, hw, h	7-SLB, 6-A
Dle-hpse2a	tc(n)	F: gag act cat gga atc tac ag R: cca tgc ttg gtt taa tgc tct gc	bw, pw, gw, hw	3-SLB, 4-A
Dle-brca1	gt(n)	F: ctc tga ata tcc ttt tgt gcc ttg R: gtg tct ctt gcc ctc aaa ccc tc	bw, pw, gw, hw	1-SLB, 1-A

Name	Repeat	Sequence	Amplified for	Number of alleles
Dle-ptenb	ca(n)	F: cta cct tta tga tta cta gta ctg R: tcc ttt ctc ctg gtt acc aa ttt g	bw, pw, gw, hw, h, c	5-SLB, 5-SLB
Dle-ptena	ca(n)	F: ggt aca gaa cag ttc tca agc tat gtc R: ctg tga cca gct ggt aat atg aaa c	bw, pw, gw, hw, h	1-SLB, 2-A
Dle-brca2b	gt(n)	F: gca gaa aga cca aga atc aga ac R: cta aac aac aat gcc tga tac	bw, pw, gw, hw, c	5-SLB, 5-SLB
Dle-brca2a	gt(n)	F: gtg taa aat ggt tgg ttc tct cct c R: ggt gaa ttg agc ata ctc acc tac	bw, pw, gw, hw, h	1-SLB, 1-A

Name	Repeat	Sequence	Amplified for	Number of alleles
Dle-rb1b	gt(n)	F: cag ttt caa ata gtg ctg cac ctt g R: ttg aaa cga tgg aga tga cg	bw, pw, gw, hw	8-SLB, 6-A
Dle-rb1a	gt(n)ga(n)	F: gca gta att ccc aag aaa agc c R: gca gta gaa ttt aca cgt gta gtt g	bw, pw, gw, hw, h, c	3-SLB, 4-A
Dle-foxo3b	ca(n)	F: cac aaa att ccc atc gag ttc cag g R: tat ctg tat gct ctc agc tat gtc	bw, pw, gw, hw, h, c	1-SLB, 1-A
Dle-foxo3a	ca(n)	F: tgg ttt gta agg gag aaa tgc c R: gga taa ggg act ttg gct tgg ac	bw, pw, gw, hw	6-SLB, 7-A
Dle-p53a		F: agt cgc agg cag aac tcg R: gtg cag gtt aca gac ttg g	bw	1-SLB

Name	Repeat	Sequence	Amplified for	Number of alleles
Dle-p53b		F: gaa cag cca agt ctg taa cc R: tcg cta tag tca gag cag c	bw	1-SLB
Dle-p53c		F: gga agg gaa ttt acg tgc tga g R: cca gtg tga tga tgg tga gg	bw	1-SLB
Dle-p53d		F: cct cac cat cat cac act gg R: gag tga aat att ctc cat cca gtg g	bw	6-SLB

Literature Cited:

Abecasis, G. R., E. Noguchi, A. Heinzmann, J.A. Traherne, S. Bhattacharyya, N.I. Leaves, *et al*. 2001. Extent and distribution of linkage disequilibrium in three genomic regions. American Journal of Human Genetics 68: 191-197.

Acevedo-Whitehouse, K. F., D. Gulland, D. Greig, and W. Amos. 2003. Inbreeding: Disease susceptibility in California sea lions. Nature 422:35.

Altschul, S., W. Gish, W. Miller, E. Myers, and D. Lipman. 1990. Basic local alignment search tool. Journal of Molecular Biology 215: 403-410.

Antoniou, A., P.D. Pharoah, S. Narod, et al. 2003. Average risks of breast and ovarian cancer associated with BRCA1 or BRCA2 mutations detected in case series unselected for family history: a combined analysis of 22 studies. American Journal of Human Genetics 72:1117-1130.

Bartsch D.K., M. Sina-Frey, S. Lang, et al. 2002. CDKN2A germline mutations in familial pancreatic cancer. Annals of Surgery 236:730-737.

Beland, P., S. Guise, C. Girard, et al. 1993. Toxic compounds and health and reproduction effects in St. Lawrence beluga whales. Journal of Great Lakes Research 19:766-775.

Boldrini, L., S. Gisfredi, S. Ursino, et al. 2008. Effect of the p53 codon 72 and intron 3 polymorphisms on non-small cell lung cancer (NSCLC) prognosis. Cancer Investigation 26:168-172.

Borg, A., T. Sandberg, K. Nilsson, et al. 2000. High frequency of multiple melanomas and breast and pancreas carcinomas in CDKN2A mutations-positive melanoma families. Journal of the National Cancer Institute 92:1260-1266.

Bossart, G. D., C. Cray, L. Solorzano, S.J. Decker, L.H. Cornell, and N.H. Altman. 1996. Cutaneous papillomaviral-like papillomatosis in a killer whale. Marine Mammal Science 12:274-281.

Bossart, G. D., G. Shin-je, M. Rehtanz, et al. 2005. Orogenital neoplasia in Atlantic bottlenose dolphins (Tursiops truncatus). Aquatic Mammals 31:473-480.

Breast Cancer Linkage Consortium. 1999. Cancer risks in BRCA2 mutation carriers. Journal of the National Cancer Institute 91:1310-1316.

Browning, H., K. Acevedo-Whitehouse, F. Gulland, et al. 2014. Evidence for a genetic basis of urogenital carcinoma in the wild California sea lion. Proceedings of the Royal Society of London B 281: 20140240.

Canbay E, O.T. Kahraman, D. Bugra, et al. 2013. Increased Gastric Cancer Risk with PTEN IVS4 Polymorphism in a Turkish Population. Genetic Testing and Molecular Biomarkers 17: 249-253.

Canbay E, O.T. Kahraman, D. Bugra, et al. 2013. Association between PTEN IVS4 polymorphism and development of colorectal cancer in a Turkish population. Expert Opinion on Therapeutic Targets 17:1-6.

Costa, S., D. Pinto, D. Pereira, H. Rodrigues, J. Cameselle-Teijeiro, R. Medeiros, and F. Schmitt. 2008. Importance of TP53 codon 72 and intron 3 duplication 16bp polymorphisms in prediction of susceptibility on breast cancer. BMC Cancer 8:32.

Dawson, D.A., O. Hanotte, C. Greig, I.R.K. Stewart, and T. Burke. 2000. Polymorphic microsatellites in the blue tit *Parus caeruleus* and their cross-species utility in 20 songbird families. Molecular Ecology 9:1941-1944.

Dommering, C., T. Marees, A. van der Hout, et al. 2012. RB1 mutations and second primary malignancies after hereditary retinoblastoma. Familial Cancer 11:225-233.

Driscoll, E. E., J.I. Hoffman, L.E. Green, G.F. Medley, and W. Amos. 2011. A preliminary study of genetic factors that influence susceptibility to bovine tuberculosis in the British cattle herd. PLoS One 6:e18806.

Easton, D.F., D. Ford, and D.T. Bishop. 1995. Breast and ovarian cancer incidence in BRCA1-mutation carriers. Breast Cancer Linkage Consortium. American Journal Human Genetics 56:265-271.

George, J.C., J. Bada, J. Zeh, L. Scott, S.E. Brown, T. O'Hara, and R. Suydam. 1999. Age and growth estimates of bowhead whales (*Balaena mysticetus*) via aspartic acid racemization. Canadian Journal of Zoology 77:571-580.

Geraci, J., N. Palmer, and D. St Aubin. 1987. Tumors in cetaceans: analysis and new findings. Canadian Journal of Fisheries and Aquatic Sciences 44:1289-1300.

Gemignani, F., V. Moreno, V., Landi, et al. 2004. A TP53 polymorphism is associated with increased risk of colorectal cancer and with reduced levels of TP53 mRNA. Oncogene 23:1954-1956.

Gulland, F. M., J.G. Trupkeiwicz, T.R. Spraker, and L.J. Lowenstine. 1996. Metastatic carcinoma of probable transitional cell origin in 66 free-living California sea lions (Zalophus californianus), 1979 to 1994. Journal of Wildlife Disease 32:250-258.

Herbst, L. H. 1994. Fibropapillomatosis of marine turtles. Annual Review of Fish Diseases 4:389-425.

Jarrett, W. F. H., P.E. Mcneil, W.T.R. Grimshaw, I.E. Selman, and W.I.M. Mcintyre. 1978. High incidence area of cattle cancer with a possible interaction between an environmental carcinogen and a papilloma virus. Nature 274:215-217.

Jee, S. H., S.Y. Won, J.E. Yun, J.E. Lee, J.S. Park, and S.S. Ji. 2004. Polymorphism p53 codon-72 and invasive cervical cancer: a meta-analysis. International Journal of Gynecology & Obstetrics 85:301-308.

Keane, M., J. Semeiks, A. Webb, et al. 2015. Insight into the evolution of longevity from the bowhead whale genome. Cell Reports 10:112-122.

Lambertsen, R., B.A. Kohn, J.P. Sundberg, and C.D. Buergelt. 1987. Genital papillomatosis in sperm whale bulls. Journal of Wildlife Diseases 23:361-367.

Lesueur, F., H. Song, and S. Ahmed. 2006. Single-nucleotide polymorphisms in the Rb1 gene and associated with breast cancer in the British population. British Journal of Cancer 94: 1921-1926.

Levine, A., J. Momand, and C. Finlay. 1991. The p53 tumor suppressor gene. Nature 351:453-456.

Lohmann, D., B. Brandt, W. Hopping, E. Passarge, B. Horsthemke. 1996. The spectrum of RB1 germ-line mutations in hereditary retinoblastoma. American Journal Human Genetics 58:940-949.

Mantelli M., M. Barile, P. Ciotti, et al. 2002. High prevalence of the G101W germline mutation in the CDKN2A (p16(ink4a)) gene in 62 Italian malignant melanoma families. American Journal of Medical Genetics 107:214–221.

Martineau, D., K. Lemberger, A. Dallaire, P. Labelle, T. Lipscomb, P. Michel, and I. Mikaelian. 2002. Cancer in wildlife, a case study: Beluga from the St. Lawrence estuary, Quebec, Canada. Environmental Health Perspectives 110:285-292.

McAloose, D. and Newton, A.L. 2009. Wildlife cancer: a conservation perspective. Nature Reviews Cancer 9:517-526.

McCallum, H., D.M. Tompkins, Jones, M. et al. 2007. Distribution and impacts of Tasmanian devil facial tumor disease. EcoHealth 4:318-325.

Mills, G.B., Y. Lu, X. Fang, et al. 2001 The role of genetic abnormalities of PTEN and the phosphatidylinositol 3-kinase pathway in breast and ovarian tumorigenesis, prognosis, and therapy. Seminars in Oncology 28:125-141.

Patenaude, N., J. Quinn, P. Beland, M. Kingsley, and B White. 1994. Genetic variation of the St. Lawrence beluga whale population assessed by DNA fingerprinting. Molecular Ecology 3: 375-381.

Petronzelli F., D. Sollima, G. Coppola, M.E. Martini-Neri, G. Neri, and M. Genuardi. 2001. CDKN2A germline splicing mutations affecting both P16ink4 and P14arf RNA processing in a melanoma/ neurofibroma kindred. Genes Chromosomes Cancer 31:398-401.

Pyecroft, S.B, A. Pearse, R. Loh, et al. 2007. Towards a case definition for devil facial tumour disease: what is it? EcoHealth 4:346-351.

Rohlen, N., C. Doring, M. Hansmann, et al. 2014. FOXO3a Polymorphism rs4946936 and its interaction with vitamin D in differentiated thyroid cancer. Experimental and Clinical Endocrinology and Diabetes 122:84.

Schlötterer, C., B. Amos, and D. Tautz. 1991. Conservation of polymorphic simple sequence loci in cetacean species. Nature 354:63-65.

Siddle, H., A. Kreiss, M.D.B. Eldridge, M., et al. 2007. Transmission of a fatal clonal tumor by biting occurs due to depleted MHC diversity in a threatened carnivorous marsupial. Proceedings of the National Academy of Sciences of the United States of America 104: 16221-16226.

Spaulding, M., M.A. O'Leary, and J. Gatesy. 2009. Relationships of Cetacea (Artiodactyla) among mammals: Increased taxon sampling alters interpretations of key fossils and character evolution. PloS One 4(9):e7062.

Tachibana I., J. Smith, K. Sato, S.M. Hosek, D.W. Kimmel, and R.B. Jenkins. 2000. Investigation of germline PTEN, p53, p16(INK4A)/p14(ARF), and CDK4 alterations in familial glioma. American Journal of Medical Genetics 15:136–141.

Taillon-Miller, P., I. Bauer-Sardina, N.L. Saccone, et al. 2000. Juxtaposed regions of extensive and minimal linkage disequilibrium in human Xq25 and Xq28. Nature Genetics 25:324-328.

Thakur, N., S. Hussain, V. Nasare, B. Das, S. Basir, and M. Bharadwaj. 2011. Association analysis of p16 (CDKN2A) and RB1 polymorphisms with susceptibility to cervical cancer in an Indian population. Molecular Biology Reports 39(1): 407-414.

Thompson D., and D.F. Easton. 2002. Cancer incidence in BRCA1 mutation carriers. Journal of the National Cancer Institute 94:1358-1365.

Van Bressem, M. F., K. Van Waerebeek, G.E. Pierard, and C. Desaintes. 1996. Genital and lingual warts in small cetaceans from coastal Peru. Diseases of Aquatic Organisms 26, 1–10.

Van Bressem, M. F., R.A. Kastelein, P. Flamant, and G. Orth. 1999. Cutaneous papillomavirus infection in a harbour porpoise (*Phocoena phocoena*) from the North Sea. Veterinary Record 144:592–593.

Van Bressem, M. F., K. Van Waerebeek, and J.A. Raga. 1999. A review of virus infections of cetaceans and the potential impact of morbilliviruses, poxviruses and papillomaviruses on host population dynamics. Diseases of Aquatic Organisms 38:53–65.

Van Bressem, M. F., K. Van Waerebeek, U. Siebert, A. Wunschmann, L. Chavez-Lisambart, and J.C. Reyes. 2000. Genital diseases in the Peruvian dusky dolphin (*Lagenorhynchus obscurus*). Journal of Comparative Pathology 122:266-277.

Van Bressem, M.F., P. Cassonnet, A. Rector, C. Desaintes, K. Van Waerebeek, J. Alfraro-Shigueto, M. Van Ranst, G. Orth. 2007. Genital warts in Burmeister's porpoises: characterization of *Phocoena spinipinnis* papillomavirus type 1 (PsPV-1) and evidence for a second, distantly related PsPV. Journal of General Virology 88:1928-1933.

Wang, Y., L. Zhou, J. Chen, et al. 2014. Association of the 3'UTR FOXO3a polymorphism rs4946936 with an increased risk of childhood acute lymphoblastic leukemia in a Chinese population. Cell Physiology and Biochemistry 34:325-332.

Wang-Gohrke, S. W. Weikei, H. Risch et al. 1999. Intron variants of the p53 gene are associated with increased risk for ovarian cancer but not in carriers of BRCA1 or BRCA2 germline mutations. British Journal of Cancer 81:179-183.

Wierdl, M., M. Dominska, and T. Petes. 1997. Microsatellite instability in yeast: dependence on the length of the microsatellite. Genetics 146:769-779.

Whibley, C., P. Pharoah, and M. Hollstein. 2009. p53 polymorphisms: cancer implications. 2009. Nature Reviews 9:95-107.

Wren J.D., E. Forgacs, J.W. Fondon et al. 2000. Repeat polymorphisms within gene regions: Phenotypic and evolutionary implications. The American Journal of Human Genetics 67:345-356.

Yim, H., Y.S. Cho, X. Guang, et al. 2013. Minke whale genome and aquatic adaptation in cetaceans. Nature Genetics 46:88-92.

Chapter 3:

Are genetic factors influencing cancer susceptibility and mortality patterns in the St. Lawrence Beluga?

Abstract:

The objective of this project was to examine how genetic factors may be influencing cancer susceptibility and other mortality patterns in the St. Lawrence beluga (SLB) population. First, to assess whether variation along key genes was influencing cancer development, I compared the DNA of cancer individuals to those that died of other causes (control group). Specifically, I sequenced the p53 gene, and conducted microsatellite association analysis on seven additional candidate genes. I did not find any association between variation along the p53 gene and the presence of cancer. I also did not find any credible association between specific microsatellite alleles and cancer, or a credible combined effect of microsatellite alleles across loci. However, with the microsatellite data, a few patterns emerged that warrant further investigation as more necropsy data for more individuals become available. Second, to assess how inbreeding may be influencing mortality patterns, I compared the level of homozygosity (a proxy for measuring inbreeding) among individuals who died from different causes. Results showed that although there was no association between the level of inbreeding and cancer susceptibility, bacteria-induced mortality was correlated with a higher level of inbreeding, while parasite-induced mortality was correlated with a lower level of inbreeding, compared to other mortality categories. Future tests should continue to look into how inbreeding may be influencing bacterial and parasite susceptibility. This in turn can help assess the relative impact that a low level of genetic variability may have on population dynamics and recovery potential in the SLB.

Introduction:

In recent years, cancer has emerged as a conservation threat for several wildlife populations (McAloose and Newton, 2009). Although difficult to detect, increased research and monitoring efforts have identified a high prevalence of benign and malignant tumours across a range of mammalian taxa (McAloose and Newton, 2009). A prominent example of a population that suffers from a high rate of cancer is the St. Lawrence beluga (SLB). This population has been drastically hunted in the past, which is presumed to have decreased their population size by 90% (DFO, 2012). However, despite receiving protection 1979 (Lesage and Kingsley, 1998), the SLB have shown no signs of

recovery, with a recent population estimate of only 889 individuals (Mosnier et al., 2015). Although the factors limiting recovery are not explicitly known, a high rate of cancer has been suggested as a threat to population growth (Martineau et al., 2002). For example, necropsy data from 1983 to 2006 (n=175) show that cancer accounted for 20% of the mortalities found in the adult population (DFO, 2012). Although the prevalence of cancer in the SLB has decreased in recent years (Mosnier et al., 2015), understanding the etiology and spread of this disease is important to effectively manage this population, and to mitigate future outbreaks.

Previous research on the etiology of cancer in the SLB has focused on the effect of environmental factors. Because the St. Lawrence beluga live downstream of the highly industrialized Great Lakes Region, many scientists presume that their high cancer rate is caused by anthropogenic contaminants found in their environment (De Guise et al., 1994; Martineau et al., 2002; Newman and Smith, 2006). Indeed, toxicological studies have shown that the tissues of St. Lawrence belugas have high levels of organochlorines (OCs) and polycyclic aromatic hydrocarbons (PAHs) (Martineau et al. 1988). The correlation of contaminants with the presence of neoplasm and other lesions have led some scientists to suggest that the contaminants are either directly acting as carcinogens, or indirectly leading to cancer by weakening the immune systems of belugas (DFO, 2012). However, although all belugas in this region are presumably subjected to similar amounts of contaminants, there is a wide range of responses. For example, many individuals are dying of cancer at a relatively young age, while others are living and reproducing well

into old age. This differential response, despite similar exposure, suggests that genetic characteristics may play an important role in cancer susceptibility.

The main purpose of this project is to understand if, and to what extent, genetic factors may be influencing cancer susceptibility in the SLB. To address this, I focused on understanding whether variation in key cancer-related genes are linked to the presence or absence of cancer. I did this by sequencing the p53 gene, as well as by conducting microsatellite association analysis on the following 7 genes: p16, RB1, BRCA1, BRCA2, PTEN, HPSE2, and FOXO3. We chose to focus on the p53 gene because the p53 pathway plays a critical role in tumor prevention (e.g. Kang, 2009) and is the most frequently mutated gene in the presence of (human) tumors. Similarly, p16, BRCA1, PTEN, BRCA2, RB1 and FOXO3 are also tumor suppressor genes known to play a critical role in cancer resistance. Lastly, variation along the HPSE2 gene has recently been linked to cancer susceptibility in the California sea lion (*Zalophus californianus*) population (Browning et al. 2014). Given that belugas and sea lions are both marine mammals living in similar environments, this gene may play a similar role in influencing cancer development in the SLB as well. Further justification for the choice of these specific genes can be found in Chapter 2.

The second aim of this project is to understand how inbreeding may be influencing cancer susceptibility and other mortality patterns in the SLB. A small population comprised of 889 individuals (Mosnier et al., 2015) may make inbreeding more likely than would be expected in a larger population. Increased inbreeding is known to

negatively influence various fitness-associated traits (Forstmeier et al., 2012), and can further reduce the recovery potential of small populations (e.g. Saccheri et al., 1998). Because the degree of inbreeding has been linked to cancer susceptibility in some species (e.g. California sea lion, Acevedo-Whitehouse et al., 2003), I first tested for an association between inbreeding and cancer susceptibility. I further tested for associations between inbreeding and other mortality categories, including juvenile mortality, parasite and bacterial infection, dystocia (i.e., difficult delivery), and neonatal birth. Although inbreeding and/or low genetic variation is listed as a potential threat to the SLB population (COSEWIC, 2004), this is the first study to look into how inbreeding may be influencing the different mortality patterns documented in this population.

Taken together, the overall aim of this project is to answer the following two questions: 1) Does variation in key cancer-related genes influence cancer susceptibility, and 2) Does degree of inbreeding influence cancer susceptibility and/or other mortality patterns in the SLB population? Determining the influence of genetic factors on cancer susceptibility and mortality patterns is needed to assess if, and to what degree, genetic factors are influencing health and limiting the recovery of the SLB. For example, variation in genetic characteristics may make some individuals more prone than others to developing neoplasms when exposed to environmental carcinogens. Such an effect may help explain the intrapopulation variation in cancer susceptibility, despite individuals living in the same environment. Further, understanding how inbreeding may be influencing mortality patterns can help us understand the extent to which genetic factors may be limiting recovery for this population. Disentangling the role that genetic factors may play in

influencing cancer development and other mortality patterns can help guide conservation decisions, providing information on the relative impacts of genetic and environmental factors on individual health, and thus indicating where conservation actions can most effectively be directed in the future.

Materials and Methods:

Sample Collection and Preparation:

Necropsy data (including age and cause of death) were provided by the Carcass Monitoring Program which is carried out by Fisheries and Oceans Canada (DFO), Parks Canada, St. Lawrence National Institute of Ecotoxicology, and several other partners (DFO, 2012). A majority of the data from this program comes from the analysis of beluga carcasses that are found along the shores of the St. Lawrence. From 1983 to 2008 alone, 389 beluga carcasses were found, which averages to about 15 carcasses per year (DFO, 2012). Skin samples from necropsied individuals were provided by the Group for Research and Education on Marine Mammals (GREMM), which has been conducting long-term research on the St. Lawrence beluga population for the last 25 years. In total, I obtained skin samples from **139** individuals collected from 1983 to 2014. For these individuals, necropsy data showed the following causes of death: cancer (n=27), bacterial infection (n=14), parasite infection (n=31), neonatal (n=9), undetermined (n=29), trauma (n=6), starvation (n=4), degenerative changes (n=3), dystocia (n=14), fishing gear (n=2),

and viral infection (n=1). DNA was extracted from these samples using a standard phenol:chloroform protocol as described in Wang et al. (2008).

Statistical Approach:

I used Bayesian methods for all statistical analyses to gain the most reliable information about the data (Wade, 2000; Kruschke, 2013), and to avoid problems associated with p-values (Johnson, 1999; Halsey et al., 2015). Bayesian methods provide an alternative way to analyze data that remedies many of the problems encountered with standard null-hypothesis significance testing. Whereas null hypothesis testing estimates the probability of data having occurred given a particular hypothesis, Bayesian analysis provides a measure of the probability of an hypothesis being true, given the data (Wade, 2000). Bayesian statistical inference can be especially useful in conservation research and management (Ellison 1996, Ludwig 1996, Wolfson et al. 1996). Because this approach analyzes the probability of a hypothesis and explains the amount of certainty in each parameter estimate, it provides a more comprehensive understanding of ecological patterns or management implications than a frequentist approach based on p-values (e.g. Wade, 2000). Not relying on p-values is also important for my analyses because I tested many hypotheses. Interpretation of statistical analyses based on p-values can become very tricky as the number of tests increases, where proposed "corrections" for multiple tests can greatly reduce power, and are not universally appropriate. However, there is no need for such corrections in Bayesian analyses (Kruschke, 2015). Further, because the etiology of cancer can be multifactorial, and likely influenced be many factors, making a

simple yes/no decision using an arbitrary cut-off at p = 0.05 did not seem like an appropriate approach. Instead, the goal was to see what hypotheses had the highest probabilities given the data, rather than coming to yes/no decisions, particularly since I am working with a small data set.

I conducted all analyses using a combination of the R statistical and programing environment (R Core Team, 2015), the JAGS program for running and sampling Markov Chain Monte Carlo processes (Plummer, 2003), and the R package runjags for communicating between the two. Further details regarding each specific analysis are provided under the different subheadings below, within the context of each relevant molecular marker and/or metric.

Bayesian inference first requires the input of prior probabilities, or the degree of confidence we have in each hypotheses before the data is seen. Prior probabilities I chose for parameters under different statistical models are shown in Appendix 3.1.

p53 Sequencing:

I amplified four sections of the p53 gene: (1) beginning of exon 2 to the end of exon 4, (2) beginning of exon 4 to the end of exon 5, (3) beginning of exon 6 to the end of exon 7, and (4) beginning of exon 7 to the end of exon 9 (Figure 2.1). The regions amplified include both the DNA binding domain (exon 5-8) as well as regions where previously studied cancer-linked polymorphisms reside (e.g. codon 72 in exon 4, Whibley et al.

2009). To amplify the desired regions, the following PCR conditions were used for all four primer sets: each 15µl PCR cocktail included 2µL (10ng) of template DNA, 1× PCR Buffer, 0.2 mM of each dNTP, 1.5 mM $MgCl_2$, 0.3 mg/mL bovine serum albumin, and 0.05 U/µL Taq DNA polymerase (Promega Corporation). The PCR amplification conditions were as follows: an initial denaturing step of 95°C for 5 min; followed by 30 cycles of 95°C for 30 sec, an annealing temperature of 55°C for 1 min, 72°C for 1 min; with a final extension period of 60°C for 45 min. Any excess dNTPs and primers were removed via an enzymatic reaction containing 5 µl amplified DNA, 0.65µl Antarctic phosphatase buffer (50 mM Bis-Tris-Propane-HCl, 1 mM $MgCl_2$, 0.1 mM $ZnCl_2$, pH 6.0), 0.1µl Antarctic phosphatase (New England Biolabs), and 0.03µl exonuclease I for each sample. Samples were then incubated for 15 minutes at 37°C and for 15 minutes at 80°C. For the sequencing reaction each 15µl reaction included the 5.78µl cocktail from the previous step, in addition to 1.5µl of Reaction Mix (Life Technologies), 3µl of Sequencing Buffer, and 1µl (at 5µM) of the forward primer. The PCR sequencing conditions were as follows: 96°C for 2 min; followed by 30 cycles of 96°C for 20 sec, 50°C for 20 sec, and 60°C for 4 min. After the sequencing reaction, the samples were desalted via an ethanol precipitation as described in Irwin et al. (2003). Sequencing products were then resuspended in 10µl of HiDi formamide (Life Technologies) for sequencing.

PCR products were size-separated and visualized on an ABI 3500xl Genetic Analyzer (Applied Biosystems). I initially sequenced 10 beluga biopsy samples in both the forward and reverse direction to identify which primer (if any) produces better sequencing results.

Sequencing with either primer worked equally well, and the remaining necropsy samples were sequenced using the forward primer. Sequences were then manually trimmed and edited using 4Peaks (Mekentosj, Amsterdam) and aligned using MEGA (version 6, Kumar *et al*., 2008).

I sequenced 27 individuals who died of cancer (cancer group) and compared them to 30 individuals that died of other causes (control group). Three individuals from the cancer group, and two from the control group were excluded from analyses as their sequences were not clear. This resulted in a total of 24 and 28 individuals for the cancer and control groups, respectively. The control group was limited to individuals that were at least 45 years old because younger individuals may have developed cancer had they lived longer, and therefore may not be adequate "non-cancer" controls. I then tested if observed single nucleotide polymorphisms (SNPs) were linked with cancer susceptibility. For these analyses I used a Bayesian alternative to a contingency test, in which individuals were organized into two groups for each locus. The first grouping had two levels, designated as cancer and control. The second grouping indicated the frequency of each allele in each group. This grouping therefore had the same number of levels as there were alleles for each SNP locus. Here, the frequency of each allele in the cancer and non-cancer groups were compared against expectations under the null hypothesis of no relationship, to see if one or more allele(s) was over- or under-represented in each group. I also conducted the same analyses using haplotypes, which took into consideration all variable sites along the exon 2 to exon 9 region of the p53 gene.

Microsatellite Association Analysis:

One way to determine whether polymorphisms within a gene are linked to cancer development is to directly sequence the gene of interest. However, this can become very costly and time consuming, particularly if the relevant gene(s) has not yet been identified, resulting in many potentially relevant genes being sequenced. Alternatively, an effective first step to narrow down the number of genes that may be involved is through analyses of candidate microsatellite loci. Microsatellite association analysis works under the following premise: if an allele at a microsatellite region is associated with the presence of cancer, it may be because it is linked to a mutation in a nearby genomic region that is influencing cancer susceptibility. Microsatellite markers are ideal for association studies as they are highly polymorphic, relatively cheap and easy to analyze, and can exhibit cross-species utility (Schlötterer et al., 1991; Davis & Strobeck, 1998). As a result, many studies have successfully used microsatellite markers to identify genes influencing genetic diseases (e.g., Driscoll et al., 2011). Here, I applied this approach to narrow down genes that may play a role in cancer susceptibility in the SLB.

I amplified 13 microsatellite loci occurring along the introns of the following genes: BRCA1, BRCA2, FOXO3, p16, PTEN, RB1, and HPSE2 (Figure 2.1). I developed multiplex reactions that allowed all 13 microsatellite loci to be amplified within 4 reactions (Appendix 3.2). Each 15µl PCR cocktail included 2µL (10ng) of template DNA, 1× PCR Buffer, 0.2 mM of each dNTP, 1.5 mM $MgCl_2$, 0.3 mg/mL bovine serum albumin, and 0.05 U/µL Taq DNA polymerase (Promega Corporation). The PCR

amplification conditions were as follows: an initial denaturing step of 95°C for 5 min; followed by 30 cycles of 95°C for 30 sec, an annealing temperature (varied for each multiplex reaction, Appendix 3.2) for 1 min, 72°C for 1 min; with a final extension period of 60°C for 45 min. Samples were prepared for capillary electrophoresis and fragment analysis on an ABI 3500xl Genetic Analyzer (Applied Biosystems) by combining 2μL of PCR product, 0.25μL GeneScan-600 LIZ size standard (ABI) and 10μL HiDi formamide (ABI). Microsatellites were scored using the GeneMarker software (SoftGenetics), and all calls were edited and/or confirmed by eye.

As with the p53 gene, to test whether specific alleles at any of these microsatellite loci were linked with cancer susceptibility, I used a Bayesian alternative to a contingency test. Individuals were organized into two groups for each locus. The first grouping had two levels, designated as cancer (n=27) or control (n=30). The second grouping indicated the frequency of each allele in each group. For these analyses, each locus was analyzed independently. This organization provides clear visualization of difference in allele frequencies (or lack thereof) between the cancer and control groups.

Haplotype association analysis:

Increased cancer susceptibility may be due to the combined effects of more than one gene. To check for "haplotype association", or to see if the combination of alleles across loci might be contributing to cancer susceptibility, I looked for structuring between the

cancer and control groups. I did this by using the R package ‘adegenet’ to conduct a discriminant analysis of principal components (DAPC) analysis (Jombart et al., 2010) of 9 polymorphic cancer-linked microsatellite loci. DAPC can be used to infer the number of clusters of genetically similar individuals. It employs a multivariate statistical approach where variance in the sample is partitioned into a between-group and within-group component, to maximize discrimination between groups. DAPC aims to provide an efficient description of genetic clusters using a few synthetic variables. There are constructed as linear combinations of the original variables (alleles) which have the largest between-group variance, and the smallest within-group variance. In DAPC, genetic data are first transformed using a principal components analysis (PCA), after which subsequent clusters are identified using discriminant analysis (DA). I used a-score optimization to determine the optimal number of principal components (=4) to maximize power of discrimination while also minimizing the risk of over-fitting, after which discriminant analysis of principle components was conducted.

I conducted a logistic regression to determine whether the groupings obtained from the DAPC analysis separated cancer individuals from the control group. The y-variable was the DAPC cluster assignment, coded as 1 or 0 indicating whether or not a pair of individuals was or was not in the same DAPC cluster, respectively. There were two x-variables: x1 was also a dichotomous variable (1 or 0) indicating whether or not a pair of individuals was in the same “cancer” group (cancer or control); and x2 was a metric variable containing the relatedness of each pair based on the neutral microsatellite loci. I then conducted this regression analysis to assess the effect of being in the same group

(cancer or control) on DAPC cluster assignment, while controlling for the effect of relatedness (estimated from the 16 neutral microsatellite loci, see below). It was important to control for relatedness in this model as this is a major factor driving the genetic similarity of individuals, and will therefore also likely influence cluster assignment.

Inbreeding Analysis:

To obtain estimates of inbreeding for each individual, I genotyped each sample at 16 presumably neutral microsatellite loci. I developed multiplex reactions that fit all 16 loci into 6 reactions (Appendix 3.3). PCR conditions, genotyping preparation, and scoring were the same as described for the analyses of the cancer-linked microsatellite loci. I used neutral microsatellites to assess level of inbreeding because the properties associated with inbreeding (e.g. increased homozygosity) of these loci should be representative of genome-wide patterns (Szulkin et al., 2010; Forstmeier et al., 2012).

To estimate the level of inbreeding, I used the R package GENHET (Coulon, 2010) to calculate the homozygosity-by-loci (HL, Aparicio et al., 2006) and internal relatedness (IR, Amos et al., (2001) metrics for each sample, which are two widely used methods for estimating the relative degree to which individuals are homozygous within a population. Since inbred individuals are expected to be relatively homozygous throughout the genome, marker heterozygosity can be used as a proxy for assessing the level of inbreeding (Forstmeier et al., 2012).

To test if the degree of inbreeding is associated with cancer susceptibility, I compared the distributions of the inbreeding metrics between the cancer and control groups. To try to narrow down any potential patterns between inbreeding and cancer development, I applied a Bayesian quadratic logistic regression in which the predicted variable was either cancer or control, and the predictor variable was the inbreeding coefficient. This approach can help us find patterns that may not be seen just by comparing between means (e.g., if cancer is associated with both low and high levels of inbreeding). To assess the effect of inbreeding on general mortality patterns, I used a Bayesian alternative to a single factor ANOVA, where the response variable was the inbreeding coefficient, and predictor categorical variables included different causes of death: cancer, bacterial infection, parasite, neonatal, and dystocia. This approach can show if there are any credible differences between the different mortality classes with respect to inbreeding, as well as allow us to compare each mortality class to all remaining (or combined) mortality classes. Lastly, to test if inbreeding influenced juvenile mortality, I regressed age on inbreeding coefficient using a Bayesian simple linear regression. I hypothesized that if juvenile mortality was influenced by the level of inbreeding, then there would be a credible negative association between increased inbreeding and age.

To compare the level of inbreeding from the St. Lawrence population to that of other beluga populations, I also genotyped 10 Arctic beluga (sampled from the same location but could be from different populations) samples at 16 neutral microsatellite loci. I conducted a separate GENHET analysis where I included these additional 10 samples.

Comparing the level of inbreeding (HL and IR values) between the Arctic and SLB group can elucidate if the entire SLB population is inbred and/or has an overall low amount of variation (in inbreeding) between individuals. Such a scenario of limited variation in inbreeding coefficients would impact the ability of my analyses to detect correlations between inbreeding and causes of mortality. If this were the case, I would expect to see higher inbreeding coefficients and lower variability for the SLB population, as compared to Arctic samples. I first compared inbreeding values between 10 Arctic samples and 139 SLB samples to gain an overall picture of the level and variance of inbreeding for both groups. I also randomly sampled 10 individuals from the SLB population a total of 10 times, and compared it with the 10 Arctic individuals. Having equal samples sizes of 10 gave a better estimate of the variation in inbreeding when comparing between groups.

Results:

p53 gene:

No variation was found for the regions spanning exons 2-4, 4-5, or 6-7 of the p53 locus. Although past research on humans and mice identified polymorphisms linked to cancer susceptibility in intron 3 and exon 4 (Dumont et al., 2003; Hu et al., 2007; Whibley et al., 2009), I did not find variation at these sites. I did, however, find three variable sites spanning exons 7-9. Specifically, position 127 (intron 7) was variable for C/T, position 202 (intron 7) was variable for C/T, and position 355 (exon 8) was variable for G/A. The single nucleotide polymorphism (SNP) occurring in exon 8 is a silent mutation that

resulted in the same amino acid (Ala) for both variants. Analyses did not show any patterns linking observed variation at the p53 locus to cancer susceptibility, for either specific alleles, or combined haplotypes (Table 3.1).

Microsatellite association results:

Nine of the cancer-linked microsatellite loci were polymorphic, with the remaining 4 (Dle-foxo3b, Dle-brca2a, Dle-brca1a, and Dle-ptena) being monomorphic for all genotyped individuals (see Figure 2.1). Of the 9 polymorphic loci, 3 potential relationships were found between specific alleles at some of these loci and cancer susceptibility. For example, the FOXO3 locus (Dle-foxo3a) has two alleles that may be affecting cancer susceptibility in opposing ways. Specifically, the 277 allele was only found in individuals with cancer, and the 279 allele is primarily found in individuals without cancer (Table 3.2). The HPSE2 locus (Dle-HPSE2b) also shows a pattern for which the 276 allele seems to be less associated with the control group (Table 3.3). Finally, analysis of the BRCA2 locus (Dle-brca2b) suggests a pattern of association in which the 303 allele is less likely to be associated with cancer, while the 305 allele is more likely to be associated with cancer (Table 3.4).

For all loci examined, statistical analyses did not show a credible difference between cancer and control groups (e.g. Table 3.2; Table 3.3; Table 3.4). This may be because our sample size is too small. With only a few individuals per allele group, a small or medium effect size may not be apparent with our analysis. For example, for groups with our

sample sizes, power analyses (not shown) found that the allele frequencies between groups would have to differ by counts of at least 10 for the 95% HDI to not encompass zero. However, the frequencies of the 277 allele at the FOXO3 locus were only 5 and 0 for the cancer and control groups, respectively, indicating that sample size was indeed too small to detect a credible difference. Therefore, we will continue microsatellite association analyses on promising loci as samples and necropsy data from more individuals become available. If a strong pattern emerges (with credible statistical differences between cancer and control groups), the next step would be to directly sequence the gene of interest, and compare between the cancer and control group, much like we did with the p53 gene.

Haplotype association results:

DAPC analysis of 9 polymorphic cancer-linked loci showed four distinct clusters (Figure 3.1). A logistic regression accounting for the effect of relatedness did not show a credible difference in separating cancer and control individuals based on clustering. However, although not "credible" the posterior probabilities were suggestive of an effect of cancer genotype on cluster assignment (Figure 3.2). If there is a true effect, increasing sample size may help gain better resolution of this pattern.

Inbreeding results:

Inbreeding did not show any clear relationship with cancer susceptibility. This is reflected in both a comparison of distributions between the cancer and control groups (Figure 3.3) as well as in a quadratic logistic regression comparing the two groups (Figure 3.4). However, a single factor ANOVA comparing mortality classes did show a credible difference for both the 'bacteria' and 'parasite' groups. Belugas that died of bacterial infection had an overall higher level of inbreeding, whereas belugas that died from parasite loads had an overall lower level of inbreeding, compared to other mortality groups (Figure 3.5; Figure 3.6). There was no association found between level of inbreeding and age (Figure 3.7), which in turn may suggest that juvenile mortality is not influenced by increased inbreeding. When each mortality class was analyzed separately, a pattern between age and inbreeding was still not seen.

Comparing 10 Arctic beluga samples to 139 samples from the SLB showed that the SLB population is more inbred than Arctic belugas (Figure 3.8). Comparing 10 Arctic samples to 10 randomly selected samples from the St. Lawrence population also showed the same pattern, and suggested that the variation in inbreeding (based on HL and IR values) is similar for both groups (Figure 3.8).

Discussion:

Genes and cancer susceptibility:

The results presented here suggest that variation along the p53 gene is not influencing cancer susceptibility in the SLB population. There were only 3 sites of variation found at this locus, with two occurring along intron 7 (presumably neutral) and the third resulting in a silent mutation at exon 8 (also presumably neutral). None of these SNPs were linked to the presence of cancer, even when considering the haplotype of all three SNP combinations (Table 3.1). It may be hypothesized that the entire SLB population is fixed at an allele (at this gene) that in turn may make the entire population more susceptible to cancer than other beluga populations. However, the p53 sequences of all examined individuals from the St. Lawrence population were the same as the cDNA of two arctic belugas (which also included the same polymorphism in exon 8) (GenBank Accession Number AF475081, Xu et al., 2002), suggesting that the SLB population is not fixed for a particularly susceptible allele at this gene. Not finding an association between p53 gene polymorphisms and cancer susceptibility is an important finding because this gene is known for playing a crucial role in cancer development for various species (e.g. Hainaut et al., 1997). Polymorphisms along this gene have previously been hypothesized to influence the high incidence of tumorigenesis in the SLB (Xu et al., 2002), but our findings suggest otherwise. Future work studying the etiology of cancer can now focus on identifying other candidate genes and pathways that may be influencing cancer development for this population.

Microsatellite association analysis along 7 candidate genes also did not show any credible difference between cancer and control groups. This was true when looking at the effects of individual alleles (Table 3.2; Table 3.3; Table 3.4) as well as combined effects across

loci (Figure 3.2). However, this lack of association may be due to reasons other than there being no true effect. Given that our sample size was small (comprising of approximately 30 individuals for both cancer and control group), this may mean that a smaller effect size will go undetected by our analysis. For example, power analyses showed that with our sample size we would be not be able to detect a credible difference unless the difference in the frequency of an influential allele was at least 10 between the cancer and control groups. Therefore, increasing the sample size may give us a better indication as to whether certain alleles are linked to the presence or absence of cancer. Although our microsatellite association analysis does not allow us to state anything with a high degree of confidence, it does provide a strong starting point. For example, microsatellite alleles at the FOXO3 (Table 3.2), HPSE2 (Table 3.3), and BRCA2 (Table 3.4) locus show some general patterns of association with cancer presence/absence. DAPC analysis also suggests that there may be a combined effect of alleles on cancer susceptibility based on these candidate genes (Figure 3.2). We will continue to analyze these regions as more samples from necropsied individuals become available in the coming years. This in turn, will give us a more conclusive understanding of whether these candidate genes are influencing cancer susceptibility for this population.

Overall, there are three main hypotheses that can explain the variation we see in cancer susceptibility within the SLB population: (1) some individuals carry cancer susceptible allele(s) (2) the entire population is fixed at cancer susceptible allele(s) or (3) cancer development is driven predominantly by environmental factors. Because there are hundreds of genes and pathways that can be influencing cancer development, we cannot

say conclusively whether genetic variation along genes are influencing cancer development based on analysis on a few candidate loci. However, it can be argued that if genetic factors were playing a role in differentially influencing cancer development, those with susceptible genes may develop cancer at a younger age. Since belugas who die from cancer in the St. Lawrence population are dying at an old age (average age of death being 48), this may indicate a strong role of environmental factors. It is possible that cancer development in this population is caused by chance mutations accumulated over time through continued contaminant exposure (hypothesis #3). For example, high levels of Benzo(a)pyrene (BaP) exposure are currently proposed as an important factor in cancer development in the SLB. The most common tumors found in this population are gastrointestinal epithelial cell tumors, and this has been hypothesized to be caused by feeding on BaP contaminated invertebrates, which then concentrates carcinogens in the gastrointestinal area (Martineay et al., 2002). This hypothesis is supported by laboratory studies done on mice that show that the route of BaP intake determines where the tumor develops (Culp et al., 1998). However, is it important to note that even if hypothesis #1 or #2 above hold true, this does not exclude the negative effects of environmental contaminants. On the contrary, individuals carrying inherited cancer susceptible genes are more likely to feel the impacts of carcinogens in their environment, since fewer environmental mutations may be needed to drive cancer development in these individuals. Therefore, regardless of the relative impact of environmental vs. genetic factors, it is imperative to reduce the contaminant load in the St. Lawrence estuary. Conservation initiatives should focus on reducing the carcinogenic contaminants within

the St. Lawrence river to lower cancer rates and prevent future outbreaks in this population.

Inbreeding:

From a conservation genetics perspective, a small population size may make the SLB genetically vulnerable due to increased inbreeding, which can subsequently reduce an individual's overall fitness, as well as lower their ability to adapt to novel changes (Amos and Balmford, 2001). Compared to other Canadian beluga populations, past studies have shown that the SLB has reduced genetic diversity, suggesting that inbreeding may influence the genetic characteristics of this population (Patenaude et al., 1994; Murray et al., 1999; de March and Postma, 2003). Increased inbreeding has previously been hypothesized as a factor limiting recovery (COSEWIC, 2004). However, our study is the first to look into how inbreeding may be influencing some of the major mortality patterns documented in this population. With respect to cancer susceptibility, our results showed no clear relationship with the level of inbreeding (Figure 3.3; Figure 3.4). Although past studies on other marine mammal populations (California sea lion) have shown that inbred individuals are more prone to cancer development (Acevedo-Whitehouse et al., 2003), the same effect was not found for the SLB. There is also not a clear link between inbreeding and dystocia, inbreeding and neonatal birth (Figure 3.5) or between inbreeding and age (Figure 3.7). However, comparing mortality categories (Figure 3.5) did show that individuals who died from bacterial infection were more inbred (Figure

3.6), and in contrast, those that died from parasite infection were less inbred (Figure 3.6), compared to other groups.

Our results suggest that increased inbreeding may lead to decreased bacterial resistance in the SLB. Theoretically, it is widely accepted that inbreeding can lower pathogen resistance, and that exposure to infectious agents can exacerbate the negative fitness consequences of inbreeding (e.g., Coltman et al., 1999; Keller and Waller, 2002). Since inbred individual are expected to be more homozygous, they may experience a bacteria-mediated reduction in fitness if: (1) they are unable to recognize a wide breadth of pathogens as more heterozygous individuals, or (2) if pathogens are part of an environment that selects against individuals expressing deleterious recessive alleles. Although several studies have shown an association between inbreeding and pathogen resistance across various species (e.g. Cassinello et al., 2001; Acevedo-Whitehouse et al., 2003; Valsecchi et al., 2004; Whiteman et al., 2006), with a few focusing specifically on bacterial pathogens (e.g. Ilmonen et al., 2008), our results appear to be one of the first to suggest a role between inbreeding and bacterial resistance in a cetacean population.

Bacterial infection is currently a major cause of mortality for the SLB population. For example, data from 1983 to 2006 showed that bacterial infection accounted for 18% of mortalities (DFO, 2012). Therefore, understanding the relative impacts of inbreeding on bacterial resistance is needed to assess how this relationship may influence population trends and recovery rates. Future research should look into specific bacterial loads in necropsied individuals to further disentangle the relationship between level of inbreeding

and bacterial resistance. The different types of bacterial pathogens should also be identified, and their relative impacts on survival quantified. As our sample size comprised only 14 individuals for the 'bacteria' group, more data should also be analyzed as more necropsy data becomes available in the coming years. This will hopefully give us a better understanding of this relationship, which is turn can help us better assess how inbreeding and bacterial infections may be negatively influencing this population.

An interesting pattern found when looking at the effects of inbreeding in the SLB is that parasite infection seems to be associated with a lower level of inbreeding when compared to other individuals (Figure 3.5; Figure 3.6). The large negative impacts that inbreeding can have on general fitness (e.g. Saccheri et al., 1998) combined with past research highlighting the positive relationship between homozygosity and parasite susceptibility (e.g. Coltman et al., 1999; Cassinello et al., 2001; Acevedo-Whitehouse et al., 2003) lead me to hypothesize that if a trend between parasite infection and inbreeding was found, it would be in the opposite direction of our results. However, a few other studies have also shown that in some populations, homozygous individuals can have higher parasite resistance than those with higher heterozygosity (e.g. Rikjs et al., 2008; Giese and Hedrick, 2003). For example, an experimental study on the endangered Gila topminnow (*Poecilliopsis occidentalis*) showed that inbred fish had lower mortality rates than more outbred fish when exposed to a novel pathogen (Gise and Hedrick, 2003). One hypothesis that may explain the underlying mechanism behind this trend is the presence of specific beneficial alleles. Because inbred individuals are more likely to be homozygous at specific beneficial alleles, this may give them an advantage over individuals that are more

heterozygous. Since parasite infection accounts for 20% of mortalities in this population (based on data from 175 necropsied individuals examined from 1983 to 2006, DFO, 2012), future work should look into why reduced genetic diversity may be harmful with respect to parasite-induced mortality. It is interesting that we saw no effect of inbreeding on cancer susceptibility, and opposing trends for bacterial and parasite infection. Taken together, these findings highlight that inbreeding can have varied consequences with respect to different diseases in the same population.

Another aspect of our inbreeding analysis involved comparing the St. Lawrence population to belugas from the Arctic. Past studies have shown that the SLB have lower genetic diversity than Arctic beluga populations (Patenaude et al., 1994; Murray et al., 1999; de March and Postma, 2003). Therefore, one of the hypotheses as to why we may not have been able to detect an effect of inbreeding on cancer susceptibility (or other mortality patterns) is because the population as a whole may be inbred and/or have low variation between individuals. However, comparing 10 Arctic belugas to 139 belugas from the St. Lawrence population showed that although the mean level of inbreeding is higher for the SLB, the entire population is not inbred as there is significant variation between individuals (Figure 3.8). Comparing 10 Arctic belugas to 10 randomly selected individuals from the St. Lawrence population also showed similar levels of variation between individuals from each group (Figure 3.8). Therefore, not finding an association between inbreeding and some of the mortality classes (e.g. cancer) cannot be attributed to the population being mostly inbred and/or having low variability.

Overall, our inbreeding analyses can guide conservation initiatives by helping identify how inbreeding is influencing some of the mortality patterns seen in this population. Although we were not able to detect an effect of inbreeding on cancer susceptibility, we did notice that individuals dying from bacterial and parasite infection have relatively high and low levels of inbreeding, respectively. Future research should add to these findings by addressing questions relating to the underlying mechanisms causing these patterns (e.g. why is homozogosity related to lower parasite-induced mortality?), as well as asking more detailed questions about these relationships (e.g. how do different types of bacterial infection influence mortality rates in this population?). By continuing to test key hypotheses, and determining (or ruling out) limiting factors, we can gain a more comprehensive understanding of the factors limiting recovery, and take the necessary steps to mitigate some of the threats that may be facing this population.

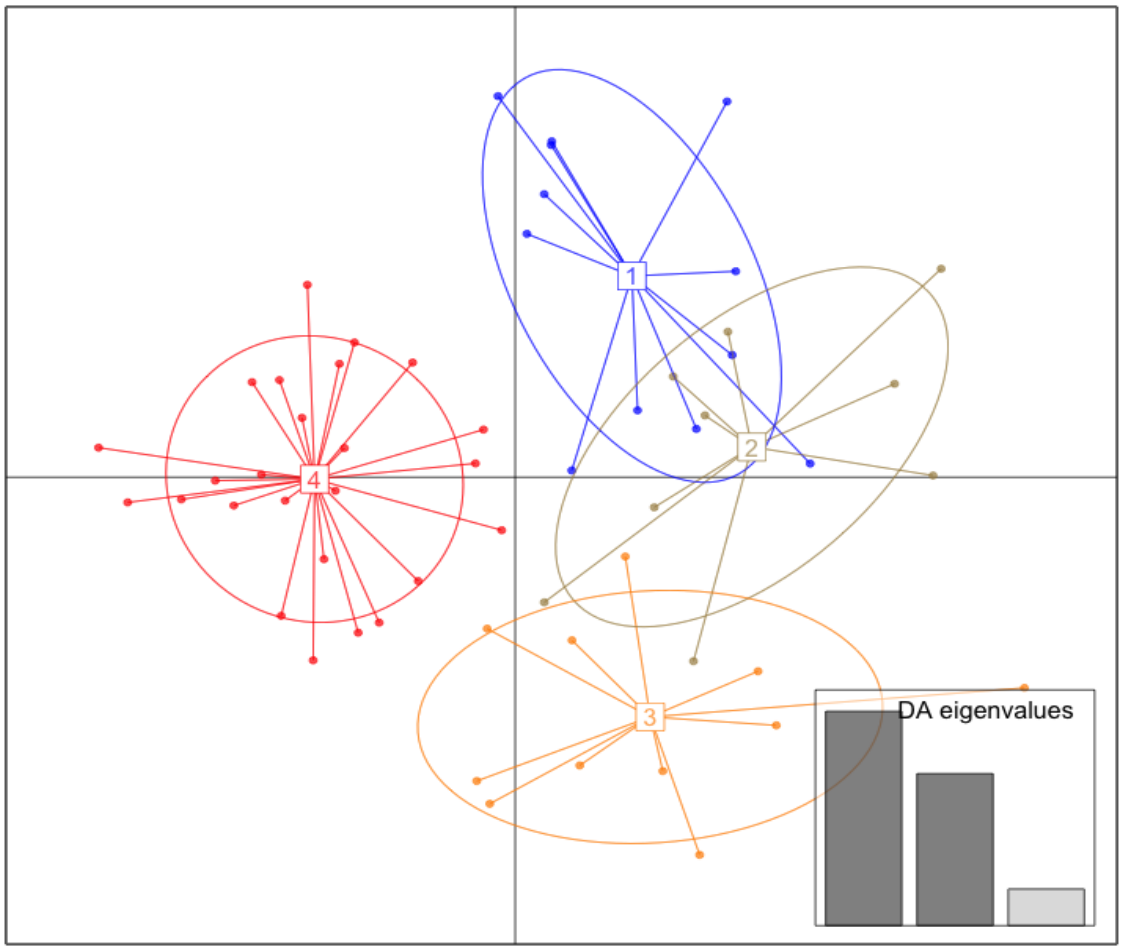

Figure 3.1. Results from DAPC analysis for 9 polymorphic cancer-linked loci from the St. Lawrence beluga, showing four genetic clusters. Individuals are represented by dots and the clusters as inertia ellipses. Eigenvalues (which shows the number of discriminant functions to retain, which in this case is 3) of the analysis are displayed in inset.

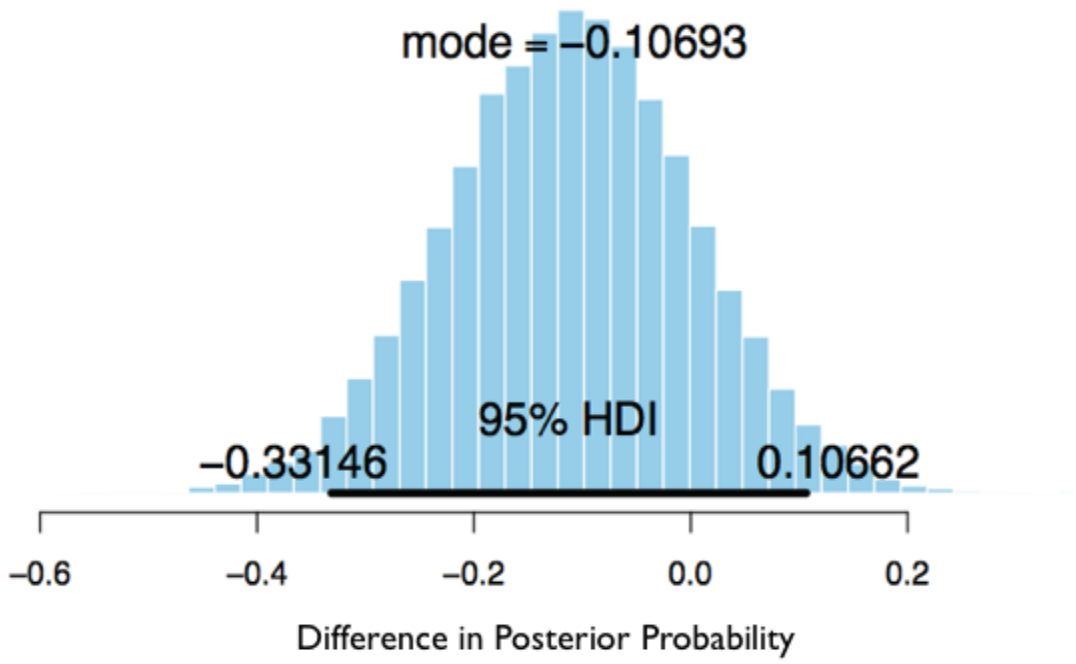

Figure 3.2. Posterior probability distribution from logistic regression showing the difference between being in the same group (cancer or control) vs. being in a different group (cancer or control) for pairs that were in the same cluster (based on DAPC analysis of cancer-linked microsatellites). A 95% HDI interval that does not encompass 0 would show a credible difference. However, although the results are not credible, the probability distribution is not centered around 0, suggesting an effect of cancer genotype on cluster assignment.

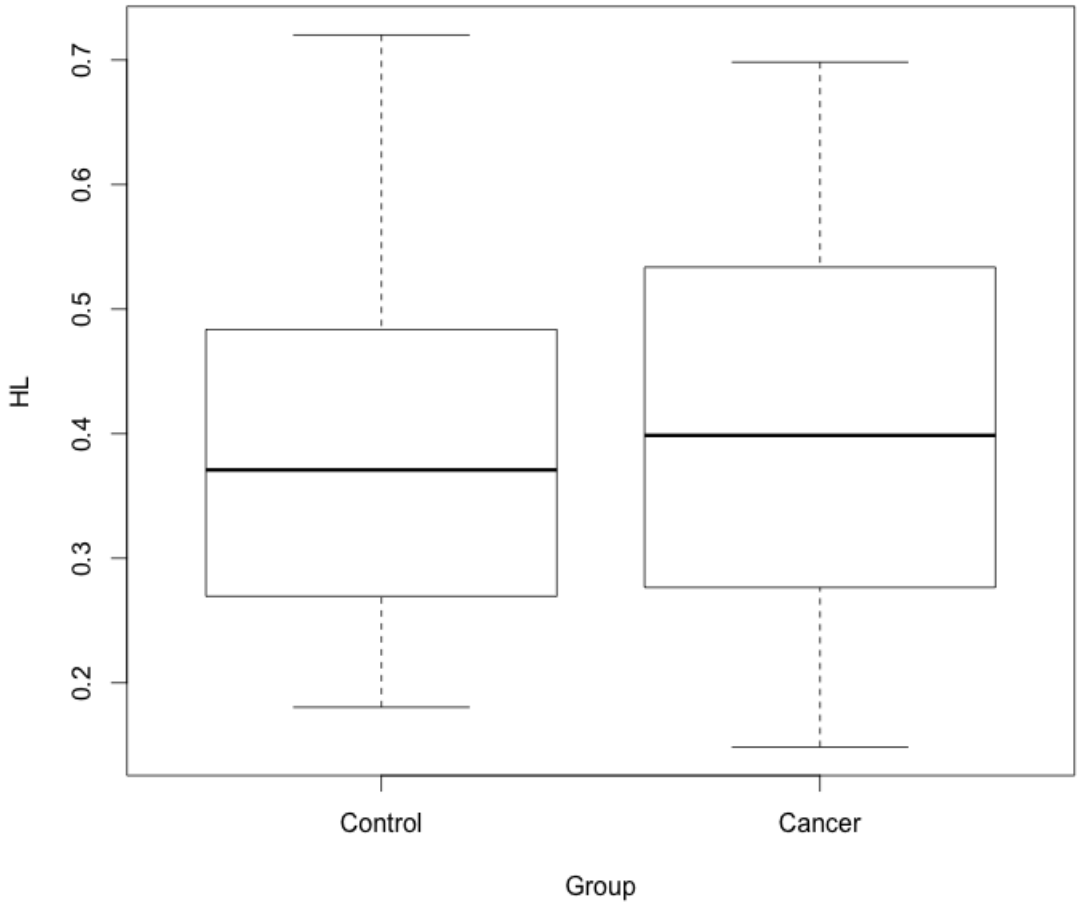

Figure 3.3. Boxplot showing the inbreeding coefficient (HL values at neutral loci) of cancer and control group. There was no credible difference between the means of the two groups.

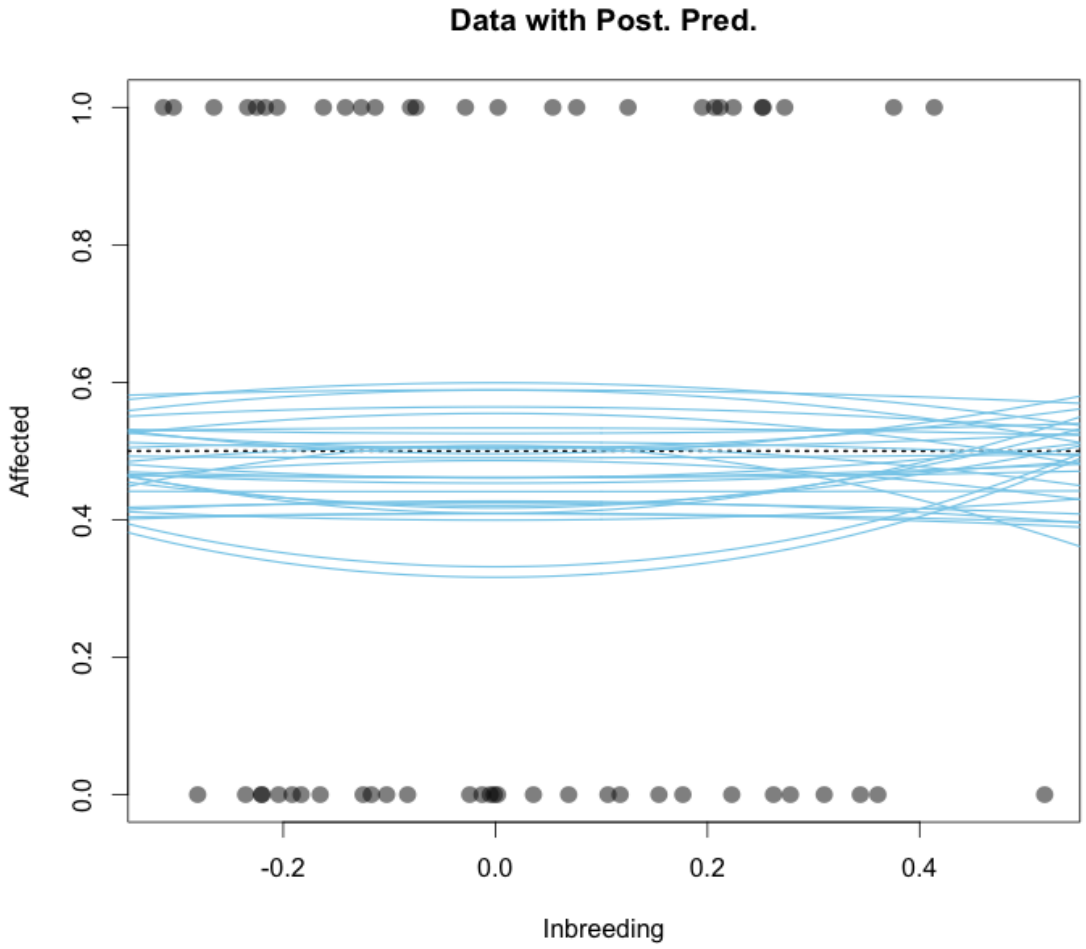

Figure 3.4. Regression lines from the posterior probability depicting the relationship between inbreeding and cancer presence/absence. Inbreeding coefficient is represented by HL values (of neutral loci), cancer is represented by 1 and the control group is represented by 0. There are no clear patterns found when comparing between the two groups.

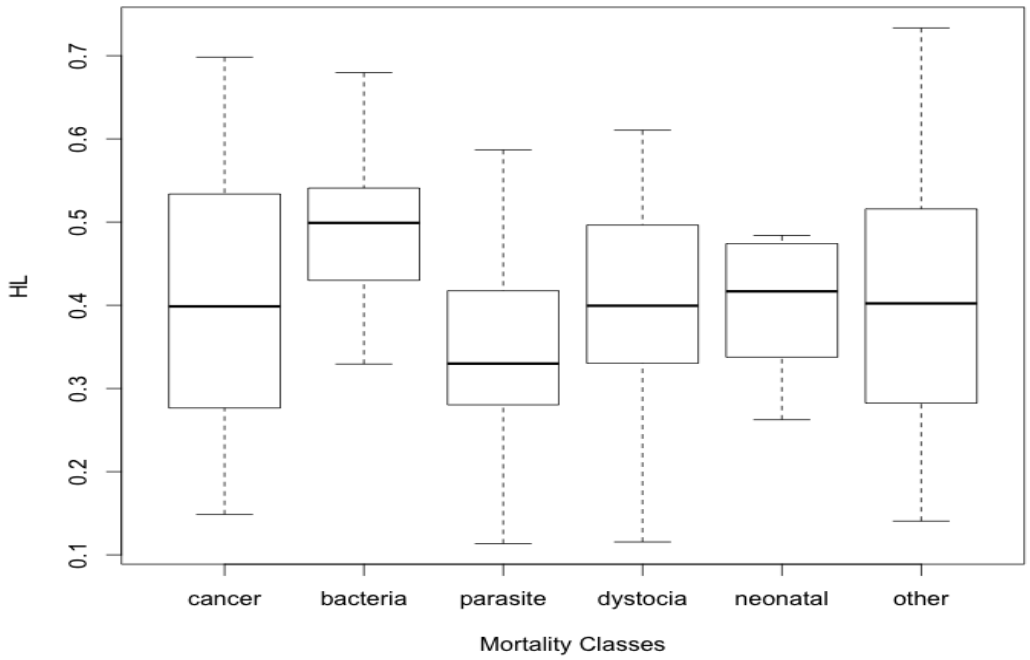

Figure 3.5. Boxplots depicting the level of inbreeding (HL at neutral loci) found in different mortality classes. 'Other' represents individuals that died from the following causes: degenerative changes, fishing gear, undetermined, starvation, and trauma. A single factor ANOVA showed that the 'bacteria' group is more inbred, while the 'parasite' group is less inbred than the remaining individuals.

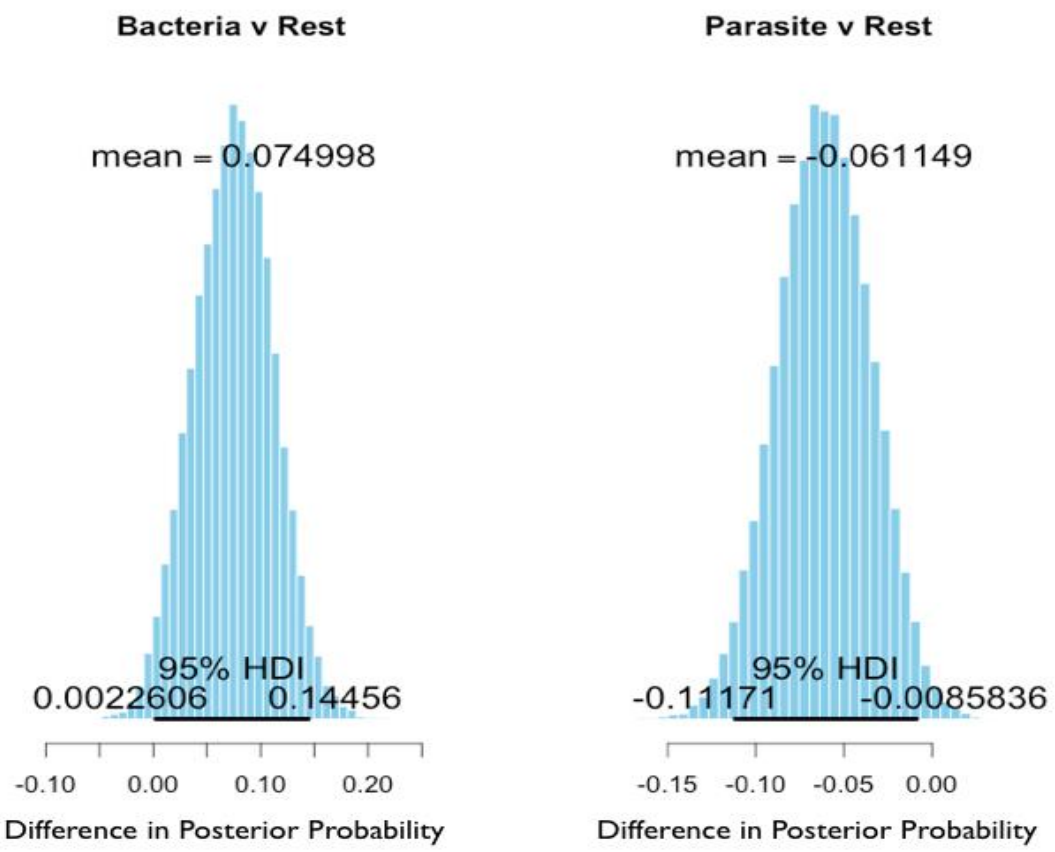

Figure 3.6. A single factor ANOVA comparing the inbreeding level (HL at neutral loci) of different mortality classes. Individuals that died from bacterial infection had a higher level of inbreeding when compared to individuals from all other mortality classes. One the other hand, individuals who died from parasite infection had a lower level of inbreeding when compared to those that died from all other mortality classes. This figure shows the difference between posteriors when comparing bacteria and parasite groups with individuals from all remaining mortality classes. Analysis of IR values at neutral loci also showed the same patterns.

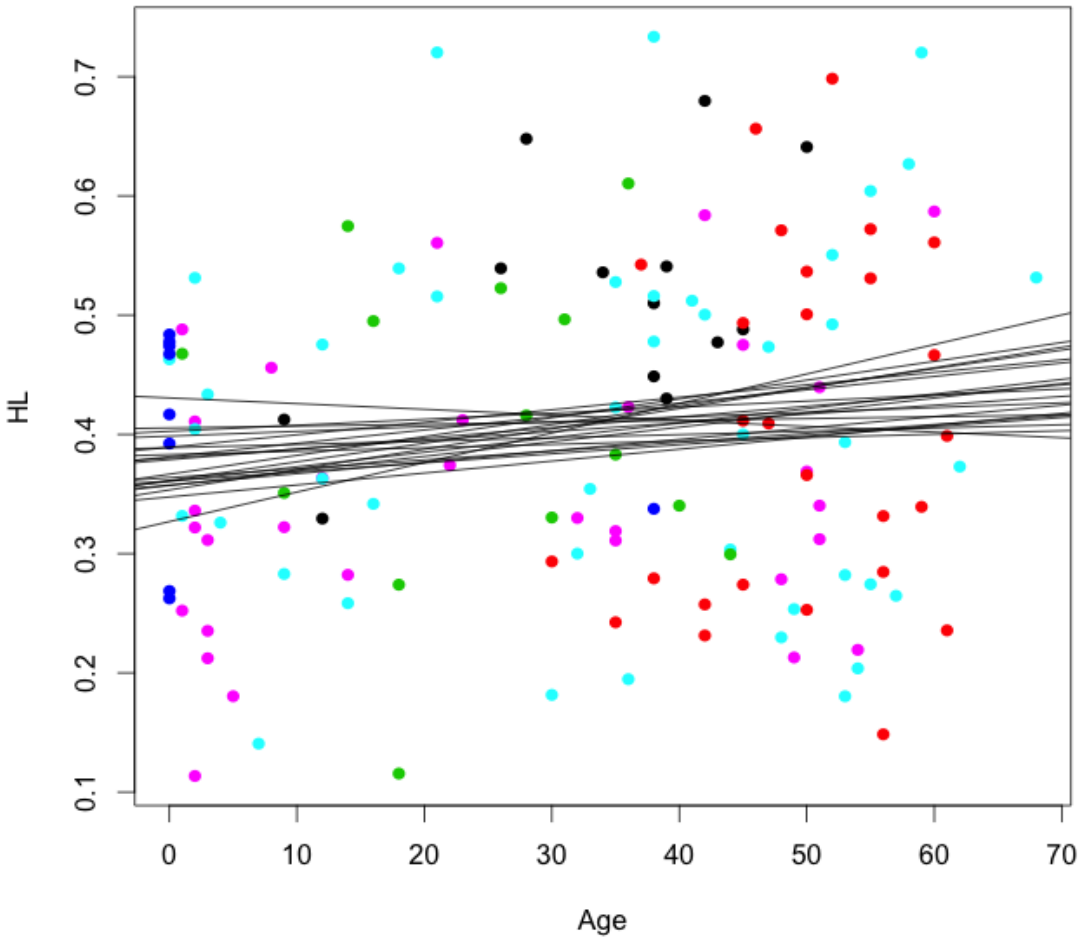

Figure 3.7. Plot of inbreeding coefficient (HL at neutral loci) regressed on age, with 20 regression lines from the posterior probability. Although a slight trend is seen where the level of inbreeding increases with age, the posterior probability of the slope was centered around 0 (mode: 0.0009; 95% HDI confidence interval: -0.00026-0.0020) showing that the age of death does not seem to be influenced by the level of inbreeding. Further, when separated by mortality class, there still was no credible association detected with respect to HL and age. Different colours correspond to different mortality classes: red = cancer, black = bacteria, purple = parasite, green = dystocia, dark blue = neonatal and light blue = other.

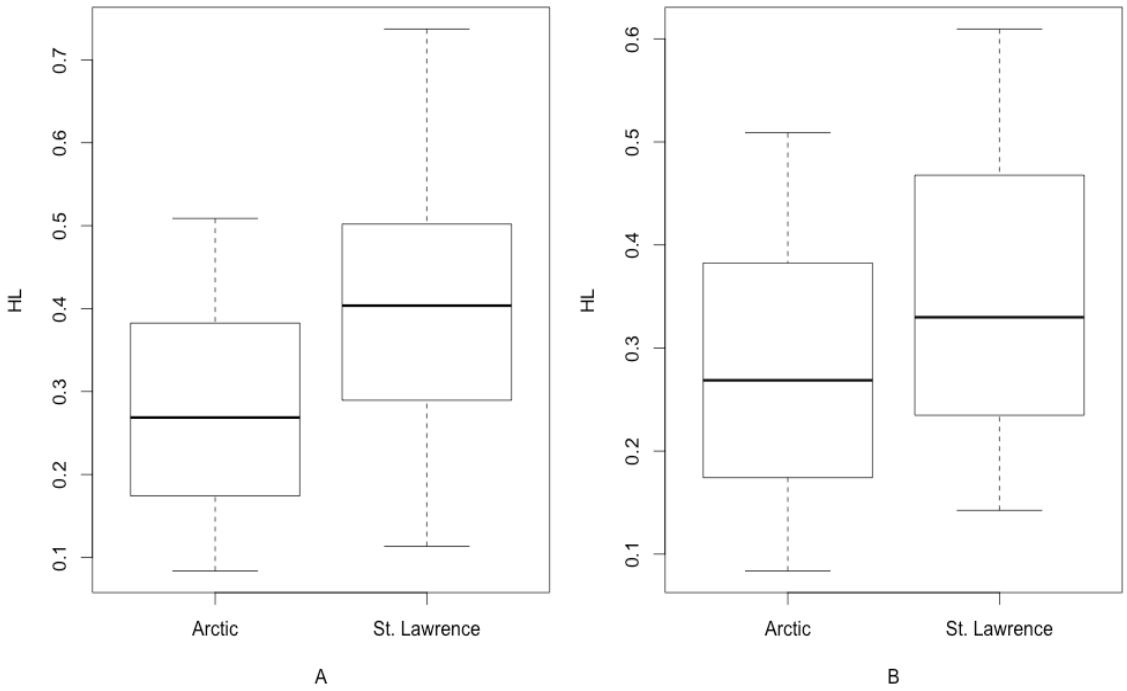

Figure 3.8. (A) Boxplot of inbreeding coefficient (HL at neutral loci) comparing 10 Arctic belugas to 131 belugas from the St. Lawrence population. Comparing the difference between the means of the two groups showed a credible difference, with the St. Lawrence population being more inbred. (B) Boxplot of inbreeding coefficient (HL at neutral loci) comparing 10 Arctic belugas to 10 randomly selected samples from the St. Lawrence population. This comparison was done 10 different times (each time comparing the same 10 Arctic samples to 10 randomly chosen samples from the SLB population). All comparisons showed the same general trend, where the variation in inbreeding for the two populations were similar, but the overall inbreeding coefficient was higher for the St. Lawrence population.

Table 3.1. Frequencies of p53 alleles and haplotypes that fall under cancer and control groups. The probability distribution is the mode (and associated 95% HDI confidence interval) from the posterior probability representing the likelihood of each allele or haplotype to be associated with the cancer group.

Frequency	Group	Allele	Posterior Probability Distribution
32	Cancer	C (site 127)	0.0002 (-0.172, 0.172)
37	Control	C (site 127)	
16	Cancer	T (site 127)	-0.0002 (-0.184, 0.172)
19	Control	T (site 127)	
16	Cancer	C (site 202)	0.0006 (-0.185, 0.172)
19	Control	C (site 202)	
32	Cancer	T (site 202)	-0.0006 (-0.171, 0.185)
37	Control	T (site 202)	
9	Cancer	A (site 355)	0.007 (-0.115, 0.408)
6	Control	A (site 355)	
39	Cancer	G (site 355)	-0.007 (-0.408, 0.114)
50	Control	G (site 355)	
28	Cancer	C-T-G (haplotype)	-0.001 (-0.429, 0.127)
29	Control	C-T-G (haplotype)	
6	Cancer	C-T-A (haplotype)	0.004 (-0.190, 0.782)
1	Control	C-T-A (haplotype)	
13	Cancer	T-C-G (haplotype)	-0.0001 (-0.470, 0.160
14	Control	T-C-G (haplotype)	
9	Cancer	T-C-A (haplotype)	0.001 (-0.313, 0.356)

Frequency	Group	Allele	Posterior Probability Distribution
6	Control	T-C-A (haplotype)	

Table 3.2. Frequency of alleles at the FOXO3 locus (Dle-foxo3a) in cancer and control groups. The probability distribution shows the likelihood (mode and 95% HDI confidence interval) of each allele to be associated with cancer is also shown (based on results from the contingency table analysis).

Frequency	Group	Allele	Posterior Probability Distribution
7	Cancer	273	-0.014 (-0.792, 0.215)
12	Control	273	
5	**Cancer**	**277**	**0.035 (-0.173, 1.997)**
0	**Control**	**277**	
3	**Cancer**	**279**	**-0.01 (-1.123, 0.130)**
9	**Control**	**279**	
0	Cancer	281	0.002 (-1.655, 0.656)
1	Control	281	
34	Cancer	283	0.000 (-0.458. 0.353)
35	Control	283	
3	Cancer	285	0.014 (-0.416, 1.18)
1	Control	285	

Table 3.3. Frequency of alleles at the HPSE2 locus (Dle-hpse2b) in cancer and control groups. The probability distribution shows the likelihood (mode and 95% HDI confidence interval) of each allele to be associated with cancer is also shown (based on results from the contingency table analysis).

Frequency	Group	Allele	Posterior Probability Distribution
50	Cancer	272	0.006 (-0.541, 0.614)
49	Control	272	
2	**Cancer**	**276**	**-0.027 (-1.355, 0.186)**
8	**Control**	**276**	
1	Cancer	282	0.0353 (-0.651, 1.970)
0	Control	282	
1	Cancer	284	-0.001 (-0.935, 0.895)
1	Control	284	

Table 3.4. Frequency of alleles at the BRCA2 locus (Dle-brca2b) in cancer and control groups. The probability distribution shows the likelihood (mode and 95% HDI confidence interval) of each allele to be associated with cancer is also shown (based on results from the contingency table analysis).

Frequency	Group	Allele	Posterior Probability Distribution
35	**Cancer**	**303**	**-0.037 (-1.107, 0.230)**
47	**Control**	**303**	
16	**Cancer**	**305**	**0.008 (-0.736, 0.652)**
10	**Control**	**305**	
0	Cancer	343	-0.018 (-2.608, 1.645)
1	Control	343	
2	Cancer	339	-0.002 (-0.564, 2.304)
0	Control	339	
1	Cancer	329	0.018 (-0.863, 2.125)
0	Control	329	

Appendix 3.1. Prior probabilities chosen for different models used in this project.

Prior probabilities chosen for a simple linear regression. The equation representing this model is: $y = b0 + b1x$. Metric data were first standardized (centered around 0 with a SD of 1) to avoid issues with autocorrelation.

Parameter	Priors	Explanation
Y	comes from a t distribution with a mean represented by (b0 + b1x), tau of $1/\sigma^2$, and nu of nu1 + 1	A t-distribution is useful when data are mostly normally distributed, but have a few outliers (accommodated by the heavy tails)
	tau comes from a folded t distribution with a mean of 0, tau of 0.001, and nu of 2	A folded t-distribution with these values allows tau to have a higher likelihood of taking smaller values (lower precision in my estimate)
	nu1 has a rate of 0.5	If I assume a t-distribution that is more likely to have a heavier tail, a smaller nu value is more likely. Therefore, a rate of 0.5 made sense as this starts at 0 and exponentially decreases. However, since nu has to take the value of at least 1, I added 1 to this value
b0	b0 is normally distributed with a mean of 0 and tau of $1/\sigma^2$	If I assume no initial relationship between the predicted and predictor variable, then b0 should be representative of the y data. Since I standardized all metric data prior to analysis, the mean of the y data (and therefore b0) should be centered around 0

Parameter	Priors	Explanation
	tau comes from a folded t distribution with a mean of 0, tau of 0.001, and nu of 2	A folded t-distribution with these values allows tau to have a higher likelihood of taking smaller values (lower precision in my estimate)
b1	b1 is normally distributed with a mean of 0 and tau of $1/\sigma^2$	If I assume no relationship between the predicted variable and b1, then b1 should be centered around 0
	tau comes from a folded t distribution with a mean of 0, tau of 0.001, and nu of 2	A folded t-distribution with these values allows tau to have a higher likelihood of taking smaller values (lower precision in my estimate)

Prior probabilities chosen for comparing two means (metric predicted variable on two groups). The two groups being compared is represented by y1 and y2.

Parameter	Priors	Explanation
y1	normally distributed with a mean (mu1) that is normally distributed and tau coming from a gamma distribution	Data sets that I compared were normally distributed
	mu1 is is normally distributed with a mean of the combined data set and a tau of 1/ (1000 x σ^2)	If I assume no difference between the two groups, then mu1 could be normally distributed and centered around the mean value of the combined y data set. Choosing a SD 1000x higher then that of the combined dataset ensured a loose estimate of SD.

Parameter	Priors	Explanation
	tau1 is a gamma distribution with an α of 0.001 and a β of 0.001	These values result in a higher likelihood of tau taking on lower values (=lower precision)
y2	normally distributed with a mean (mu1) that is normally distributed and tau coming from a gamma distribution	Data sets that I compared were normally distributed
	mu1 is is normally distributed with a mean of the combined data set and a tau of 1/ (1000 x σ^2)	If I assume no difference between the two groups, then mu1 could be normally distributed and centered around the mean value of the combined y data set. Choosing a SD 1000x higher then that of the combined dataset ensured a loose estimate of SD.
	tau1 is a gamma distribution with an α of 0.001 and a β of 0.001	These values result in a higher likelihood of tau taking on lower values (=lower precision)

Prior probabilities chosen for a contingency table analysis. The equation for this model can be represented by $\lambda = \exp(b0 + br + bc + brc + ...)$.

Parameter	Priors	Explanation
Y	comes from a poisson distribution with a λ distribution represented by the equation	When modeling count data, it is appropriate to use the poisson distribution which allows for positive integers and the parameter lambda

Parameter	Priors	Explanation
b0	normally distributed with a mean of yLogMean (log mean of the y data) and a tau of $1/\sigma^2$ (where σ^2 is represented by 2x the yLogSD)	If I assume no relationship between predicted and predictor variables, then b0 should represent the log mean of the y data. I obtained a loose estimate of SD by taking the log of the SD and multiplying by 2.
predictor variables	normally distributed with a mean of 0 and a tau of $1/\sigma^2$	If I assume no relationship between predicted and predictor variables, then all predictor variables should be centered around 0.
	tau comes from a folded t distribution with a mean of 0, tau of 0.001, and nu of 2	A folded t-distribution with these values allows tau to have a higher likelihood of taking smaller values (lower precision in my estimate)

Priors chosen for a single factor ANOVA. The equation for this model can be represented by y = b0 + Bjx, where b0 is the baseline value for the population, and Bj is the degree of deflection above or below this baseline that is associated with being a member of each group. Metric data was standardized prior to analysis in order to avoid problems with autocorrelation.

Parameter	Priors	Explanation
Y	comes from a t distribution with a mean represented by (b0 +∑bjx), nu of (nu1+1), and a tau of of $1/\sigma^2$	A t-distribution is useful when data are mostly normally distributed, but have a few outliers (accommodated by the heavy tails)

Parameter	Priors	Explanation
	nu1 has a rate of 0.5	If I assume a t-distribution that is more likely to have a heavier tail, a smaller nu value is more likely. Therefore, a rate of 0.5 made sense as this starts at 0 and exponentially decreases. However, since nu has to take the value of at least 1, I added 1 to this value
	σ^2 is from a uniform distribution ranging from 0 to 10	Since data was standardized a value from 0 to 10 should capture the SD.
b0	normally distributed with a mean of 0 and a tau of $1/\sigma^2$ (where σ^2 is 10x the SD of the y data)	If I assume no relationship between the predictor and predicted variables, then b0 should be representative of the mean of the y data. Since the y data is standardized and centered around 0, I chose a normal distribution with a mean of 0 to represent b0.
Bjx	normally distributed with a mean of 0 and a tau of $1/\sigma^2$	If I assume no relationship between the predictor and predicted variables, then bjx should be centered around 0.
	tau comes from a folded t distribution with a mean of 0, tau of 0.001, and nu of 2	A folded t-distribution with these values allows tau to have a higher likelihood of taking smaller values (lower precision in my estimate)

Priors for a logistic regression. A logistic regression is represented by the following equation: $y = \text{logistic}(b0 + \sum bjxj)$ where y is represented by either 0 or 1.

Parameter	Priors	Explanation
Y	comes from a Bernoulli distribution taking a value of 1 or 0	A Bernoulli distribution is used when the y value can take on a dichotomous value of either 1 or 0
b0	normally distributed with a mean of 0 and a tau of $1/\sigma^2$ (where σ^2 takes a value of 2)	
$\sum bjxj$	normally distributed with a mean of 0 and a tau of $1/\sigma^2$	If I assume no relationship between predictor and predicted variables, then predictor variables should be centered around 0.
	tau is from a folded t distribution with a mean of 0, tau of 0,001, and nu of 2	A folded t-distribution with these values allows tau to have a higher likelihood of taking smaller values (lower precision in my estimate)

Appendix 3.2. The reaction conditions for cancer-linked microsatellite markers. Included are the reaction number, locus name, locus colour, primer concentration, and reference.

Reaction #	Locus	Colour	Annealing Temperature (°C)	Concentration	Reference
1	Dle-p16A	Black (NED)	56	0.08uM	Chapter 2
1	Dle-ptenA	Red (PET)	56	0.3uM	Chapter 2
1	Dle-ptenB	Black (NED)	56	0.3uM	Chapter 2
1	Dle-brca2A	Green (VIC)	56	0.16uM	Chapter 2
2	Dle-p16B	Black (NED)	60	0.055uM	Chapter 2

Reaction #	Locus	Colour	Annealing Temperature (°C)	Concentration	Reference
2	Dle-brca1A	Black (NED)	60	0.19uM	Chapter 2
2	Dle-rb1B	Blue (6FAM)	60	0.22uM	Chapter 2
2	Dle-foxo3A	Red (PET)	60	0.16uM	Chapter 2
3	Dle-hpse2A	Green (VIC)	50	0.2uM	Chapter 2
3	Dle-hpse2B	Black (NED)	50	0.3uM	Chapter 2
3	Dle-brca2B	Blue (6FAM)	50	0.2uM	Chapter 2
3	Dle-foxo3B	Red (PET)	50	0.2uM	Chapter 2
4	Dle-rb1A	Green (VIC)	50	0.22uM	Chapter 2

Appendix 3.3. The reaction conditions for neutral microsatellite markers. Included are the reaction number, locus name, locus colour, primer concentration, and reference.

Reaction #	Locus	Colour	Annealing Temperature (°C)	Concentration	Reference
1	Ev37Mn	Green (VIC)	54	0.3uM	Valsecchi and Amos (1996)
1	FCB17	Red (PET)	54	0.7uM	Buchanan et al. (1996)
1	FCB5	Yellow (NED)	54	0.45uM	Buchanan et al. (1996)
1	FCB10	Blue (6FAM)	54	0.65uM	Buchanan et al. (1996)
2	FCB4	Red (PET)	58	0.25uM	Buchanan et al. (1996)
3	Ev14Pm	Blue (6FAM)	58	0.3uM	Valsecchi and Amos (1996)
3	SW19	Red (PET)	52	0.4uM	Richard et al. (1996)
3	RW48	Green (VIC)	52	0.35uM	Waldick et al. (1999)
4	Ev94Mn	Blue (6FAM)	56	0.5uM	Valsecchi and Amos (1996)
4	FCB6	Yellow (NED)	56	1.2uM	Buchanan et al. (1996)
4	FCB1	Green (VIC)	56	0.45uM	Buchanan et al. (1996)
5	TexVet5	Yellow (NED)	52	0.15uM	Rooney et al. (1999)

Reaction #	Locus	Colour	Annealing Temperature (°C)	Concentration	Reference
5	MK6	Green (VIC)	52	0.15uM	Krutzen et al. (2001)
5	TexVet19	Blue (6FAM)	52	0.3uM	Rooney et al. (1999)
6	FCB14	Green (VIC)	55	0.2uM	Buchanan et al. (1996)
6	RW34	Blue (6FAM)	55	0.09uM	Waldick et al. (1999)

Literature Cited:

Acevedo-Whitehouse, K., F. Gulland, D. Greig, and W. Amos. 2003. Inbreeding: Disease susceptibility in California sea lions. Nature 422: 35.

Amos, W. and A. Balmford. 2001. When does conservation genetics matter? Heredity 87: 257-265.

Amos, W., J. W. Wilmer, K. Fullard, T. M. Burg, J. P. Croxall, D. Bloch and T. Coulson. 2001. The influence of parental relatedness on reproductive success. Proceedings of the Royal Society B: Biological Sciences 268(1480): 2021–2027.

Aparicio, J. M., J. Ortego and P. J. Cordero, 2006. What should we weigh to estimate heterozygosity, alleles or loci? Molecular Ecology 15(14): 4659–4665.

Buchanan, F. C., M. K. Friesen, R. P. Littlejohn and J. W. Clayton, 1996. Microsatellites from the beluga whale *Delphinapterus leucas*. Molecular Ecology 5(4): 571–575.

Cassinello, J., M. Gomendio, and E. Roldan. 2001. Relationship between coefficient of inbreeding and parasite burden in endangered gazelles. Conservation Biology 15: 1171–1174.

Coltman, D., J. Pilkington, J. Smith, J.M. Pemberton. 1999. Parasite-mediated selection against inbred soay sheep in a free-living, island population. Evolution 53(4): 1259-1267.

COSEWIC. 2004. COSEWIC assessment and update status report on the beluga whale *Delphinapterus leucas* in Canada. Committee on the Status of Endangered Wildlife in Canada. Ottawa. 70p.

Coulon, A., 2010. genhet: an easy-to-use R function to estimate individual heterozygosity. Molecular Ecology Resources 10(1): 167–169.

Culp, S., D. Gaylor, W. Sheldon, L. Goldstein and F. Beland. 1998. A comparison of the tumors induced by coal tar and benzo[a]pyrene in a 2-year bioassay. Carcinogenesis 19: 117-124.

Davis, C. S. and C. Strobeck, 1998. Isolation, variability, and cross-species amplification of polymorphic microsatellite loci in the family Mustelidae. Molecular Ecology 7(12): 1776– 1778.

De Guise, S., A. Lagace and P. Beland. 1994. Tumors in St. Lawrence beluga whale (*Delphinapterus leucas*). Vet Pathology 31: 444-449.

de March, B. G. E. and L. D. Postma. 2003. Molecular genetic stock discrimination of belugas (*Delphinapterus leucas*) hunted in eastern Hudson bay, northern Québec, Hudson strait, and Sanikiluaq (Belcher island), Canada, and comparisons to adjacent populations. Arctic. 56:111-124.

DFO. 2012. Recovery strategy for the beluga whales (*Delphinapterus leucas*) St. Lawrence estuary population in Canada. Species at Risk Act Recovery Strategy Series. Fisheries and Oceans Canada, Ottawa.

Driscoll, E. E., J. I. Hoffman, L. E. Green, G. F. Medley and W. Amos. 2011. A Preliminary Study of Genetic Factors That Influence Susceptibility to Bovine Tuberculosis in the British Cattle Herd. PLoS ONE 6(4): e18806–9.

Ellison, A. M. 1996. An introduction to Bayesian inference for ecological research and environmental decision-making. Ecological Applications 6: 1036-1046.

Forstmeier, W., H. Schielzeth, J. C. Mueller, H. Ellegren and B. Kempenaers, 2012. Heterozygosity-fitness correlations in zebra finches: microsatellite markers can be better than their reputation. Molecular Ecology 21(13): 3237–3249.

Giese A.R. and P.W. Hedrick. 2003. Genetic variation and resistance to a bacterial infection in the endangered Gila topminnow. Animal Conservation 6: 369–377.

Halsey, L. G., D. Curran-Everett, S. L. Vowler and G. B. Drummond, 2015. The fickle P value generates irreproducible results. Nature Methods 12(3): 179–185.

Ilmonen P., D.J. Penn, K. Damjanovich, J. Clarke, D. Lamborn, L. Morrison, L. Ghotbi and W.K. Potts. 2008. Experimental infection magnifies inbreeding depression in house mice. Journal of Evolutionary Biology 21: 834–841.

Johnson, D. 1999. The insignificance of statistical significance testing. The Journal of Wildlife Management 63(3): 763–772.

Jombart T., S. Devillard and F. Balloux. 2010 Discriminant analysis of principal components: a new method for the analysis of genetically. BMC Genet. 11: 94.

Irwin D.L., K.R. Mitchelson, and I. Findlay. 2003. PCR product cleanup methods for capillary electrophoresis. BioTechniques 34: 932-934.

Kruschke, J. K., 2013. Bayesian estimation supersedes the t test. Journal of Experimental Psychology 142(2): 573–603.

Kru¨tzen, M., E. Valsecchi, R. C. Connor and W. B. Sherwin. 2001. Characterization of microsatellite loci in Tursiops aduncus. Molecular Ecology Notes 1(3): 170–172.

Lesage, V. and M. C. S. Kingsley. 1998. Updated status of the St. Lawrence river population of the beluga, *Delphinapterus leucas*. Canadian Field Naturalist. 112: 98-114.

Ludwig, D. 1996. Uncertainty and the assessment of extinction probabilities. Ecological Applications 6(4): 1067-1076.

Martineau, D., Lagace, A., Beland, P., Higgins, D., Armstrong, L., Shugart, R. 1988. Pathology of stranded beluga whales (*Delphinapterus leucas*) from the St. Lawrence estuary, Quebec, Canada. Journal of Comparitive Pathology 98: 287-311.

Martineau, D., K. Lemberger, A. Dallaire, P. Labelle, T. Lipscomb, P. Michel, and I. Mikaelian. 2002. Cancer in wildlife, a case study: Beluga from the St. Lawrence estuary, Quebec, Canada. Environmental Health Perspectives 110:285-292.

McAloose, D. and A.L. Newton. 2009. Wildlife cancer: a conservation perspective. Nature Reviews Cancer 9:517-526.

Mosnier, A., T. Doniol-Valcroze, J. F. Gosselin, V. Lesage, L. N. Measures and M. O. Ham- mill, 2015. Insights into processes of population decline using an integrated population model: The case of the St. Lawrence Estuary beluga (*Delphinapterus leucas*). Ecological Modelling 314:15–31.

Murray, B. W., R. Michaud and B. N. White. 1999. Allelic and haplotype variation of major histocompatibility complex class II DRB1 and DQB loci in St. Lawrence beluga, (*Delphinapterus leucas*). Molecular Ecology. 8:1127-1139.

Newman, S., Smith, S. 2006. Marine mammal neoplasia: a review. Vet pathology 43: 865-880.

Patenaude, N., J. Quinn, P. Beland, M. Kingsley and B. White. 1994. Genetic variation of the St. Lawrence beluga whale population assessed by DNA fingerprinting. Molecular Ecology 3: 375-381.

Plummer, M., 2003. JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. Proceedings of the 3rd International Workshop on Distributed Statistical Computing March 20–22:Vienna, Austria.

R Core Team, 2015. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Richard, K. R., H. Whitehead and J. M. Wright, 1996. Polymorphic microsatellites from sperm whales and their use in the genetic identification of individuals from naturally sloughed pieces of skin. Molecular Ecology 5(2): 313–315.

Rijks, J., J. Hoffman, T. Kuiken, A. Osterhaus and W. Amos. 2008. Heterozygosity and lungworm burden in seals (*Phoca vitulina*). Heridity 100: 587-593.

Rooney, A. P., D. B. Merritt and J. N. Derr, 1999. Microsatellite diversity in captive bottlenose dolphins (*Tursiops truncatus*). The Journal of heredity 90(1):228–231.

Sambrook, J. and D. W. Russell. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, July 2001.

Sanger F., S. Nicklen and A.R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. PNAS 74: 5463–5467.

Schlo¨tterer, C., B. Amos and D. Tautz, 1991. Conservation of polymorphic simple sequence loci in cetacean species. Nature 354(6348):63–65.

Slate, J., P. David, K.G. Dodds, B.A. Veenvliet, B.C. Glass, T.E. Broad. and J.C. McEwan. 2004. Understanding the relationship between the inbreeding coefficient and multilocus heterozygosity: theoretical expectations and empirical data. Heredity 93: 255-265.

Spielman D., B.W. Brook, D.A. Briscoe and R. Frankham. 2004. Does inbreeding and loss of genetic diversity decrease disease resistance? Conservation Genetics 5: 439–448.

Szulkin, M., N. Bierne and P. David, 2010. Heterozygosity-fitness correlations: a time for reappraisal. Evolution 64(5): 1202–1217.

Valsecchi, E. and W. Amos, 1996. Microsatellite markers for the study of cetacean populations. Molecular Ecology 5(1): 151–156.

Valsecchi, E., W. Amos, J. Raga, M. Podesta and W. Sherwin. 2004. The effects of inbreeding on mortality during a morbillivirus outbreak in the Mediterranean striped dolphin (*Stenella coeruleoalba*). Animal Conservation 7: 139–146.

Wade, P. R., 2000. Bayesian methods in conservation biology. Conservation Biology 14(5): 1308–1316.

Waldick, R. C., M. W. Brown and B. N. White, 1999. Characterization and isolation of microsatellite loci from the endangered North Atlantic right whale. Molecular Ecology 8(10): 1763–1765.

Wang, J. Y., T. R. Frasier, S. C. Yang and B. N. White. 2008. Detecting recent speciation events: the case of the finless porpoise (genus *Neophocaena*). Heredity 101: 145-155.

Whibley, C., P. Pharoah and M. Hollstein. p53 polymorphisms: cancer implications. 2009. Nature Reviews 9: 95-107.

Whiteman, N.K., K.D. Matson, J.L. Bollmer and P.G. Parker. 2006. Disease ecology in the Galapagos Hawk (*Buteo galapagoensis*): host genetic diversity, parasite load and natural antibodies. Proceedings of the Royal Society of London B: Biological Science 273: 797–804.

Xu, N., T. Shiraki, T.Yamada, M. Nakajima, J. Gauthier, C. Pfeiffer and S. Sato. 2002. Nucleotide sequence of the p53 cDNA of beluga whale (*Delphinapterus leucas*). Gene 288: 159-166.

Chapter 4:

Conclusion

Main Findings:

The overall purpose of this project was to try to understand if genetic factors are influencing cancer susceptibility and other mortality patterns in the St. Lawrence Beluga (SLB) population. As previously mentioned, the SLB is known for having one of the highest noted cancer rates in any reported wildlife species (e.g. Martineau et al., 2002; DFO, 2012). Past research has attributed the high prevalence of benign and malignant tumours to the contaminants found within their immediate environment (De Guise et al., 1994; Martineau et al., 2002; Newman and Smith, 2006). However, relatively little research has focused on understanding the genetic etiology of cancer development in this population.

By comparing the DNA between cancer and control groups in the SLB population our study found that:

1. Genetic variation along the p53 locus does not seem to influence cancer susceptibility;
2. Microsatellite linkage analysis on 7 candidate genes (p16, RB1, BRCA1, BRCA2, PTEN, HPSE2, FOXO3) did not detect any strong associations between specific microsatellite alleles and cancer susceptibility. However, there were a few associations identified that warrant further research. The combined effect of microsatellite alleles across loci was also suggestive (though not credible) of an effect on cancer susceptibility;
3. Degree of inbreeding does not seem to be associated with cancer susceptibility

In addition to examining the effect of degree of inbreeding on cancer development, I also assessed the effect of inbreeding on other mortality patterns. The results showed:

4. Bacteria-induced mortality is associated with a higher level of inbreeding, while parasite-induced mortality is associated with a lower level of inbreeding, relative to other mortality categories.

Taken together, I did not find strong associations between genetic variation along key cancer genes or any effect of inbreeding with respect to cancer susceptibility. However, I did detect that inbreeding may be influencing bacteria and parasite resistance in this population.

Problems Encountered:

Sample Size:

One of the major problems with this study is that my sample sizes were small for the cancer and control groups. Skin samples from necropsied individuals were provided by the Group for Research and Education on Marine Mammals (GREMM), who have been conducting long-term research on the St. Lawrence beluga population for the last 25 years. In total, I obtained skin samples from 139 individuals collected from 1983 to 2014. However, only 30 samples were from those that died of cancer. Furthermore, because our

control group was composed of individuals that died of other causes and were also over the age of 45 (in order to assume that they would not have died of cancer had they lived longer), this resulted in approximately 30 individuals for the control group as well. A small sample size is not ideal as it lowers the statistical power of our analysis, thereby reducing the probability of detecting a true effect.

The issues surrounding a small sample size became most apparent for our microsatellite association analysis. Although we did not detect any credible difference between cancer and control groups, this may be because a smaller effect size may not be detectable with our current sample size. To determine what effect size is required given our sample size and statistical analysis, I conducted a power analysis on generated data sets (an example of which is shown in Table 4.1.). This showed that with our sample size we would be not be able to detect a credible difference unless the difference in the frequency of an influential allele was at least 10 between the cancer and control groups.

Unfortunately, since the etiology of cancer is often multifactorial, and can involve a complex interplay between environmental, genetic, and epigenetic factors, it is likely that if there is a genetic influence, the effect size of this influence will likely not be large for any particular locus, but the effect may be influenced across many loci. Therefore, even though our analyses did not show any credible differences between cancer and control groups, this could be for one of three reasons: (1) there is no difference between cancer and control groups, (2) the effect size is too small to detect with our current sample size, or (3) the sample size is too small to capture the true nature and variation in each group.

To mitigate this problem, we will continue to analyze necropsy data as more become available in the upcoming years. For example, we noticed a few interesting patterns with our microsatellite linkage analysis. Specifically, although we detected no credible differences between cancer and control groups, certain alleles at the FOXO3, HPSE2, and BRCA2 locus were more or less associated with the cancer group than would be expected by chance (see Chapter 3). These loci show enough promise that we will continue these analyses as samples and necropsy data from more individuals become available. If we do see a credible difference with increased sample size, our next step will be to sequence the gene of interest (much like we did with the p53 locus) to see if these specific microsatellite alleles are linked to polymorphisms within the coding region of the gene. A larger sample size will also give us more resolution for our p53 and haplotype association analysis.

Implications of this project:

Molecular markers for the study of cancer:

Primers developed for the sequencing of the p53 gene, as well as for microsatellite association analysis for 7 key cancer related genes (p16, RB1, BRCA1, BRCA2, PTEN, HPSE2, FOXO3) may be used to study the genetic etiology of cancer across several mammalian species. Microsatellite primers were shown to work across both baleen and toothed whales, and approximately half amplified in members of Ungulata (cow and

horse). Thus, they may prove useful to other studies on the genetic basis of cancer in other cetacean and ungulate populations. Cancer has recently emerged as a conservation threat for many wildlife populations, and a high prevalence of benign and malignant tumours has already been noted in several species. For example, in one study, benign genital papillomas were present in 66.7% of dusky dolphins and 48.5% of Burmeister's porpoises, and were considered important enough to interfere with copulation in 10% of Burmeister's porpoises (Van Bressem et al., 1996). A high incidence of cancer has previously been noted in a cattle population as well (Jarrett et al. 1978). Increased monitoring and research efforts may detect the presence of cancer in subsequent cetacean and ungulate species, and our primers may be used as a strong starting point for elucidating the role genetic factors may be playing on influencing cancer development in specific populations.

Impact of Inbreeding on the SLB:

This project looked at the impact increased inbreeding may have on some of the major mortality patterns seen in the SLB population. Low levels of genetic variability and/or inbreeding have previously been hypothesized as factors limiting recovery in the SLB (COSEWIC, 2004), however, previous studies have not directly tested for the effects of inbreeding on individual or population health. Our results are the first to show that inbreeding may be influencing disease resistance in this population. While the level of inbreeding did not seem to influence cancer susceptibility, individuals that died from bacterial infection seem to be be relatively inbred, while individuals that died from

parasite infection seem relatively less inbred than the remaining population (see Chapter 3). These relationships are important and warrant further investigation since bacterial and parasitic pathogens are the two of the leading causes of death in this population, accounting for 18%, and 20% of mortalities (based on data from 175 necropsied individuals examined from 1983 to 2006, DFO, 2012).

The relationship seen between increased homozygosity and bacteria-induced mortality suggests that the current level of inbreeding may have negative consequences for the population. Since the population is already small (most recent estimate of 889 individuals [Mosnier et al., 2015]), the likelihood of inbreeding and inbreeding depression is higher than would be expected for a larger population. Therefore, it is imperative to fully understand the impact increased inbreeding may be having on bacterial resistance. Currently, our data set is limited as we only have 14 samples from necropsied individuals who died from bacterial infection, and no information on the species of bacteria affecting individuals, or the bacterial load found in infected individuals. Future research can add to our current understanding of this relationship by: (1) analyzing more necropsied individuals who died from bacterial infection, (2) identifying the different types of bacterial pathogens present in the population, and (3) comparing varying bacterial pathogen loads with the level of inbreeding. Successfully quantifying the effect inbreeding may have on bacterial resistance is required to fully understand impacts on individual health, as well as to come up with effective conservation strategies for the population as a whole.

Finding a negative association between homozygosity and parasite-induced mortality may also have conservation implications for the SLB. While increased inbreeding has been linked to decreased pathogen resistance in several wildlife populations (e.g. Cassinello et al., 2001; Acevedo-Whitehouse et al., 2003; Valsecchi et al., 2004; Whiteman et al., 2006), very few studies have shown the reverse trend (Rijks et al., 2008; Giese and Henrick, 2003). Our study is the first to show that relatively inbred individuals may have an advantage with respect to parasite resistance in this population, presumably due to the presence of homozygous advantageous alleles. However, because bacterial-induced mortality shows the opposite trend, this may mean that the impacts of inbreeding on disease susceptibility and individual health may be more complex than assumed. Future work should aim to identify the underlying mechanism driving the negative relationship between inbreeding and parasite susceptibility. For example, one could try to identify potential alleles within the genome that may influence increased parasite resistance in their homozygous state. A broader understanding of this pattern will not only help determine how inbreeding may be differentially influencing disease susceptibility in the SLB, but can also improve theory on why genome-wide homozygosity may sometimes be advantageous with respect to pathogen resistance.

Future Direction:

One of the overarching themes of this project was to assess if, and to what extent, inbreeding was influencing mortality patterns in the St. Lawrence beluga population. Although we looked at the effects of inbreeding on cancer susceptibility, pathogen

resistance, dystocia, neonatal birth, and juvenile mortality, inbreeding can also impact many other fitness associated traits. A complementary project to our study could be to look into the impact increased inbreeding may have on reproductive performance in the St. Lawrence beluga. Our collaborators at GREMM have been collecting long-term field data from female belugas, noting how many calves they birth through-out their lifetime. We could use these data to assess whether increased inbreeding is resulting in reduced reproductive performance. Many studies have found that genetic impacts may be more readily found in association with reproductive performance than mortality (e.g., Chapman et al., 2009), and a few studies on wildlife populations have shown an effect of inbreeding on reproductive success (e.g. Wildt et al., 1982). Tackling this question can help us gain a broader understanding of how genetic factors, in particular, inbreeding, may be influencing population and reproductive dynamics in the St. Lawrence beluga.

Table 4.1. Hypothetical microsatellite locus with 4 different alleles, and their associated frequencies under cancer and control groups. Contingency table analysis of this data suggests that in order to detect a credible difference between cancer and control group with a small data set (n=28 individuals for each group), you would need a difference in the order of 14-0.

Frequency	Group	Allele	Result
14	Cancer	100	credible difference
0	Control	100	
7	Cancer	200	no credible difference
7	Control	200	
7	Cancer	300	no credible difference
7	Control	300	
7	Cancer	400	no credible difference
7	Control	400	

Literature Cited:

Acevedo-Whitehouse, K. F. Gulland, D. Greig, and W. Amos. 2003. Inbreeding: Disease susceptibility in California sea lions. Nature 422: 35.

Cassinello, J., M. Gomendio, and E. Roldan. 2001. Relationship between coefficient of inbreeding and parasite burden in endangered gazelles. Conservation Biology 15: 1171–1174.

Chapman, J. R., S. Nakagawa, D. W. Coltman, J. Slate, and B. C. Sheldon. 2009. A quantitative review of heterozygosity-fitness correlations in animal populations. Molecular Ecology 18: 2746-2765.

COSEWIC. 2004. COSEWIC assessment and update status report on the beluga whale *Delphinapterus leucas* in Canada. Committee on the Status of Endangered Wildlife in Canada. Ottawa. 70p.

De Guise, S., Lagace, A., Beland, P. 1994. Tumors in St. Lawrence beluga whale (Delphinapterus leucas). Vet Pathology 31: 444-449.

DFO. 2012. Recovery strategy for the beluga whales (Delphinapterus leucas) St. Lawrence estuary population in Canada. Species at Risk Act Recovery Strategy Series. Fisheries and Oceans Canada, Ottawa. 88 pp + X pp.

Giese A.R. and P.W. Hedrick. 2003. Genetic variation and resistance to a bacterial infection in the endangered Gila topminnow. Animal Conservation 6: 369–377.

Jarrett, W. F. H., P.E. Mcneil, W.T.R. Grimshaw, I.E. Selman, and W.I.M. Mcintyre. 1978. High incidence area of cattle cancer with a possible interaction between an environmental carcinogen and a papilloma virus. Nature 274: 215-217.

Martineau, D., Lemberger, K., Dallaire, A., Labelle, P., Lipscomb, T., Michel, P., Mikaelian, I. 2002. Cancer in wildlife, a case study: Beluga from the St. Lawrence estuary, Quebec, Canada. Environmental Health Perspectives.

Mosnier, A., T. Doniol-Valcroze, J. F. Gosselin, V. Lesage, L. N. Measures and M. O. Ham- mill, 2015. Insights into processes of population decline using an integrated population model: The case of the St. Lawrence Estuary beluga (Delphinapterus leucas). Ecological Modelling 314: 15–31.

Newman, S., Smith, S. 2006. Marine mammal neoplasia: a review. Vet pathol 43: 865-880.

Rijks, J., J. Hoffman, T. Kuiken, A. Osterhaus and W. Amos. 2008. Heterozygosity and lungworm burden in seals (Phoca vitulina). Heridity 100: 587-593.

Valsecchi, E., W. Amos, J. Raga, M. Podesta and W. Sherwin. 2004. The effects of inbreeding on mortality during a morbillivirus outbreak in the Mediterranean striped dolphin (Stenella coeruleoalba). Animal Conservation 7: 139–146.

Van Bressem, M. F., Waerebeek, K., Pierard, G., Desaintes, C. 1996. Genital and lingual warts in small cetaceans from coastal Peru. Disease of Aquatic Organisms 26: 1-10.

Wenwei, H., Zhaohui, F., Teresky, A., Levine, A. 2007. p53 regulates maternal reproduction through LIF. Nature Letter 450: 721-725.

Whiteman, N.K., K.D. Matson, J.L. Bollmer and P.G. Parker. 2006. Disease ecology in the Galapagos Hawk (Buteo galapagoensis): host genetic diversity, parasite load and natural antibodies. Proc. R. Soc. Lond. B Biol. Sci. 273: 797–804.

Wildt, D.E., Baas, E.J., Chakraborty, P.K., Wolfle, T.L., Stewart, A.P. 1982. Influence of inbreeding on reproductive performance, ejaculate quality and testicular volume in the dog. Theriogenology 17(4): 445-452.