Understanding Genetic Limitations to Population Recovery in Endangered Marine Mammals

By
Carla Anne Crossman

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

August 2024, Halifax, Nova Scotia

Approved by: Dr. Timothy Frasier
Supervisor
Professor, Saint Mary's University

Approved by: Dr. Zhongmin Dong
Committee Member
Professor, Saint Mary's University

Approved by: Dr. Daniel Ruzzante
Committee Member
Professor, Dalhousie University

Approved by: Dr. Michael McGowen
External Examiner
Research Scientist & Curator of Marine Mammals, Smithsonian National Museum of Natural History

Date: August 7, 2024

Understanding Genetic Limitations to Population Recovery in

Endangered Marine Mammals

by Carla Anne Crossman

Abstract

In small populations, genetic factors can both inhibit their recovery and provide insight into their past. Genomic tools can therefore provide a way to better understand limitations to population recovery in endangered species. In my dissertation I present four clear examples of how genomic tools can directly address questions outlined in recovery planning documents for endangered marine mammals in Canada.

I used methylation patterns at genomic positions that have been previously found to respond differently to stressors to demonstrate that methylation patterns may be a tangible means to quantify cumulative effects of stress in wildlife populations using resident killer whales (*Orcinus orca*) as a case study. I used whole genome sequencing to demonstrate that North Atlantic right whales (*Eubalaena glacialis*) have lived for thousands of years with smaller effective population sizes than southern right whales (*E. australis*) from the Southwest Atlantic. These results suggest that basing recovery goals in the North Atlantic on the successful recovery seen in the Southern Oceans may not be appropriate. I used ddRAD sequencing to show that in North Atlantic right whales, a female's inbreeding coefficient did not correlate with her reproductive fecundity, however all individuals had higher heterozygosity than expected suggesting inbreeding may be leading to increased fetal loss in the population – potentially limiting the growth of the population as a whole. Finally, I designed a GT-Seq panel to improve ongoing genetic monitoring of North Atlantic right whales and demonstrated its increased power to resolve familial relationships in the species.

Genomic tools can provide invaluable insight to directly address knowledge gaps outlined in recovery planning documents in Canada. While these provide just a few examples for a pair of endangered species, the utility of genomics to inform conservation should be better acknowledged and incorporated in recovery planning for Canada's Species at Risk.

August 7, 2024

Table of Contents

List of Tables

List of Figures

Acknowledgements

First and foremost, I would like to thank my supervisor Dr. Timothy Fraiser for his incredible support, guidance and assistance over the past six years. I also extend thanks to my current and former lab mates in the Frasier lab who provided assistance, feedback and support. I also wish to thank my committee members Dr. Zhongmin Dong and Dr. Daniel Ruzzante for their advice and guidance over the course of my degree.

There have been many others without whom this work would not be possible in particular Philip Hamilton and those who contributed data and samples to the North Atlantic Right Whale Consortium. I thank Dr. Michael Fontaine for hosting me in his lab during my internship and helping me to take on my first whole genome dataset. I also thank the members of his lab in 2022, especially Josquin Daron for welcoming me and providing invaluable tips and tricks.

Thank you to my friends (many of whom have gone through this process before) for listening and empathizing with my struggles and always providing supporting advice or well needed distractions.

I thank my husband, Dusan, and my family for their endless patience and understanding for the past six years. Finally, I cannot be more thankful for my son Max and his soon-to-be brother who have been my biggest motivation to succeed and do my best to have my work ultimately affect conservation efforts, while simultaneously teaching me how, in the bigger picture, this work is completely insignificant.

Preface

Chapters of this thesis have been published (or accepted for publication) in peer-reviewed journals as follows:

Chapter 2:

Crossman, C.A., Barrett-Lennard, L.G. and Frasier, T.R., 2021. An example of DNA methylation as a means to quantify stress in wildlife using killer whales. *Scientific Reports*, 11(1), p.16822. doi.org/10.1038/s41598-021-96255-1

Chapter 3:

Crossman, C.A., Fontaine, M.C. and Frasier, T.R., 2023. A comparison of genomic diversity and demographic history of the North Atlantic and Southwest Atlantic southern right whales. *Molecular Ecology,* 00:1-17. doi.org/10.1111/mec.17099

Chapter 4:

Crossman, C.A., Hamilton, P.K., Brown, M.W., Conger, L.A., George, R.C., Jackson, K.A., Radvan, S.N., Frasier, T.R., 2024. Effects of inbreeding on reproductive success in endangered North Atlantic right whales. *Royal Society Open Science*, 11:240490. doi.org/10.1098/rsos.240490

1 Introduction

1.1 The impact of non-lethal threats to wildlife populations

Effective conservation requires far more than eradicating lethal threats. By focusing on lethal threats, or the ultimate causes of death, there is a much needed, immediate payoff – fewer individuals die, slowing a population decline, or perhaps even aiding in recovery. However, in this process, we often overlook contributing factors or other ultimate causes of mortality facing these populations. Many anthropogenic activities that affect wildlife are not lethal themselves but can lead to negative fitness consequences that could interfere with recovery. For example, exposure to endocrine disrupting chemicals has been found to decrease fertility, delay sexual maturity or decrease offspring viability in many different taxa (Marlatt et al. 2022). Many cetaceans (whales, dolphins, and porpoises) use acoustics to locate prey and/or to communicate over long distances with other individuals – including for reproduction. Underwater noise can mask these calls and therefore have a number of different consequences in both the short or long term for individuals and populations including reducing mating opportunities (Weilgart 2007). Reproductive success is correlated with food availability in songbirds (Grames et al. 2023) and therefore changes in prey abundance could

impact their recovery or population growth. These non-lethal threats could greatly affect future population growth and should therefore be of great concern for small populations.

Having a small population size is in itself an additional non-lethal stressor on a population. Small populations can inherently retain less genetic diversity and suffer from increased effects of genetic drift. Genetic drift represents one of the primary mechanisms through which genetic diversity is lost over time, and its effects are amplified in small populations (Wright, 1931). In small populations, there are fewer options for mates leading to higher rates of inbreeding. Inbreeding can lead to severe fitness consequences for the offspring (Charlesworth and Charlesworth 1987), furthering the potential for decreased population sizes. In very small populations, especially those that face ongoing anthropogenic mortalities, the ongoing risks associated with small population size, can further contribute to population decline leading into a positive feedback loop known as the extinction vortex (Gilpin and Soulé 1986; Caughley 1994).

Unfortunately, for many species, it is rare to be affected by a single anthropogenic stressor. The flora and fauna that inhabit the Galápagos Islands are simultaneously facing impacts of marine pollution, fishing pressure and threats

associated with climate change (reviewed in Alava et al. 2023). Belugas from the St. Lawrence estuary population are threatened by vessel noise and disturbance, and contaminant exposure among other threats (reviewed in Lesage 2021). The cumulative effects of stressors on individuals can be difficult to measure and/or assess. Their combined effects can be categorized as additive, antagonistic or synergistic. All three of these combined effects were identified in a comprehensive assessment of cumulative effects in an algal turf community (Fong et al. 2018). Most concerningly are synergistic effects where the cumulative effects of two stressors are greater than the sum of their parts. Synergistic effects have been found in honey bees (*Apis mellifera*) where the effects of pathogens and pesticides or parasites act synergistically to negatively impact the health and survival of bees (Di Prisco et al. 2016; Grassl et al. 2018). Synergistic interactions between climate change, habitat loss and harvesting threaten the persistence of marten (*Martes americana*) and lynx (*Lynx canadensis*) in Southeastern Canada and the Northeastern United States (Carroll 2007). Effective conservation efforts will require viewing extinction through the lens of synergistic processes (Brook et al. 2008).

Understanding the cumulative impact of multiple non-lethal threats is important for conservation, but these can be difficult to measure empirically. Increasingly complex models are being developed to estimate cumulative effects including population

consequence of disturbance (PCoD) or population consequences of multiple stressors (PCoMS) models (Pirotta et al. 2018, 2022). A recent study applied a PCoMS framework to 50 years of North Atlantic right whale (*Eubalaena glacialis*) life history data and described how threats like blunt vessel strikes and serious entanglement are likely to have the greatest health implications on individual health, but prey limitations can also have prolonged effects (Pirotta et al. 2023).

Like all models, PCoD and PCoMS models have limitations as they rely on many assumptions and therefore may not accurately describe the realized impacts felt by individuals. Individual heterogeneity in responses to stressors is often not considered in these models (Keen et al. 2021. Supp Mat). Data on physiological responses to disturbance are harder to obtain, requiring samples such as blood, feces or tissues from living animals, and are often not included in model framework (Pirotta et al. 2018). New methods using genomic, epigenomic and/or transcriptomic data are being developed to describe realized impacts of stressors to individuals, and offer a promising outlook on our ability to truly understand the impact of non-lethal stressors to individuals.

1.2 A role for genomics in applied conservation

Since the 1970s, biologists have recognized the importance of conserving genetic diversity (Frankel, 1974). Genetic diversity acts as a reservoir of standing variation that may be required for species to adapt to changing environments. Without this diversity, populations may not be as adaptable to new environments, including exposure to new or modified diseases (Hilbish & Koehn, 1987; Lacy, 1997; Lande & Shannon, 1996). As technology has changed, we are able to use genomics to understand far more about an individual's or a population's biology than ever before.

Transitioning from a handful of markers to genome-wide panels or even whole genomes has greatly increased the resolution of data we can obtain from both a single sample, and from a population as a whole. The benefits of genomics for conservation have been reviewed in great detail (see Allendorf et al. 2010; Ouborg et al. 2010; Brandies et al. 2019; and Theissinger et al. 2023 for more focused reviews on their utility). A vastly greater number of genomic markers provides more accurate estimates of genome-wide diversity and inbreeding (Allendorf et al. 2010). Hoffman et al. (2014) identified inbreeding depression in a population of harbour seals (*Phoca vitulina*) with a panel of >14,000 SNPs that was not clearly detected with 27 microsatellite loci. Genomic data can improve our ability to detect population structure and delineate conservation units (Ouborg et al. 2010; Funk et al. 2012) as was demonstrated in Atlantic Lobster

(*Homarus americanus*) where the Northern group was thought to be panmictic after an assessment with microsatellites, but six populations were identified in the Northern community based on a large SNP-based study (Benestan et al. 2015). Whole genome sequence data can allow for reconstruction of historical demographic changes and accurate assessment of effective population size (Theissinger et al. 2023). Effective population size of North American gray wolves (*Canis lupus*) recently estimated from genomic data over the past few decades has better informed conservation management identifying that while the populations are not of immediate risk of extinction due to inbreeding depression, there are still concerns for their long-term persistence (vonHoldt et al. 2024). Better genetic profiles can now be obtained from degraded DNA or eDNA samples enabling more questions to be answered with less DNA and less invasive sampling (Andrews et al. 2018). This can improve the effectiveness of conservation actions by early detection of invasive species such as in Illinois, USA where eDNA was used to delineate the invasion front of non-native Asian Carp (*Hypophthalmichthys sp.*) throughout the canals and waterways (Jerde et al. 2011).

We are now able to answer questions with genomics that we couldn't dream of even a decade ago - particularly those of utmost importance for conservation. Still, however, a gap remains between the broad acceptance of the importance of genomics in

conservation and its utilization in management (Shafer et al. 2015; Haig et al. 2016; Cook and Sgrò 2017).

1.3 Marine mammals

In many respects, marine mammals may be considered 'difficult' to study, and arguably more difficult to manage. However, for many species, the individual-level life history data we can obtain, allows for a unique understanding of many aspects of their biology. Many marine mammals are individually identifiable by external colouration or markings (e.g. killer whale saddle patches: Bigg et al. 1976 & Bigg et al. 1986; humpback whale tail colouration: Katona et al. 1979; right whale callosity patterns Kraus et al. 1986), allowing for a vast amount of individual-level life history data to be collected. As long-lived species, from approximately 10 years for harbour porpoises (*Phocoena phocoena*; Read and Hohn 1995) to over 200 years for bowhead whales (*Balaena mysticetus*; George et al. 1999), field observations typically only span a few generations, even for the many dedicated long-term field studies on various species that collect excellent population-level ecological data (reviewed in Mann and Karniski 2017). Pairing this fine scale life history information with genomic data collected from blow,

feces, skin samples etc. can provide an extremely powerful dataset to better understand the conservation needs of a species.

Many conservation concerns for marine mammals arise from threats associated with non-lethal stressors throughout their lifetime. Many cetaceans have a largely coastal or near coastal distribution; meaning their ranges (often including their migratory routes) heavily overlap with anthropogenic activities. Marine mammals, especially cetaceans, are impacted by a variety of anthropogenic stressors including, but not limited to, physical disturbance (such as non-lethal vessel strikes or harassment), acoustic disturbance (which may impact their ability to locate prey), exposure to chemicals and contaminants, and competition for prey resources. Their long lifespans, mean that the cumulative, lifelong effects of non-lethal stressors are likely great and underappreciated as they are difficult to fully detect and quantify even with long-term studies.

Over 35% of cetacean species are listed by the IUCN as Near Threatened or at greater risk of extinction and even more cetaceans are considered at risk at a population level (IUCN 2023; Committee on Taxonomy 2024). Therefore, in addition to the impact of non-lethal stressors through their lifetimes, the genetic consequences of living with small population sizes are also likely impacting many species/populations.

1.4 Marine mammal conservation in Canada

Canada is surrounded by three oceans, and its waters inhabited by over one third of the world's 135 species of marine mammals (Coastal Ocean Research Institute 2016; Oceans North Conservation Society and World Wildlife Fund Canada Ducks Unlimited Canada 2018; Committee on Taxonomy 2024; Marine Animal Response Society 2024). Marine mammals are charismatic sentinels for ocean health, making it easier to gain public buy-in for conservation efforts and meaning successful management of marine mammals will likely have cascading effects that benefit conservation of many other species.

In Canada, legal protection of wildlife largely falls under the Species at Risk Act (SARA). Twenty-three (non-extirpated) marine mammal populations are protected under Schedule 1 of SARA (Government of Canada 2002). The decision to list a species/population takes into account their socioeconomic impacts to Canadians, in addition to scientific evidence of their status as assessed by an independent scientific body: the Committee on the Status of Endangered Wildlife in Canada (COSEWIC) (Government of Canada 2002; Findlay et al. 2009). Differences between these two status listings/assessments (SARA and COSEWIC) are well known, and reflect lags in listing and the complex nature of species protection in Canada where listing decisions are not dictated exclusively by scientific information (see Findlay et al. 2009; Mooers et al. 2010; and Turcotte et al. 2021 for a greater discussion on these issues).

Under SARA, one of the first steps for management of species listed as Endangered, Threatened or Extirpated is the preparation of a Recovery Strategy (Government of Canada 2002). A Recovery Strategy "identifies what needs to be done to arrest or reverse the decline of a species. It sets goals and objectives and identifies the main areas of activities to be undertaken" (Government of Canada 2022). Recovery of a species under SARA means returning the population size to its natural condition in Canada before human activities impacted the species (Environment and Climate Change Canada 2020). Fisheries and Oceans Canada has adopted a precautionary approach to setting recovery goals for the management of many species whereby goals for species/populations are set to 70% of historic maximum population size (DFO 2006; Hammill and Stenson 2007). For many cetaceans that were impacted by whaling, most historic sizes have been estimated based upon whaling records that are often considered incomplete, or inconclusive at the species/population level (Stevick et al. 2001; Jackson et al. 2016). There has been little acknowledgement to date for the power of genomics to help refine these goals using genomic estimates of effective population size over time.

In SARA recovery planning documents pertaining to marine mammals, the cumulative effects of non-lethal stressors are also rarely mentioned. The Recovery Strategy for Blue, Fin and Sei Whales (*Balaenoptera musculus, B. physalus, and B. borealis*)

in Pacific Canadian Waters (Gregr et al. 2006) noted that stressors could be interacting synergistically as is shown in other species, but no further mention of methods to measure or assess cumulative effects were identified in either the Recovery Strategy or the Action Plan (Fisheries and Oceans Canada 2017a). Genetic effects limiting recovery in small populations was noted as a possible impediment to recovery in a review of the recovery actions for St. Lawrence beluga where it was acknowledged that the low genetic diversity in the small population may impede reproductive rates (Fisheries and Oceans Canada 2020) and in the Recovery Strategy for blue whales in the North Atlantic where genetic risks associated with a population bottleneck are discussed (Beauchamp et al. 2009). Understanding these limitations and how they could impact recovery is important for many other small populations/species yet it is rarely mentioned in their respective recovery documents.

All of this is not to say that independent research groups or conservation practitioners do not appreciate genomics as a tool to address these issues, or that future conservation planning won't incorporate more genomic tools, but the ability of genomics to help inform conservation planning of marine mammals appears to be underutilized in Canada.

1.5 Summary of thesis

The goal of my thesis research is to demonstrate how marine mammal conservation management in Canada can be improved by employing genomic tools. The applications presented in my thesis are not meant to be a comprehensive list of ways genomics can aid in conservation, but rather to serve as clear examples how Recovery Strategies and Action/Management Plans can be informed by genomics while simultaneously addressing knowledge gaps for Species at Risk in Canada. The specific objectives of my research are to:

- Demonstrate how the cumulative impact of stressors on an individual may be assessed and quantified using epigenetics.
- Demonstrate how population recovery goals should consider the historic population demography of a species that can be reconstructed from whole genome sequencing.
- Demonstrate how population growth limitations can be identified with genomics.
- Demonstrate how improved resolution from genomic markers can aid in long-term population monitoring.

To meet these objectives, I focus my research on two of Canada's most endangered species/populations of marine mammals: Southern Resident killer whales (*Orcinus orca*) and North Atlantic right whales (*Eubalaena glacialis*). These species/population serve as the best models to clearly meet my study objectives and as endangered populations, are likely to see some of the greatest benefits to having these conservation concerns addressed at this time.

1.5.1 Southern Resident killer whales

Southern Resident killer whales (*Orcinus orca*) are a population of fish-eating killer whales that primarily inhabit the coastal waters of British Columbia and Washington State (Ford et al. 2000). Over 60 killer whales in British Columbia and Washington were removed or killed during captures for the aquarium trade in the 1960s and 1970s, the majority likely coming from the Southern Resident population (Bigg and Wolman 1975). The population has been slow to recover ever since. As of July 2023, the population had only 75 individuals, and only one or two new calves have been added to the population each year for the past five years (Orca Network 2023; Center for Whale Research 2024).

Chinook salmon (*Oncorhynchus tshawytscha*) are a critical component of the Southern Resident killer whale diet. One study found 96% of fish kills by resident killer whales were salmonids, and 65% of those were identified as Chinook salmon (Ford et al. 1998). Ford et al. (2010a) found a very strong link between killer whale mortality and abundance of Chinook salmon. Using aerial photogrammetry, nutritional stress has been identified in Southern resident killer whales where poor body condition was correlated with low prey availability and poor survivorship (Fearnbach et al. 2018; Stewart et al. 2021). There is less concern about nutritional stress in Northern Resident killer whales who forage on a greater number of Chinook stocks (Ford and Ellis 2006; Hanson et al. 2021) and the individual fish consumed by Northern Residents are also thought to be larger than those consumed by Southern Residents (Ford et al. 2010b; Hanson et al. 2021).

Acoustic and physical disturbance are also of concern for Southern Residents. Resident killer whales use acoustic cues to communicate and locate prey (Ford 1989). Anthropogenic noise leads to behavioural changes and may be masking communication and impacting their ability to hunt (Williams et al. 2006; Lusseau et al. 2009). Southern Resident killer whales live in an area with heavy vessel traffic. Over 11,000 large ships transit through Haro Strait, British Columbia, Canada each year (Nuka Research and Planning Group LLC. et al. 2013) and during the summer months a large whale

watching industry – comprised of more than 30 companies (Pacific Whale Watching Association 2017), many of which have several vessels – spend time in close proximity to these whales. The physical disturbance and underwater noise from these vessels likely represent large stressors for Southern Residents.

Population monitoring of Southern Resident killer whales has been ongoing since the 1970s (Bigg 1982), but monitoring of stress in wild killer whale populations has been limited. Stress hormones have been measured from fecal samples and are linked to food supply (Ayres et al. 2012). This echoes other reports documenting nutritional stress in the population (Hilborn et al. 2012; Matkin et al. 2017). Most studies that consider stress focus on acute or seasonal stress, but fail to capture the effects of chronic, long-term and potentially heritable stress. One approach listed in the Action Plan for the Northern and Southern Resident Killer Whale in Canada (Fisheries and Oceans Canada 2017b) is to "Determine the short and long-term effects of chronic and immediate forms of disturbance, including vessels and noise, on the physiology, foraging and social behaviour of Resident Killer Whales" and a separately proposed recovery measure is to "Assess cumulative effects of potential anthropogenic impacts on Resident Killer Whales using an appropriate impact assessment framework for aquatic species". Both of these tasks can at least in part be informed by using genomics to investigate the long-term realized effects of stress on individuals.

1.5.2 North Atlantic right whales

North Atlantic right whales were one of the first targeted species during the whaling era and their name even originates from this history as they were considered the 'right' whale to kill (Kraus and Rolland 2007). It is estimated that over 3,000 North Atlantic right whales were killed between 1634 and 1951 in the western North Atlantic (Reeves et al. 2007). Some accounts suggest that North Atlantic right whale hunting ceased, not because they became less desirable, but because they were seldom seen (Reeves et al. 1978). Models by Reeves et al. (1992) predict that the population may have been limited to fewer than 50 animals by the early 1700s (although there is considerable uncertainty around the nadir of their abundance). Since that time, North Atlantic right whales have struggled to recover, with the most recent population estimate of only 356 individuals remaining (Pettis and Hamilton 2024). Their counterpart, the Southern right whale (*E. australis*), was also a target of whaling; however, unlike North Atlantic right whales, there are estimated to be over 10,000 Southern right whales currently and their population is increasing (Best et al. 2001; Kraus et al. 2005).

North Atlantic right whales face threats from human activity. Entanglement in fishing gear and accidental vessel strikes are the leading causes of known mortalities (Daoust et al. 2017). Changes to fishing practices and vessel speed restrictions have been implemented to help mitigate these risks (Fisheries and Oceans Canada 2018a, 2018b;

Transport Canada 2018). However, while threat mitigation may prevent whale deaths, it does not address the poor reproductive success being exhibited by these animals. Long-term studies enable us to recognize individual whales and track the calving history of females (Hamilton et al. 2007). In comparison to the closely related Southern right whales, North Atlantic right whales are ultimately not reproducing as quickly as expected (Kraus et al. 2001; Browning et al. 2009; Frasier et al. 2023).

Past studies have found that North Atlantic right whales have extremely low genetic diversity (Waldick et al. 2002). The extremely low levels of heterozygosity, and low population size suggest that inbreeding depression could be a likely cause for their poor reproductive success (Schaeff et al. 1997). In an extensive paternity study, Frasier et al. (2013) found that mate choice seems to be random with respect to individuals; however, successful pregnancies that result in a viable calf are more often the result of genetically different gametes. Their results suggest post-copulatory selection in North Atlantic right whales is acting to retain genetic variation (Frasier et al. 2013).

COSEWIC clearly notes in the most recent Status Report for the species that "[p]opulation growth may also be limited by parasites, disease, contaminants, industrial activities, the genetic and demographic effects of small population size, and nutritional stress" (COSEWIC 2013). The need for genetic studies to better understand

recovery was specifically listed in the Action Plan developed for North Atlantic right whales which lists "[c]ollect[ing] tissue, blow, and fecal samples to support hormonal and genetic studies of reproduction, reproductive health, nutritional state, and stress levels" as a specific action item to address the recovery objectives of "monitor[ing] population and threats" and "increase[ing] understanding of life history characteristics, low reproductive rate, habitat and threats to recovery through research" (Fisheries and Oceans Canada 2021).

1.5.3 Outline of thesis research

Chapter 2: An example of DNA methylation as a means to quantify stress in wildlife using killer whales. In this chapter, I demonstrate how methylation profiles at genes involved in stress response can be used as a quantifiable metric of realized cumulative stress levels. I tested the hypothesis that methylation profiles between Northern and Southern resident killer whales would differ, where Southern resident killer whales who seemingly experience greater anthropogenic stressors will exhibit changes to methylation patterns at genes related to stress response that are consistent with those patterns found in controlled studies on rodents.

Chapter 3: A comparison of genomic diversity and demographic history of the North Atlantic and Southwest Atlantic southern right whales. In this chapter, I used whole genome resequencing to help resolve historic effective population sizes of North Atlantic right whale and Southern right whales and quantify contemporary levels of genetic diversity. I tested the hypothesis that the species will exhibit extremely low genome-wide diversity echoing results from fewer markers.

Chapter 4: Effects of inbreeding on reproductive success in endangered North Atlantic right whales. In this chapter, I compared genomic inbreeding coefficients generated with double-digest Restriction Site Associated DNA sequencing (ddRADseq) with a measure of individual fecundity based on field observations to investigate the relationship between inbreeding and reproductive success in female North Atlantic right whales. I tested the hypothesis that if inbreeding depression is affecting female reproductive success, a female's reproductive fecundity will have a negative relationship with her inbreeding coefficient.

Chapter 5: A SNP panel designed for monitoring North Atlantic right whales. In this chapter I designed a panel of single nucleotide polymorphism (SNP) markers that can

be used to genotype individual North Atlantic right whales at over 230 loci simultaneously with genotyping-in-thousands by sequencing (GT-Seq; Campbell et al. 2015). This chapter was not hypothesis driven, but rather aimed to develop a panel of markers that would increase resolution over microsatellites to assign parentage and to be used for ongoing genetic monitoring of the species.

1.6 References

Alava, J. J., K. McMullen, J. Jones, M. J. Barragán-Paladines, C. Hobbs, A. Tirapé, P. Calle, et al. 2023. Multiple anthropogenic stressors in the Galápagos Islands' complex social–ecological system: Interactions of marine pollution, fishing pressure, and climate change with management recommendations. Integrated Environmental Assessment and Management 19:870–895.

Allendorf, F. W., P. A. Hohenlohe, and G. Luikart. 2010. Genomics and the future of conservation genetics. Nature Reviews Genetics 11:697–709.

Andrews, K. R., M. De Barba, M. A. Russello, and L. P. Waits. 2018. Advances in Using Non-invasive, Archival, and Environmental Samples for Population Genomic Studies. Pages 63–99 in P. A. Hohenlohe and O. P. Rajora, eds. Population Genomics: Wildlife.

Ayres, K. L., R. K. Booth, J. A. Hempelmann, K. L. Koski, C. K. Emmons, R. W. Baird, K. Balcomb-Bartok, et al. 2012. Distinguishing the Impacts of Inadequate Prey and Vessel Traffic on an Endangered Killer Whale (*Orcinus orca*) Population. (B. Fenton, ed.) PLoS ONE 7:e36842.

Beauchamp, J., H. Bouchard, P. de Margerie, N. Otis, and J.-Y. Savaria. 2009. Recovery strategy for the blue whale (*Balaenoptera musculus*), Northwest Atlantic population, in Canada [FINAL]. Ottawa.

Benestan, L., T. Gosselin, C. Perrier, B. Sainte-Marie, R. Rochette, and L. Bernatchez. 2015. RAD genotyping reveals fine-scale genetic structuring and provides powerful

population assignment in a widely distributed marine species, the American lobster (*Homarus americanus*). Molecular Ecology 24:3299–3315.

Best, P. B., A. Brandão, and D. S. Butterworth. 2001. Demographic parameters of southern right whales off South Africa. J. Cetacean Res. Manage. Spec Issue 2:161–169.

Bigg, M. 1982. An assessment of killer whale (*Orcinus orca*) stocks off Vancouver Island, British Columbia. Rep. Int. Whal. Commn 32:655–666.

Bigg, M. A., Balcomb, K. C., & Ellis, G. M. (1986). The photographic identification of individual cetaceans. Whalewatcher: Journal of the American Cetacean Society, 20(2), 10–12.

Bigg, M. A., I. B. MacAskie, and G. Ellis. 1976. Abundance and movements of killer whales off eastern and southern Vancouver Island with comments on management. Ste. Anne de Bellevue, Quebec.

Bigg, M. A., and A. A. Wolman. 1975. Live-capture killer whale (*Orcinus orca*) fishery, British Columbia and Washington, 1962-73. Journal of the Fisheries Board of Canada 32:1213–1221.

Brandies, P., E. Peel, C. J. Hogg, and K. Belov. 2019. The Value of Reference Genomes in the Conservation of Threatened Species. Genes 10:846

Brook, B. W., N. S. Sodhi, and C. J. A. Bradshaw. 2008. Synergies among extinction drivers under global change. Trends in Ecology and Evolution 23:453–460.

Browning, C. L., R. M. Rolland, and S. D. Kraus. 2009. Estimated calf and perinatal mortality in western North Atlantic right whales (*Eubalaena glacialis*). Marine Mammal Science 26:648–662.

Campbell, N. R., S. A. Harmon, and S. R. Narum. 2015. Genotyping-in-Thousands by sequencing (GT-seq): A cost effective SNP genotyping method based on custom amplicon sequencing. Molecular Ecology Resources 15:855–867.

Carroll, C. 2007. Interacting effects of climate change, landscape conversion, and harvest on carnivore populations at the range margin: Marten and Lynx in the northern Appalachians. Conservation Biology 21:1092–1104.

Caughley, G. 1994. Directions in Conservation Biology. The Journal of Animal Ecology 63:215–244.

Center for Whale Research. 2024. Orca Population. www.whaleresearch.com <Accessed March, 5, 2024>

Charlesworth, D., and B. Charlesworth. 1987. Inbreeding depression and its evolutionary consequences. Annuals Reviews in Ecology and Systematics 18:237–268.

Coastal Ocean Research Institute. 2016. Mariner's Guide to Whales, dolphins and Porpoises of Western Canada. Vancouver, Canada.

Committee on Taxonomy. 2024. List of marine mammal species and subspecies. Society for Marine Mammalogy. www.marinemammalscience.org <Accessed March, 5, 2024>

Cook, C. N., and C. M. Sgrò. 2017. Aligning science and policy to achieve evolutionarily enlightened conservation. Conservation Biology 31:501–512.

COSEWIC. 2013. COSEWIC Assessment and Status Report on the North Atlantic Right Whale *Eubalaena glacialis*. Committee on the Status of Endangered Wildlife in Canada. Ottawa. xi + 58 pp.

Daoust, P. Y., E. L. Couture, T. Wimmer, and L. Bourque. 2017. Incident Report: North Atlantic right whale mortality event in the Gulf of St. Lawerence, 2017. Collaborative Report Produced by: Canadian Wildlife Health Cooperative, Marine Animal Response Society, and Fisheries and Oceans Canada. 256 pp.

DFO. 2006. A Harvest Strategy Compliant with the Precautionary Approach. DFO Canadian Science Advisory Secretariat Science Advisory Report 2006/023.

Di Prisco, G., D. Annoscia, M. Margiotta, R. Ferrara, P. Varricchio, V. Zanni, E. Caprio, et al. 2016. A mutualistic symbiosis between a parasitic mite and a pathogenic virus undermines honey bee immunity and Health. Proceedings of the National Academy of Sciences of the United States of America 113:3203–3208.

Environment and Climate Change Canada. 2020. Species at Risk Act Policies: Policy on Recovery and Survival. Species at Risk Act Policies and Guidelines Series. 9 pp.

Fearnbach, H., J. W. Durban, D. K. Ellifrit, and K. C. Balcomb. 2018. Using aerial photogrammetry to detect changes in body condition of endangered southern resident killer whales. Endangered Species Research 35:175–180.

Findlay, C. S., S. Elgie, B. Giles, and L. Burr. 2009. Species listing under Canada's species at risk act. Conservation Biology 23:1609–1617.

Fisheries and Oceans Canada. 2017a. Action plan for the blue, fin, sei and North Pacific right whales (*Balaenoptera musculus, B. physalus, B. borealis*, and *Eubalaena japonica*) in Canadian Pacific waters. Species at Risk Act Action Plan Series. Fisheries and Oceans Canada, Ottawa. iv + 28 pp.

———. 2017b. Action Plan for the Northern and Southern Resident Killer Whale (*Orcinus orca*) in Canada. Species at Risk Act Action Plan Series. Fisheries and Oceans Canada, Ottawa. v + 33 pp.

———. 2018a. Minister LeBlanc announces new protections for whales. https://www.canada.ca/ en/fisheries-oceans/news/2018/01/minister_leblancannouncesnewprotections forwhales.html <Accessed October 22, 2018>

———. 2018b. Notice of fisheries closures - Gulf region: Presence of North Atlantic Right whale (June 2, 2018). http://www.dfo-mpo.gc.ca/fm-gp/peches-fisheries/comm/ atl-arc/right-whale-baleine-noires-0206-gulf-golfe-en.html <Accessed October 22, 2018>

———. 2020. St. Lawrence Estuary Beluga. A science based review of recovery actions for three at-risk whale populations. 64pp.

———. 2021. Action Plan for the North Atlantic Right Whale (*Eubalaena glacialis*) in Canada. Species at Risk Act Action Plan Series. Fisheries and Oceans Canada, Ottawa. v + 46 pp.

Fong, C. R., S. J. Bittick, and P. Fong. 2018. Simultaneous synergist, antagonistic and additive interactions between multiple local stressors all degrade algal turf communities on coral reefs. Journal of Ecology 106:1390–1400.

Ford, J. K. B. 1989. Acoustic behaviour of resident killer whales (*Orcinus orca*) off Vancouver Island, British Columbia. Canadian Journal of Zoology 67:727–745.

Ford, J. K. B., and G. M. Ellis. 2006. Selective foraging by fish-eating killer whales *Orcinus orca* in British Columbia. Marine Ecology Progress Series 316:185–199.

Ford, J. K. B., G. M. Ellis, and K. C. Balcomb III. 2000. Killer Whales (2nd Ed.). University of British Columbia Press, Vancouver.

Ford, J. K. B., G. M. Ellis, L. G. Barrett-Lennard, A. B. Morton, R. S. Palm, and K. C. Balcomb III. 1998. Dietary specialization in two sympatric populations of killer

whales (*Orcinus orca*) in coastal British Columbia and adjacent waters. Canadian Journal of Zoology 76:1456–1471.

Ford, J. K. B., G. M. Ellis, P. F. Olesiuk, and K. C. Balcomb. 2010a. Linking killer whale survival and prey abundance: Food limitation in the oceans' apex predator? Biology Letters 6:139–142.

Ford, J. K. B., B. M. Wright, G. M. Ellis, and J. R. Candy. 2010b. Chinook salmon predation by resident killer whales: seasonal and regional selectivity, stock identity of prey, and consumption rates. DFO Can. Sci. Advis. Sec. Res. Doc. 2009/101:iv + 43 p.

Frasier, T. R., R. M. Gillett, P. K. Hamilton, M. W. Brown, S. D. Kraus, and B. N. White. 2013. Postcopulatory selection for dissimilar gametes maintains heterozygosity in the endangered North Atlantic right whale. Ecology and Evolution 3:3483–3494.

Frasier, T. R., P. K. Hamilton, and R. M. Pace. 2023. How compromised is reproductive performance in the endangered North Atlantic right whale? A proposed method for quantification and monitoring. bioRxiv 2011:2023.11.21.568115.

Funk, W. C., J. K. McKay, P. A. Hohenlohe, and F. W. Allendorf. 2012. Harnessing genomics for delineating conservation units. Trends in Ecology and Evolution 27:489–496.

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.

Gilpin, M. E., and M. E. Soulé. 1986. Minimum viable populations : processes of extinction. Pages 19–34 in M. E. Soulé, ed. Conservation Biology: The Science of Scarcity and Diversity. Sinauer Associates, Sunderland, MA.

Government of Canada Species at Risk Act / Loi sur les espèces en péril. S.C. 2002, c.29 Last Amended: December 8, 2023 (2002).

———. 2022. Species at Risk Act: recovery strategies. https://www.canada.ca/en/environment-climate-change/services/species-risk-public-registry/recovery-strategies.html <Accessed March 5, 2024>

Grames, E. M., G. A. Montgomery, C. Youngflesh, M. W. Tingley, and C. S. Elphick. 2023. The effect of insect food availability on songbird reproductive success and

chick body condition: Evidence from a systematic review and meta-analysis. Ecology Letters 26:658–673.

Grassl, J., S. Holt, N. Cremen, M. Peso, D. Hahne, and B. Baer. 2018. Synergistic effects of pathogen and pesticide exposure on honey bee (*Apis mellifera*) survival and immunity. Journal of Invertebrate Pathology 159:78–86.

Gregr, E. J., J. Calambokidis, L. Convey, J. K. B. Ford, R. Perry, L. Spaven, and M. Zacharias. 2006. Recovery Strategy for Blue, Fin and Sei Whales (*Balaenoptera musculus, B. physalus, and B. borealis*) in Pacific Canadian Waters. Species at Risk Act, Recovery Strategy Series. Fisheries and Oceans Canada. Vancouver: Fisheries and Oceans Canada. vii + 53 pp.

Haig, S. M., M. P. Miller, R. Bellinger, H. M. Draheim, D. M. Mercer, and T. D. Mullins. 2016. The conservation genetics juggling act: Integrating genetics and ecology, science and policy. Evolutionary Applications 9:181–195.

Hamilton, P. K., A. R. Knowlton, and M. K. Marx. 2007. Right whales tell their own stories: The Photo identification catalogue. Pages 75–104 in S. D. Kraus and R. M. Rolland, eds. The urban whale: North Atlantic right whales at the crossroads. Harvard University Press, Cambridge, MA.

Hammill, M. O., and G. B. Stenson. 2007. Application of the precautionary approach and conservation reference points to management of Atlantic seals. ICES Journal of Marine Science 64:702–706.

Hanson, M. B., C. K. Emmons, M. J. Ford, M. Everett, K. Parsons, L. K. Park, J. Hempelmann, et al. 2021. Endangered predators and endangered prey: Seasonal diet of Southern Resident killer whales. PLoS ONE 16:1–27.

Hilborn, R., S. P. Cox, F. M. D. Gulland, D. G. Hankin, N. T. Hobbs, D. E. Schindler, and A. W. Trites. 2012. The Effects of Salmon Fisheries on Southern Resident Killer Whales: Final Report of the Independent Science Panel. Prepared with the assistance of D.R. Marmorek and A.W. Hall, ESSA Technologies Ltd. Vancouver, B.C. for National Marine Fisheries Service (Seattle. WA) and Fisheries and Oceans Canada (Vancouver. BC). xv + 61 pp. + Appendices.

Hoffman, J. I., F. Simpson, P. David, J. M. Rijks, T. Kuiken, M. A. S. Thorne, R. C. Lacy, et al. 2014. High-throughput sequencing reveals inbreeding depression in a natural population. Proceedings of the National Academy of Sciences 111:3775–3780.

IUCN. 2023. The IUCN Red List of Threatened Species. Version 2023-1. www.iucnredlist.org <Accessed March 6, 2024>

Jackson, J. A., E. L. Carroll, T. D. Smith, A. N. Zerbini, N. J. Patenaude, and C. S. Baker. 2016. An integrated approach to historical population assessment of the great whales: Case of the New Zealand Southern right whale. Royal Society Open Science 3:150669.

Jerde, C. L., A. R. Mahon, W. L. Chadderton, and D. M. Lodge. 2011. "Sight-unseen" detection of rare aquatic species using environmental DNA. Conservation Letters 4:150–157.

Katona, S., B. Baxter, O. Brazier, S. Kraus, J. Perkins, and H. Whitehead. 1979. Identification of Humpback Whales by Fluke Photographs. Pages 33–44 in Behaviours of Marine Animals: Current Perspectives in Research.

Keen, K. A., R. S. Beltran, E. Pirotta, and D. P. Costa. 2021. Emerging themes in Population Consequences of Disturbance models. Proceedings of the Royal Society B 288:20210325.

Kraus, S. D., M. W. Brown, H. Caswell, C. W. Clark, M. Fujiwara, P. K. Hamilton, R. D. Kenney, et al. 2005. North Atlantic Right Whales in Crisis. Science 309:561–562.

Kraus, S. D., P. K. Hamilton, R. D. Kenney, A. R. Knowlton, and C. K. Slay. 2001. Reproductive parameters of the North Atlantic right whale. Journal of Cetacean Research and Management Special Is:231–236.

Kraus, S. D., K. E. Moore, C. A. Price, M. J. Crone, W. A. Watkins, H. E. Winn, and J. H. Prescott. 1986. The use of photographs to identify individual North Atlantic right whales (*Eubalaena glacialis*). Report of the International Whaling Commission Special Is:145–151.

Kraus, S. D., and R. M. Rolland. 2007. Right Whales in the Urban Ocean. Pages 1–38 in S. D. Kraus and R. M. Rolland, eds. The urban whale: North Atlantic right whales at the crossroads. Harvard University Press, Cambridge, MA.

Lesage, V. 2021. The challenges of a small population exposed to multiple anthropogenic stressors and a changing climate: The St. Lawrence estuary beluga. Polar Research 40:1–18.

Lusseau, D., D. E. Bain, R. Williams, and J. C. Smith. 2009. Vessel traffic disrupts the foraging behavior of southern resident killer whales *Orcinus orca*. Endangered Species Research 6:211–221.

Mann, J., and C. Karniski. 2017. Diving beneath the surface: long-term studies of dolphins and whales. Journal of Mammalogy 98:621–630.

Marine Animal Response Society. 2024. Marine Animals. www.marineanimals.ca <Accessed March 5, 2024>

Marlatt, V. L., S. Bayen, D. Castaneda-Cortès, G. Delbès, P. Grigorova, V. S. Langlois, C. J. Martyniuk, et al. 2022. Impacts of endocrine disrupting chemicals on reproduction in wildlife and humans. Environmental Research 208:112584.

Matkin, C. O., M. J. Moore, and F. M. Gulland. 2017. Review of recent research on Southern Resident Killer Whales to detect evidence of poor body condition in the population. Independent Science Panel Report to the SeaDoc Society. 3 pp. +Appendices

Mooers, A. O., D. F. Doak, C. Scott Findlay, D. M. Green, C. Grouios, L. L. Manne, A. Rashvand, et al. 2010. Science, Policy, and Species at Risk in Canada. BioScience 60:843–849.

Nuka Research and Planning Group LLC., B. C. M. of Environment, Nuka Research and Planning Group LLC., B. C. M. of Environment, Nuka Research and Planning Group LLC., and B. C. M. of Environment. 2013. West Coast Spill Response Study Volume 2: Vessel Traffic Study. 106 pp.

Oceans North Conservation Society, and World Wildlife Fund Canada Ducks Unlimited Canada. 2018. Canada's Arctic Marine Atlas. Ottawa, Canada.

Orca Network. 2023. Southern Resident Orca Community Demographics, Composition of Pods, Births and Deaths since 1990. https://www.orcanetwork.org/births-and-deaths <Accessed March 5, 2024>

Ouborg, N. J., C. Pertoldi, V. Loeschcke, R. K. Bijlsma, and P. W. Hedrick. 2010. Conservation genetics in transition to conservation genomics. Trends in Genetics 26:177–187.

Pacific Whale Watching Association. 2017. List of Members. https://www.pacificwhale watchassociation.com/list-of-members <Accessed January 2, 2018>

Pettis, H. M., and P. K. Hamilton. 2024. North Atlantic Right Whale Consortium 2023 annual report card. Report to the North Atlantic Right Whale Consortium, November 2014.

Pirotta, E., C. G. Booth, D. P. Costa, E. Fleishman, S. D. Kraus, D. Lusseau, D. Moretti, et al. 2018. Understanding the population consequences of disturbance. Ecology and Evolution 8:9934–9946.

Pirotta, E., R. S. Schick, P. K. Hamilton, C. M. Harris, J. Hewitt, A. R. Knowlton, S. D. Kraus, et al. 2023. Estimating the effects of stressors on the health, survival and reproduction of a critically endangered, long-lived species. Oikos 2023: e09801.

Pirotta, E., L. Thomas, D. P. Costa, A. J. Hall, C. M. Harris, J. Harwood, S. D. Kraus, et al. 2022. Understanding the combined effects of multiple stressors: A new perspective on a longstanding challenge. Science of the Total Environment 821:153322.

Read, A. J., and A. A. Hohn. 1995. Life in the fast lane: the life history of harbor porpoises from the Gulf of Maine. Marine Mammal Science 11:423–440.

Reeves, R. R., J. M. Breiwick, and E. Mitchell. 1992. Pre-exploitation abundance of right whales off the eastern United States. Pages 5–7 in J. Hein, ed. The right whale in the western North Atlantic: a science and management workshop. National Marine Fisheries Service. NEFSC Ref. Doc. 92-05, Silver Spring, Maryland.

Reeves, R. R., J. G. Mead, and S. Katona. 1978. The right whale, *Eubalaena glacialis*, in the western North Atlantic. Report of the International Whaling Commission (Vol. 28).

Reeves, R. R., T. D. Smith, and E. A. Josephson. 2007. Near-Annihilation of a Species: Right Whaling in the North Atlantic. Pages 39–74 in S. D. Kraus and R. M. Rolland, eds. The urban whale: North Atlantic right whales at the crossroads. Harvard University Press, Cambridge, MA.

Schaeff, C. M., S. D. Kraus, M. W. Brown, J. S. Perkins, R. Payne, and B. N. White. 1997. Comparison of genetic variability of North and South Atlantic right whales (*Eubalaena*), using DNA fingerprinting. Canadian Journal of Zoology 75:1073–1080.

Shafer, A. B. A., J. B. W. Wolf, P. C. Alves, L. Bergström, M. W. Bruford, I. Brännström, G. Colling, et al. 2015. Genomics and the challenging translation into conservation practice. Trends in Ecology and Evolution 30:78–87.

Stevick, P. T., J. Allens, P. J. Clapham, N. Friday, S. K. Katona, F. Larsen, J. Lien, et al. 2001. Trends in abundance of North Atlantic Humpback whales, 1979-1993. Iwc/Sc/53/Nah2 15.

Stewart, J. D., J. W. Durban, H. Fearnbach, L. G. Barrett-Lennard, P. K. Casler, E. J. Ward, and D. R. Dapp. 2021. Survival of the fattest: linking body condition to prey availability and survivorship of killer whales. Ecosphere 12(8):e03660.

Theissinger, K., C. Fernandes, G. Formenti, I. Bista, P. R. Berg, C. Bleidorn, A. Bombarely, et al. 2023. How genomics can help biodiversity conservation. Trends in Genetics 39:545-559.

Transport Canada. 2018. Protection of the North Atlantic Right Whale – Speed Restriction Measures in the Gulf of St. Lawrence (Series: Ship Safety). Bulletin No. 06/2018

Turcotte, A., N. Kermany, S. Foster, C. A. Proctor, S. M. Gilmour, M. Doria, J. Sebes, et al. 2021. Fixing the Canadian species at risk act: Identifying major issues and recommendations for increasing accountability and efficiency. Facets 6:474–494.

vonHoldt, B. M., D. R. Stahler, K. E. Brzeski, M. Musiani, R. Peterson, M. Phillips, J. Stephenson, et al. 2024. Demographic history shapes North American gray wolf genomic diversity and informs species' conservation. Molecular Ecology 33:1–16.

Waldick, R. C., S. Kraus, M. Brown, and B. N. White. 2002. Evaluating the effects of historic bottleneck events: An assessment of microsatellite variability in the endangered, North Atlantic right whale. Molecular Ecology 11:2241–2250.

Weilgart, L. S. 2007. The impacts of anthropogenic ocean noise on cetaceans and implications for management. Canadian Journal of Zoology 85:1091–1116.

Williams, R., D. Lusseau, and P. S. Hammond. 2006. Estimating relative energetic costs of human disturbance to killer whales (*Orcinus orca*). Biological Conservation 133:301–311.

2 An example of DNA methylation as a means to quantify stress in wildlife using killer whales

2.1 Introduction

Although lethal threats such as directed hunting or incidental mortality (e.g., due to habitat loss or accidental mortality such as bycatch) represent the largest threats to global biodiversity[1–3], there is increasing recognition that non-lethal threats are compromising the viability of many populations, particularly when experienced in combination and/or when they have synergistic effects[4–6]. For example, the global colony collapse of honey bees and the global declines in amphibian populations have both been attributed to the combined effects of multiple non-lethal threats[7,8]. As a result, there is a growing emphasis on the need to consider the cumulative effects of non-lethal threats in the management and conservation of biodiversity[9–11].

Many non-lethal threats—including those from anthropogenic activities such as physical disturbance and contaminant exposure—are thought to increase stress on individual animals, with subsequent negative fitness consequences. Despite being well-documented in humans and mice, quantifying the negative fitness consequences of stress in wildlife populations remains a challenge. Stress hormones can be measured in

blood, saliva, excrement and hair/feathers[12,13]; however, collection of these samples is difficult with many species and stress hormones often only capture a snapshot of an animal's recent stress exposure. Quantifying the cumulative effects of multiple stressors poses an additional obstacle, particularly if their effects are synergistic, as meta-analyses often suggest[14].

One approach that may be particularly informative for understanding the cumulative impacts of stressors is the study of methylation patterns[15,16], and in particular that of genes involved in stress response. DNA methylation is an epigenetic modification to DNA that occurs primarily at 5'-CpG-3' (cytosine-guanine dinucleotide – CpG) sites[17]. CpG sites are not evenly distributed throughout the genome, but rather are concentrated in 'CpG islands' in the promoter region of many genes where they act as important regulators of gene expression[18]. The methylation of some sites in the genome can be influenced by extrinsic factors experienced by individuals including stress exposure, and some may even be non-reversible (i.e., the biological effects of the stressor may last long after the initial stressor is gone)[19–21].

Stress-related changes in DNA methylation have been detected in genes in the hypothalamic pituitary adrenal (HPA) axis[22–26] – a main pathway regulating stress response in mammals[27]. The HPA axis is regulated by a suite of genes including the brain derived neurotrophic factor (BDNF) and the corticotropin releasing factor (CRF)[28].

CRF initiates the stress response pathway that culminates in glucocorticoids (stress hormones) being released[29]. The cellular responses to these hormones are mediated largely by the glucocorticoid receptor (NR3C1)[30]. The complexity of the stress response pathway relies on proper functioning of these and other genes and therefore altered methylation patterns could have significant downstream effects. For example, controlled laboratory experiments on mice and rats have identified changes in methylation patterns in response to stressors in genes in the HPA axis[22–26], and similar results have been identified in human studies[25,31]. Therefore, analyses of methylation patterns of stress-response genes have the potential to be particularly informative in assessing stress in wildlife populations because these patterns represent the cumulative impacts of the suite of stressors experienced by an individual, and therefore provide a method to quantify their cumulative and/or synergistic effects.

The Southern Resident population of killer whales (*Orcinus orca*) in the Northeast Pacific is an endangered population (N = 74)[32] whose main identified threats are the cumulative impacts of multiple non-lethal stressors (e.g., underwater noise, vessel disturbance, food availability and toxic contaminants)[33]. To the north is the closely related and more robust population of Northern Resident killer whales (N ≈ 300)[34]. Northern and Southern Resident killer whale populations have similar dietary preferences and inhabit adjacent, slightly overlapping ranges[35,36]. The Southern Resident

killer whale range has greater overlap with heavily inhabited coastlines and transited waterways compared to Northern Residents and as such the population experiences higher levels of exposure to anthropogenic stressors. We predict that the differential exposure to stressors will manifest themselves as differences in the amount of DNA methylation at CpG sites in genes involved in stress response.

Understanding the combined biological impacts of these non-lethal threats on individual fitness might therefore best be explored by examination of epigenetic differences in genes related to stress. As a step towards this goal, we identified the methylation patterns of stress-related genes in individual Southern Resident killer whales and compared them to those in Northern Resident individuals.

2.2 Results

Using amplicon sequencing of sodium bisulfite treated DNA extracted from biopsy samples of skin from free swimming killer whales, we identified percent methylation for CpG sites in the promoter region of three genes known to be key players in stress response: BDNF, CRF and NR3C1; as well as ACTB (β-actin) - a gene commonly used as a control as it regulates components of the cytoskeleton[37] (figure A2.1). We used a Bayesian regression model to tease apart the relative effects of individual, age, sex, population, and CpG site on methylation patterns (figure 2.1). Neither age nor sex had

appreciable effects on methylation patterns, with the mean of the posterior distributions falling close to zero (figure 2.2a), confirming that methylation of the targeted CpG sites does not change with age or under the different hormonal exposures experienced across the two sexes. Thus, any differences in methylation patterns should be due to differences in the levels of external stressors. To compare the relative effects of being in a given population on the level of DNA methylation at a given site, we took the difference in the posterior probability distributions at each site between populations. If the population of origin had little to no impact on the amount of methylation, the distribution will centre around zero, whereas a positive distribution would indicate hypermethylation in Southern Residents compared to Northern Residents and a negative distribution would be indicative of hypomethylation. As expected, CpG sites within the promoter of the control gene (ACTB) did not show different degrees of methylation across the two populations (figure 2.2b).

Of the genes examined, the CpG sites in the promoter regions of two (BDNF and NR3C1) showed similar degrees of methylation across populations (figure 2.3a,b). However, site-specific differences in methylation patterns were seen in the promoter region of CRF (figure 2.3c). The largest differences were in position -101 that is hypermethylated in Southern Resident individuals relative to the Northern Residents

and in position -95 that is hypomethylated in Southern relative to North Residents (figure 2.3c).

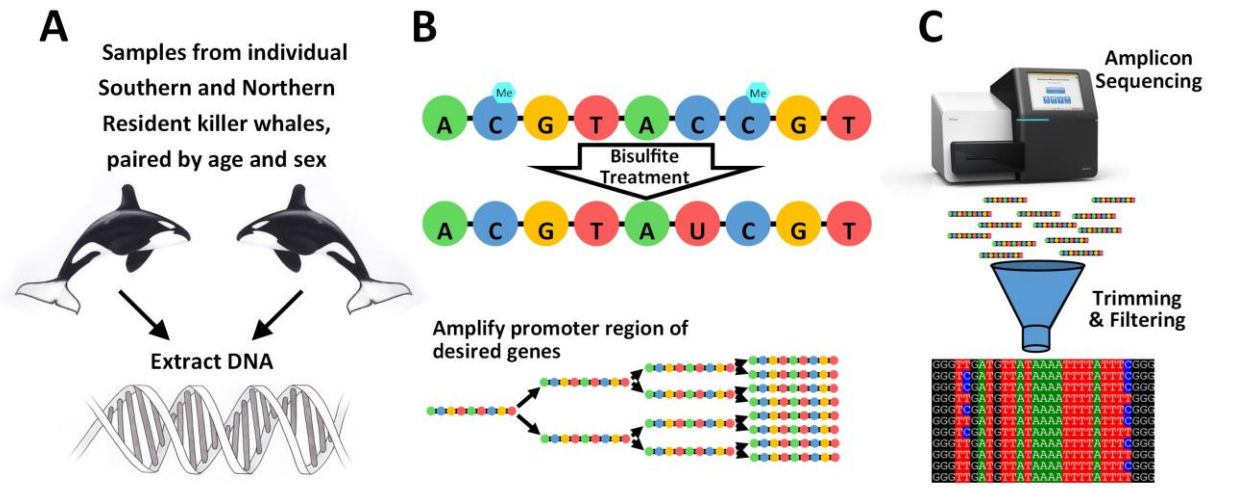

Figure 2.1 Summary of methods to generate percent of methylation at each CpG site. (A) Biopsy samples were collected from Northern and Southern Resident killer whales and DNA was extracted. (B) DNA underwent a bisulfite treatment that converts unmethylated cytosines to uracil. This transformed DNA was used as the template for amplification of four desired genes. (C) Pooled libraries were sequenced on a MiSeq. Output reads were trimmed, filtered, and aligned to reference sequences and percent methylation at each CpG site was calculated for each individual. See Appendix A2 for additional detailed methods. Image of killer whales provided by Dusan Postolovic and of MiSeq courtesy of Illumina, Inc.

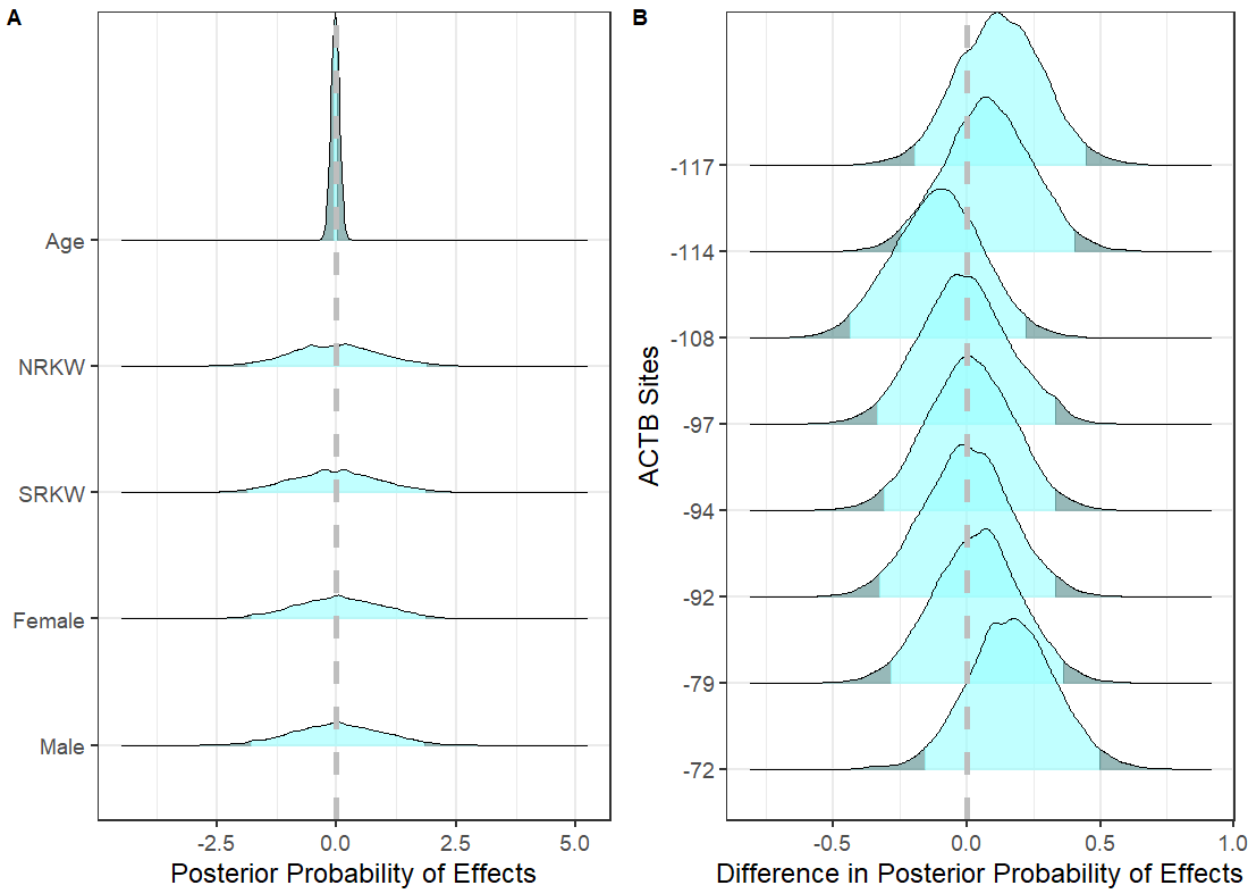

Figure 2.2 Posterior probability of effects. (A) The independent effects of age, population (Northern and Southern Resident killer whales: NRKW and SRKW) and sex on percent methylation of examined CpG sites showed little effect of total methylation patterns. (B) The control gene ACTB showed little to no difference in methylation patterns between Southern and Northern Residents. Y-axis is the number of base pairs prior to the transcriptional start site (TSS), hence the negative values. Positive X-values indicate hypermethylation in Southern Residents relative to Northern Residents and negative values indicate hypomethylation. The distribution falling outside the 95% highest density interval is indicated with the darker shading.

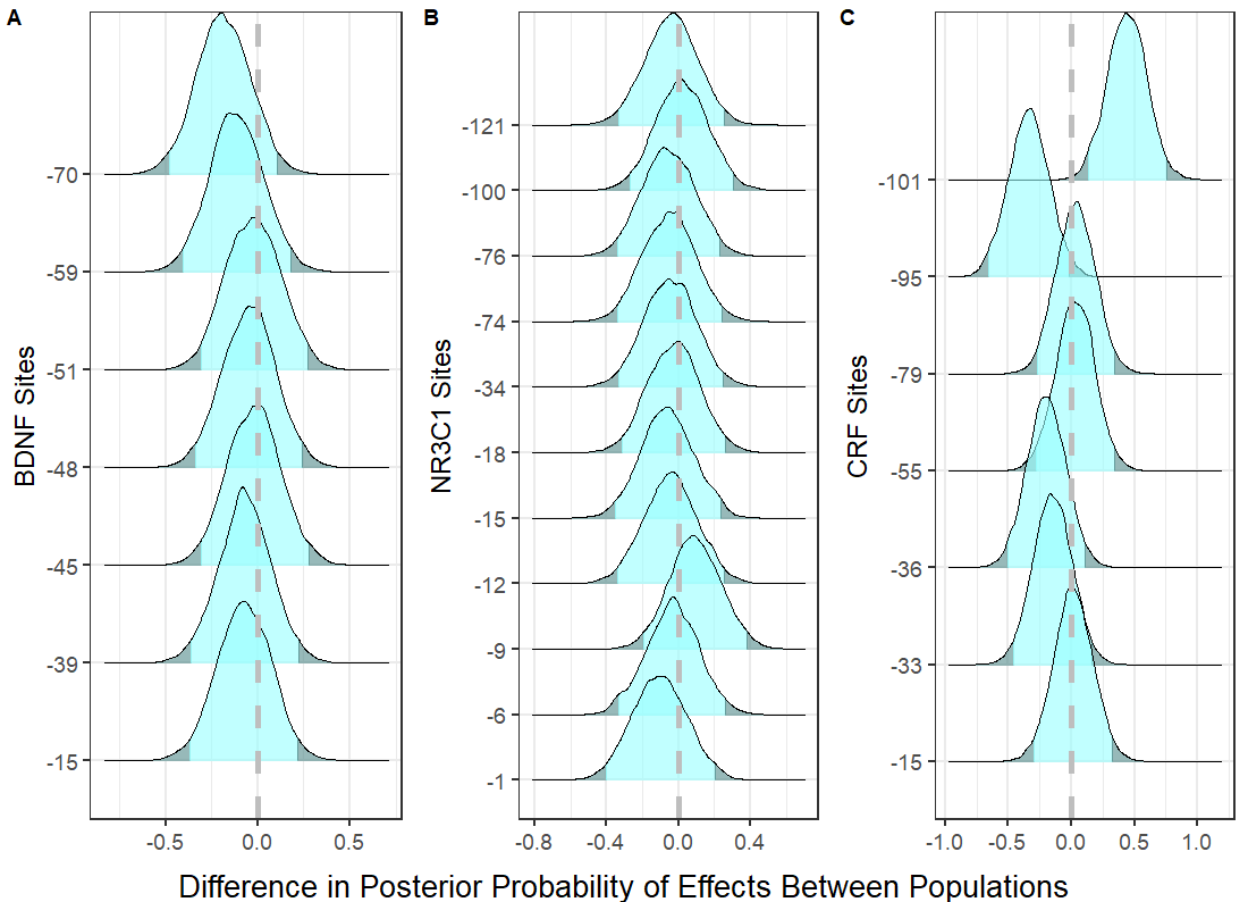

Figure 2.3 Differences in the posterior probability distributions of estimated effects between Northern and Southern Resident killer whale populations. Included are estimates for each CpG site identified by their distance from the TSS for (A) BDNF, (B) NR3C1 and (C) CRF. Positive values indicate Southern Residents are hypermethylated compared to Northern Residents and negative values indicate hypomethylation in Southern Residents compared to Northern Residents. The distribution falling outside the 95% highest density interval is indicated with the darker shading.

2.3 Discussion

Methylation patterns at stress response genes differed between Northern and Southern

Resident killer whales that could not be explained by age or sex and are likely due to

differences in stress exposure. Many studies have demonstrated that methylation patterns at some CpG sites change with age in very predictable ways and can be used in age-predictive models[38]. We chose a paired sample approach to help elucidate the possible effect of population, while also accounting for potential effects of age and sex. Our results suggest the CpG sites in the genes targeted in our study are not affected by age or sex, but rather methylation difference reflect differences in the environments experienced by Resident killer whales.

While genetically distinct[39], these two populations of killer whales are remarkably similar. The have the same dietary preferences, adjacent but overlapping distributions, and very similar physical appearances and acoustic behaviours[35,36]. One of the primary differences in the environments experienced by these two populations is the degree of exposure to anthropogenic stress. Southern Residents spend much of the year in the interior waters near the busy port cities of Vancouver and Victoria, Canada; and Seattle, USA, and therefore the whales in this area are exposed to high levels of pollution, heavy vessel traffic from recreational boaters, a large commercial whale watching fleet, and large cargo vessels causing both acoustic and physical disturbance[33]. In contrast, Northern Residents inhabit quieter, more pristine waters to the north with less commercial and recreational vessel presence.

The patterns we identified in CRF (hypermethylation at position -101 and hypomethylation at position -95) have also been observed in some tissues of rats exposed to chronic variable mild stress[26] and since the CRF promoter region in killer whales has over 95% similarity with other mammalian species (figure A2.2), this suggests that they serve the same functions across species. Additionally, altered methylation patterns in response to stress have previously been identified at position -101 in other species, but the direction of these methylation changes has differed between tissue types and studies[22,23].

Only one of these two sites (position -101) falls within a previously identified transcription factor binding region: that for the aryl hydrocarbon receptor (Ahr)[40,41]. The Ahr binding region (position -101) is particularly interesting because, in addition to being involved in stress response, aryl hydrocarbon receptors are also involved in the metabolism of polycyclic aromatic hydrocarbons (PAHs)[42], contaminants to which killer whales are frequently exposed but which do not tend to bioaccumulate and exposure is therefore difficult to assess[43].

Current practices for understanding the cumulative effects of multiple stressors on individuals rely on models that incorporate estimated effects of independent stressors and make assumptions about how these stressors interact and are experienced by the individual[44,45]. The potential to use changes in the amount of methylation at

stress-related genes within a population over time or between populations, may therefore represent a stronger approach for understanding the true biological impact and fitness consequences of multiple stressors experienced by individuals. This study represents the first step in demonstrating the application of this method for a mammalian population, which may prove to be an effective long term monitoring tool for wildlife populations.

For Southern Residents, the biological implications of these changes in methylation are yet to be determined. Further investigations into the biological or physiological implications of these methylation differences are needed, but our results suggest that long-term monitoring of levels of methylation in stress-related genes may provide a useful tool for quantifying the synergistic effects of multiple stressors, and help track the effectiveness of threat mitigation efforts on reducing stress on individual killer whales.

The cumulative and synergistic effects of anthropogenic-induced stress are likely influencing a wide range of wildlife populations, but quantifying such effects has not yet been possible. Assessing DNA methylation of genes involved in stress response may represent a promising option, as suggested by the application of this approach to critically endangered Southern Resident killer whales. Although we do not yet know

the downstream biological implications of the differential methylation patterns observed, this represents a crucial first step in making such quantification possible.

2.4 Methods

2.4.1 Library Preparation and Sequencing

DNA samples for this study are part of a long-term tissue collection not collected explicitly for this study. Skin samples used were collected from free-swimming whales for previous work using a pneumatic dart system[46], except for one sample that was collected from a freshly deceased carcass. During biopsy sampling, biopsy darts collect a small piece of skin from the dorsal surface of the animal, near its dorsal fin. This biopsy method is consistent between populations. A paired sample approach was implemented when selecting which of the existing samples to use so that for each Southern Resident sample with sufficient high-quality DNA (n = 17), one or two Northern Residents of the same sex and similar age were included (n = 30) (figure 2.4).

The DNA was previously extracted using standard phenol:chloroform methods[47] using the skin adjacent to the skin-blubber interface of each sample. 400ng of DNA from each individual was subjected to sodium bisulfite conversion using Epitect Bisulfite Kits (Qiagen). This process converted unmethylated cytosines (C) in CpG sites to uracil (U), whereas methylated cytosines in CpG sites remained C.

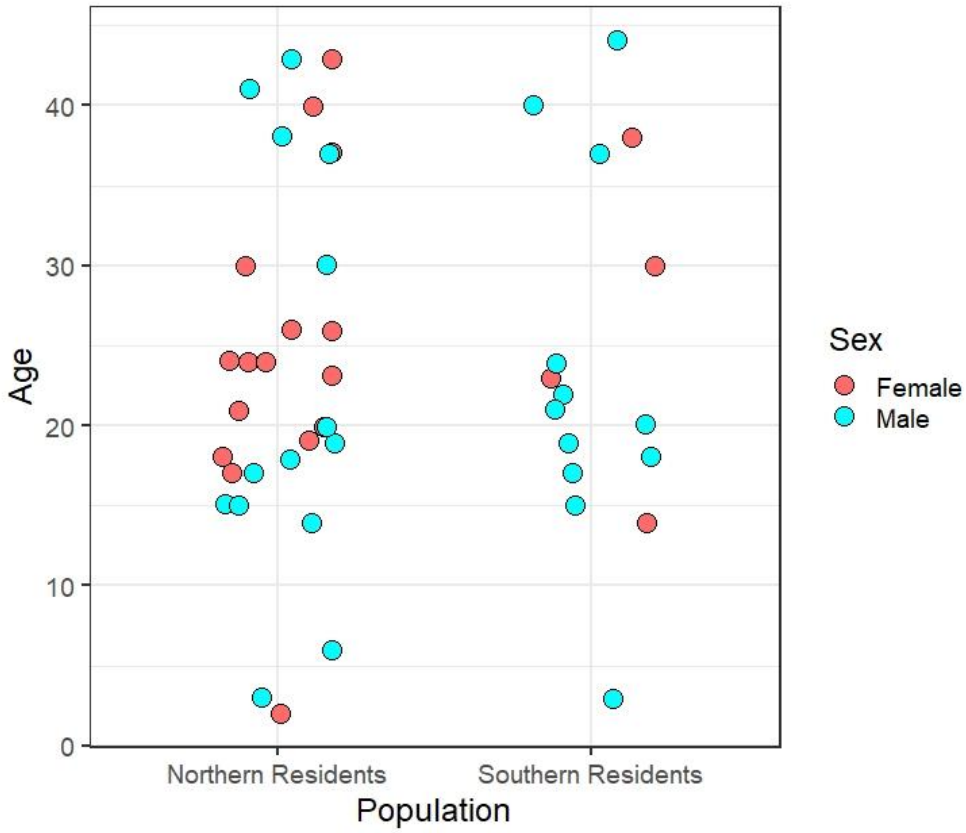

Figure 2.4 The distribution of age and sex of individuals in each population used in this study.

We identified three genes in the HPA axis and its regulation that had demonstrated methylation changes in response to stressors in controlled studies of laboratory animals: brain derived neurotropic factor (BDNF), corticotropin releasing factor (CRF) and glucocorticoid receptor (NR3C1) as well as a common control gene, β-actin (ACTB) whose methylation patterns should not be influenced by age, sex, or exposure to external stressors[37], rather interpopulation differences here could represent

genetic differences in methylation/demethylation enzymes. We identified reference sequences for the promoter region of each of these genes from the published killer whale genome (GenBank accession GCA_000331955.1). We did an in-silico sodium bisulfite conversion of our reference sequences and designed primers accounting for the DNA being single stranded and non-complementary post bisulfite treatment, avoiding CpG sites within the primer sequence if possible (see Appendix A2). We performed Sanger sequencing to test that: (i) we had amplified the correct region, and (ii) the resulting sequences would be absent of cytosine except in potentially methylated sites. Each bisulfite-treated sample was amplified using each primer pair. Our subsequent library preparation protocol was a modified version of that available for 16S sequencing[48].

We cleaned our PCR products with Ampure XP using bead:sample ratios optimized for each primer set and final elution volumes adjusted for each sample as determined by concentration on a pre-cleanup agarose gel. We performed an indexing PCR to bind sample-specific indexes using the Nextera xt Index kit (Illumina) and cleaned the products. The final libraries were quantified, pooled and sequenced on an Illumina MiSeq 250x250 PE run at the BRC Sequencing Core at the University of British Columbia (Vancouver, Canada).

2.4.2 Processing Amplicon Sequences

Each index was unique to a single individual but included amplicons from multiple loci. Separating loci and trimming primer sequences was performed in CUTADAPT 2.6 with an allowable error rate of 0.2[49]. Leading and trailing bases with a Q < 20 were trimmed and reads were trimmed when the mean quality in a 5-base sliding window fell below Q = 30 in TRIMMOMATIC 0.36[50]. Paired reads were merged using FLASH 1.2.11[51]. Aligned and indexed bam files were compiled using BWA 0.7.17-r1188 and SAMTOOLS 1.7[52,53] and aligned against the reference genes used to design our primers.

Despite normalizing our library concentrations prior to sequencing, we had differential amplification and sequencing success of each locus resulting in an uneven distribution of reads across loci. As a result, we determined minimum read depth thresholds independently for each locus. Our goal was to identify percent methylation at each site and therefore wanted to optimize read depth to increase confidence in our measures of percent methylation while still maintaining relatively consistent certainty within each locus. We visualized read depths using histograms of read depth across samples and determined the minimum threshold for each locus as: ACTB – 10 000, BDNF – 20 000, CRF – 100, NR3C1 – 500 (table A2.2 and figure A2.3). Individuals were excluded from the analysis if a minimum read depth threshold was not met. Due to low

read depths of merged pair-end reads for NR3C1, forward unpaired reads were also trimmed and included in the aligned bam file to increase read depth.

2.4.3 Analyses

Using the reference sequence for each locus, we identified potentially methylated sites (CpG sites). Methylated CpG sites in our reads would have escaped bisulfite conversion and remained as C bases whereas unmethylated CpG sites would be represented by Ts. For every individual at each gene, we calculated the proportion of C/T base calls as a metric of percent methylation at each CpG site.

We built a Bayesian regression model using the percent methylation of each site as the predicted variable where individual, age, sex, population and CpG site were the predictor variables, as well as the interaction between specific CpG sites and population. We also allowed for different standard deviations in each population. We built similar models investigating other pairwise interactions and none had an effect on our results, so we left them out of our final model. (See Appendix A2 for additional details of the model and its performance).

To better visualize and understand the differences between populations, we plotted the difference in the posterior probabilities of percent methylation at each site calculated as that from Northern Residents subtracted from that for the Southern Residents.

2.5 Approval and ethics statement

DNA samples for this study are part of a long-term tissue collection not collected explicitly for this study. Therefore, this study did not require any interaction with live animals. All biopsies were collected under relevant guidelines and regulations set out by permits issued by Fisheries and Oceans Canada and/or the US National Marine Fisheries Service. Sample collection was carried out by a number of individuals representing multiple institutions. All biopsies were collected with permission and approval from their respective ethical committee or institutional review boards. The collection of most samples was performed under approval of the University of British Columbia and their Animal Care and Use Program.

2.6 Data, code and materials

Code and methylation data are available at

https://github.com/carlacrossman/StressMethylation

2.7 References

1. Schipper, J. et al. The Status of the World's Land and Marine Mammals: Diversity, Threat, and Knowledge. Science 322, 225–230 (2008).

2. Cushman, S. A. Effects of habitat loss and fragmentation on amphibians: A review and prospectus. Biol. Conserv. 128, 231–240 (2006).

3. Díaz, S. et al. Pervasive human-driven decline of life on Earth points to the need for transformative change. Science 366, (2019).

4. Geary, W. L., Nimmo, D. G., Doherty, T. S., Ritchie, E. G. & Tulloch, A. I. T. Threat webs: Reframing the co-occurrence and interactions of threats to biodiversity. J. Appl. Ecol. 56, 1992–1997 (2019).

5. Halpern, B. S. et al. A Global Map of Human Impact on Marine Ecosystems. Science 319, 948–952 (2008).

6. Johnson, C. N. et al. Biodiversity losses and conservation responses in the Anthropocene. Science 356, 270–275 (2017).

7. Di Prisco, G. et al. A mutualistic symbiosis between a parasitic mite and a pathogenic virus undermines honey bee immunity and Health. Proc. Natl. Acad. Sci. U. S. A. 113, 3203–3208 (2016).

8. Grant, E. H. C. et al. Identifying management-relevant research priorities for responding to disease-associated amphibian declines. Glob. Ecol. Conserv. 16, e00441 (2018).

9. Schindler, D. W. The cumulative effects of climate warming and other human stresses on Canadian freshwaters in the new millennium. Can. J. Fish. Aquat. Sci. 58, 18–29 (2001).

10. Cumulative Effects in Wildlife Management: Impact Mitigation. (CRC Press, 2011). doi:10.1017/CBO9781107415324.004

11. Frid, A. & Dill, L. Human-caused disturbance stimuli as a form of predation risk. Conserv. Ecol. 6, 11 (2002).

12. Rolland, R. M., Hunt, K. E., Kraus, S. D. & Wasser, S. K. Assessing reproductive status of right whales (*Eubalaena glacialis*) using fecal hormone metabolites. Gen. Comp. Endocrinol. 142, 308–317 (2005).

13. Sheriff, M. J., Dantzer, B., Delehanty, B., Palme, R. & Boonstra, R. Measuring stress in wildlife: Techniques for quantifying glucocorticoids. Oecologia 166, 869–887 (2011).

14. Crain, C. M., Kroeker, K. & Halpern, B. S. Interactive and cumulative effects of multiple human stressors in marine systems. Ecol. Lett. 11, 1304–1315 (2008).

15. Beal, A., Rodriguez-Casariego, J., Rivera-Casas, C., Suarez-Ulloa, V. & Eirin-Lopez, J. M. Environmental epigenomics and its applications in marine organisms. in Population genomics: marine organisms (eds. Oleksiak, M. F. & Rajora, O. P.) 325–359 (Springer, Cham, 2018). doi:10.1007/13836_2018_28

16. Eirin-Lopez, J. M. & Putnam, H. M. Marine environmental epigenetics. Ann. Rev. Mar. Sci. 11, 335–368 (2019).

17. Laird, P. W. The power and the promise of DNA methylation markers. Nat. Rev. Cancer 3, 253–266 (2003).

18. Bird, A. P. CpG-rich islands and the function of DNA methylation. Nature 321, 209–213 (1986).

19. Cedar, H. & Bergman, Y. Linking DNA methylation and histone modification: patterns and paradigms. Nat. Rev. Genet. 10, 295–304 (2009).

20. Matosin, N., Cruceanu, C. & Binder, E. B. Preclinical and clinical evidence of DNA methylation changes in response to trauma and chronic stress. Chronic Stress 1, 247054701771076 (2017).

21. Radtke, K. M. et al. Transgenerational impact of intimate partner violence on methylation in the promoter of the glucocorticoid receptor. Transl. Psychiatry 1, e21-6 (2011).

22. Mueller, B. R. & Bale, T. L. Sex-Specific Programming of Offspring Emotionality after Stress Early in Pregnancy. J. Neurosci. 28, 9055–9065 (2008).

23. Elliott, E., Ezra-Nevo, G., Regev, L., Neufeld-Cohen, A. & Chen, A. Resilience to social stress coincides with functional DNA methylation of the Crf gene in adult mice. Nat. Neurosci. 13, 1351–1353 (2010).

24. Boersma, G. J. et al. Prenatal stress decreases Bdnf expression and increases methylation of Bdnf exon IV in rats. Epigenetics 9, 437–447 (2013).

25. Turecki, G. & Meaney, M. J. Effects of the Social Environment and Stress on Glucocorticoid Receptor Gene Methylation: A Systematic Review. Biol. Psychiatry 79, 87–96 (2016).

26. Sterrenburg, L. et al. Chronic stress induces sex-specific alterations in methylation and expression of corticotropin-releasing factor gene in the rat. PLoS One 6, 1–14 (2011).

27. Reeder, D. A. M. & Kramer, K. M. Stress in free-ranging mammals: Integrating physiology, ecology, and natural history. J. Mammal. 86, 225–235 (2005).

28. Jeanneteau, F. D. et al. BDNF and glucocorticoids regulate corticotrophin-releasing hormone (CRH) homeostasis in the hypothalamus. Proc. Natl. Acad. Sci. U. S. A. 109, 1305–1310 (2012).

29. Smith, S. M. & Vale, W. W. The role of the hypothalamic-pituitary-adrenal axis in neuroendocrine responses to stress. Dialogues Clin. Neurosci. 8, 383–395 (2006).

30. Turner, J. D. & Muller, C. P. Structure of the glucocorticoid receptor (NR3C1) gene 5′ untranslated region: Identification, and tissue distribution of multiple new human exon 1. J. Mol. Endocrinol. 35, 283–292 (2005).

31. Bakusic, J., Schaufeli, W., Claes, S. & Godderis, L. Stress, burnout and depression: A systematic review on DNA methylation mechanisms. J. Psychosom. Res. 92, 34–44 (2017).

32. Center for Whale Research. Population. (2020). Available at: www.whaleresearch.com. (Accessed: 11th January 2021)

33. Fisheries and Oceans Canada. Recovery Strategy for the Northern and Southern Resident Killer Whales (*Orcinus orca*) in Canada Killer Whale [Proposed] Species at Risk Act Recovery Strategy Series. Fish. Ocean. Canada x + 84 pp. (2018).

34. DFO. Population Status Update for the Northern Resident Killer Whale (*Orcinus orca*) in 2018. (2019).

35. Bigg, M. A., Olesiuk, P. F., Ellis, G. M., Ford, J. K. B. & Balcomb, K. C. Social organization and genealogy of resident killer whales (*Orcinus orca*) in the coastal waters of British Columbia and Washington State. Reports Int. Whal. Comm. 12, 383–405 (1990).

36. Ford, J. K. B. & Ellis, G. M. Selective foraging by fish-eating killer whales *Orcinus orca* in British Columbia. Mar. Ecol. Prog. Ser. 316, 185–199 (2006).

37. Chen, I.-H. et al. Selection of reference genes for RT-qPCR studies in blood of beluga whales (Delphinapterus leucas). PeerJ 4, e1810 (2016).

38. Horvath, S. & Raj, K. DNA methylation-based biomarkers and the epigenetic clock theory of ageing. Nat. Rev. Genet. 19, 371–384 (2018).

39. Hoelzel, A. R., Dahlheim, M. E. & Stern, S. J. Low genetic variation among killer whales (*Orcinus orca*) in the eastern north Pacific and genetic differentiation between foraging specialists. J. Hered. 89, 121–128 (1998).

40. Yao, M., Stenzel-Poore, M. & Denver, R. J. Structural and functional conservation of vertebrate corticotropin- releasing factor genes: Evidence for a critical role for a conserved cyclic AMP response element. Endocrinology 148, 2518–2531 (2007).

41. Aguiniga, L. M., Yang, W., Yaggie, R. E., Schaeffer, A. J. & Klumpp, D. J. Acyloxyacyl hydrolase modulates depressive-like behaviors through aryl hydrocarbon receptor. Am. J. Physiol. - Regul. Integr. Comp. Physiol. 317, R289–R300 (2019).

42. Hankinson, O. The Aryl Hydrocarbon Receptor Complex. Annu. Rev. Pharmacol. Toxicol. 35, 307–340 (1995).

43. Lundin, J. I. et al. Pre-oil spill baseline profiling for contaminants in Southern Resident killer whale fecal samples indicates possible exposure to vessel exhaust. Mar. Pollut. Bull. 136, 448–453 (2018).

44. MacDonald, L. H. Evaluating and managing cumulative effects: Process and constraints. Environ. Manage. 26, 299–315 (2000).

45. National Academies of Sciences Engineering and Medicine. Approaches to Understanding the Cumulative Effects of Stressors on Marine Mammals. (National Academies Press, 2017). doi:10.17226/23479

46. Barrett-Lennard, L. G., Smith, T. G. & Ellis, G. M. A cetacean biopsy system using lightweight pneumatic darts, and its effect on the behavior of killer whales. Mar. Mammal Sci. 12, 14–27 (1996).

47. Sambrook, J., Fritsch, E. F. & Maniatis, H. Molecular cloning: a laboratory manual. (Cold Springs Harbor Laboratory Press, 1989).

48. Illumina. 16S Metagenomic Sequencing Library Preparation. Illumina.com 1–28 (2013).

49. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 17, 10–12 (2011).

50. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120 (2014).

51. Magoč, T. & Salzberg, S. L. FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics 27, 2957–2963 (2011).

52. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–1760 (2009).

53. Li, H. et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079 (2009).

3 A comparison of genomic diversity and demographic history of the North Atlantic and Southwest Atlantic southern right whales

3.1 Introduction

Since its inception, conservation biology has been a field dedicated to using multidisciplinary approaches to protect biodiversity. Genetics has long served as an important tool for conservation biology through being able to detect population structure, monitor inbreeding and measure genetic diversity (Frankel 1974; Frankham 1995b). Advances in technology and the accessibility of genomic data have vastly increased the ways in which genomics can help inform conservation (Theissinger et al. 2023). Whole genome sequence data is easier and more accessible than ever and can help address vastly more questions with greater resolution than microsatellites, mitochondrial regions, or even reduced representation sequencing (such as RADseq – Restriction-site Associated DNA sequencing or GBS – Genotyping by Sequencing) could before including understanding how population demographic history has shaped the modern genomic landscape (Allendorf et al. 2010; Ouborg et al. 2010; Taylor et al. 2021). Whole genome sequencing data can identify runs of homozygosity more accurately and precisely than reduced-representation sequencing method with low marker density (Duntsch et al. 2021) and the ability to detect rare variants is again greatly improved

with whole genome sequencing (North et al. 2021). These advances have greatly improved the insights we can glean from a single sample from species that are otherwise difficult to study.

Marine mammals are difficult targets for conservation: their ranges usually cross political and jurisdictional boundaries, their major threats are often related to important economic industries (e.g. fishing, shipping, coastal development); and their coastal or oceanic ranges, combined with their limited activity at the surface, make them difficult to study and monitor. Despite these difficulties, efforts across the globe are underway to help protect marine mammals and minimize anthropogenic mortalities, but often fail to consider how the historical population demography of a species may be affecting its contemporary populations.

Balaenidae is a family of baleen whales that encompasses the bowhead whale (*Balaena mysticetus*), and three species of right whales: the North Atlantic right whale (*Eubalaena glacialis*), the North Pacific right whale (*E. japonica*) and the southern right whale (*E. australis*). Female right whales become reproductively active around eight years of age (Hamilton et al. 1998; Best et al. 2001) and are capable of reproducing every three years with little evidence of reproductive senescence (Knowlton et al. 1994; Best et al. 2001; Kraus et al. 2007). Right whales inhabit temperate waters around the globe

migrating to slightly warmer waters in the winter to calve and back to cooler waters in the summer to feed (Braham & Rice 1984). The high body fat content of Balaenidae (over 40%; Lockyer 1976), combined with their extremely long baleen, made them prime target species during the whaling era.

The onset of commercial whaling by the Basques before 1100 AD (Aguilar 1986; Reeves et al. 2007) began a multi-century long decimation of whale populations around the globe (Reeves et al. 2007; Reeves & Smith 2007). North Atlantic right whales in the eastern North Atlantic may have been one of the first populations of whales to be completely eradicated as the coastal nature of the species made them an easy target (Aguilar 1986; Reeves et al. 2007). With the decimation of right whales in the eastern North Atlantic, Basque whalers then moved across to the western North Atlantic where they established whaling sites along the coasts of what are now Newfoundland, Labrador and Quebec by the mid-1500s (e.g. Aguilar 1986). Although it was once thought that Basque whalers caused the greatest decline in western North Atlantic right whales, reducing them from around 10,000 individuals to perhaps one or a few thousand (e.g. Aguilar 1986; Cumbaa 1986; Gaskin 1991), more recent genetic data from bones recovered from these whaling sites show that Basques targeted bowhead whales - and not right whales - in the western North Atlantic, raising questions about the impact of Basque whaling on this population (Rastogi et al. 2004; McLeod et al. 2008, 2010).

While the impact of Basque whaling is debated, most agree that the American whaling of North Atlantic right whales that followed largely ceased as they became too rare for hunting to be profitable and were considered functionally extinct (Reeves et al. 1992, 2007; Reeves 2001). In the southern oceans, southern right whales were an equally heavily targeted species with estimates of over 140,000 individuals removed by whaling prior to the 1900s (IWC 2001).

Today, North Atlantic right whales are listed as critically endangered by the IUCN and endangered under the U.S. Endangered Species Act (NOAA 2008; IUCN 2022) with an estimated 336 individuals remaining in a single population (Pettis et al. 2022). North Pacific right whales are listed as endangered by the IUCN (IUCN 2022) and are divided into two differentiated populations on either side of the North Pacific based on mitochondrial DNA (Pastene et al. 2022). The Northeast Pacific population is made up of only ~30 individuals (Wade et al. 2011) and in the Northwest Pacific, Hakamada and Matsuoka (2016) report over 1,100 individuals, although it should be noted this estimate is based on limited sighting data and has not been reviewed by the IWC Scientific Committee. In contrast, southern right whales are listed by the IUCN as least concern with estimates of over 13,000 individuals globally (International Whaling Commission 2013; Cooke & Zerbini 2018). The major ocean basins in the southern hemisphere form two genetically distinct clusters of southern right whales in the South

Atlantic and in the Indo-Pacific (Patenaude et al. 2007). A number of winter nursing grounds have been described throughout their range (Baker et al. 1999; IWC 2001), including two genetically differentiated groups in South Atlantic found off the coasts of Argentina and South Africa (Patenaude et al. 2007; Carroll et al. 2019, 2020). Southern right whale populations have experienced different recoveries post whaling with some populations experiencing near maximal growth rate (Harcourt et al. 2019), while others remain highly endangered (Cooke 2018). The largest population in the Southwest Atlantic off the coast of Argentina was estimated to have over 4,000 individuals in 2009 (International Whaling Commission 2013) and was increasing throughout the late 2000s and early 2010s (Crespo et al. 2019).

Many of these studies assessing population structure in right whales also measured genetic diversity. Across studies, using mitochondrial DNA and microsatellites, estimates of diversity were consistently lower in North Atlantic right whales than in southern right whales from Argentina in the Southwest Atlantic (Malik et al. 2000; Waldick et al. 2002; Patenaude et al. 2007; Carroll et al. 2019, 2020). The reported levels of genetic diversity in the North Atlantic right whales are lower than the cheetah and represent one of the lowest genetic diversity levels reported for a wildlife species (Frasier et al. 2007). Interestingly, despite low overall diversity, heterozygosity in North Atlantic right whale calves is higher than expected given the genotypes of their

parents, suggesting lower success of inbred fetuses versus those that are more heterozygous (Frasier et al. 2013). To date, genome wide assessments of diversity in right whales (both whole genome and reduced representation genome sequencing) are limited and often restricted to single individuals (e.g. Wolf et al. 2022), or do not include North Atlantic right whales (Cabrera et al. 2022).

Given the large impact whaling had on many populations, but often with incomplete records of how many individuals were actually killed, particularly during the period of Basque whaling, genetic methods to infer historical demography can be particularly informative. Different methods have been developed to estimate ancestral effective population sizes from genomic sequence data based on site frequency spectrum (SFS), and identity by descent blocks (e.g. ðaði - Gutenkunst et al. 2009; MSMC - Schiffels & Durbin 2014; IBDNe - Browning & Browning 2015; Stairwayplot2 - Liu & Fu 2020). These models have been used to address conservation questions in marine mammals such as improving our understanding of killer whale population history and diversity (Foote et al. 2021), understanding the impact of whaling on fin whales (Wolf et al. 2022), and assessing the demographic history of the endangered vaquita (Morin et al. 2021; Robinson et al. 2022). Applying these models to right whales could improve our understanding of their history and inform our approaches for contemporary conservation. For example, if western North Atlantic right whales existed

at smaller population sizes than southern right whale populations throughout much of their history, then this could help explain their different patterns of recovery from whaling.

Moreover, understanding the demographic history of a population can help explain its contemporary diversity. Populations that have undergone a bottleneck event may have been subjected to a large reduction in genetic diversity that will be evident even after the population rebounds. For example, northern elephant seals (*Mirounga angustirostrus*) underwent an extreme bottleneck event over one hundred years ago where hunting reduced their numbers to only 10-20 individuals, but subsequently the species has grown to over 100,000 individuals (Hoelzel 1999). This drastic reduction in population size is still reflected in their reduced genetic diversity compared to the southern elephant seal (*M. leonina*) (Hoelzel 1999). Unlike the rapid population decline experienced by northern elephant seals, the endangered vaquita (*Phocoena sinus*) have lived with small population sizes for hundreds of generations, yet they show little evidence of inbreeding depression and possess a low burden of deleterious mutations despite decades of low numbers (Robinson et al. 2022). Demographic history can therefore provide important context for interpreting the current landscape of genetic diversity and the potential implications for conservation.

A recent study investigated how demographic changes in baleen whales, including right whales have been correlated with changes in abundance of their prey species and with large scale climatic events such as the Last Glacial Maxima (LGM; Cabrera et al. 2022). This study revealed some interesting findings related to changing ecologies of many baleen whales that could have been initiated by the LGM, focusing on the Pleistocene – Holocene transition 1kya – 30kya (Cabrera et al. 2022). The demography history results presented in Cabrera et al. (2022) are strengthened by both consistent findings across species within ocean basins and the use of genotype likelihood to account for uncertainty due to the very low coverage. However, their analyses were limited to the use of short mitochondrial markers for both Northern and southern right whales. In contrast, the nuclear dataset was limited to low-coverage RADseq data for only a few baleen whale species and included the southern right whale, but not the North Atlantic right whale. Additionally, the confidence in the site frequency spectrum (SFS) derived from RADseq data was further limited by the use of folded SFS not exploiting the full range of the SNP frequency spectrum and by low sequencing coverage, both factors which bring important limitations to the resolution that can be achieved with such a data set (Taylor et al. 2021; Mona et al. 2023). Therefore, the application of whole genome sequencing to the North Atlantic right whale provides a powerful opportunity to illustrate the additional conservation insights

WGS can provide. Here we provide estimates of genomic diversity based on high coverage whole genome sequencing of 12 North Atlantic and 10 Southwestern Atlantic southern right whales and estimate the demographic history of each from the time of their divergence to the present. Specifically, we investigated the dynamics of the species isolation process, and their long-term and recent changes in effective population sizes, assessing the impact of whaling on each population.

3.2 Methods

3.2.1 DNA and Library Preparation

The analyzed samples are part of a long-term study on right whales that began before 1980 overseen by the North Atlantic Right Whale Consortium (NARWC). The samples used in this study were collected for other studies between 1988 and 2013, across the contemporary range of North Atlantic right whales in the Northwest Atlantic (n = 12) and off the coast of Argentina (n = 9) and South Georgia (n = 1) for the southern right whale samples (See Table A3.1 for additional collection information). Samples were collected via biopsy sampling (Brown et al. 1991; Palsbøll et al. 1991) under all required regional/federal permits. We used archived DNA from the NARWC's DNA databank housed at Saint Mary's University (Halifax, Nova Scotia) from 12 North Atlantic right

whales and 10 southern right whales that had been previously extracted using standard phenol-chloroform methods (as explained in Wang et al. 2008) and stored at -20°C. A total of 1-5 μg of genomic DNA was sent to the McGill Applied Genomics Innovation Core (Montreal, Quebec, Canada) for PCR-free library preparation using a NxSeq AmpFREE kit (Lucigen). Libraries were evenly distributed and sequenced on three lanes of an Illumina NovaSeq6000 S4 (v1.5, 2 x 150 paired-ends).

3.2.2 Bioinformatic Processing

Near chromosome length reference genome assemblies were downloaded for both the North Atlantic and the southern right whale from DNAZoo (www.dnazoo.org; Dudchenko et al. 2017, 2018) (Table A3.2). We only included scaffolds longer than 1Mbp in our analyses to avoid low quality assembled contigs or scaffolds and mapping biases. After variant calling, the shortest scaffold over 1Mbp in North Atlantic right whales was only 1.12 Mbp in length and contained only 2 variant sites. It was an outlier in all preliminary analyses and was therefore also excluded. The retained scaffolds included 2.17 Gbp out of a total of 2.37 Gbp bases and 2.30 Gbp out of 2.32 Gbp in the North Atlantic and southern right whale reference assemblies respectively. As the sister species to right whales, we chose the bowhead whale as an outgroup for our analyses. We downloaded raw reads as fastq files from the bowhead whale genome assembly from NCBI (SRR1685383; Keane et al. 2015) to be processed alongside our samples.

The bioinformatic pipelines for variant calling and variant filtration with key options and output results are provided in Figure A3.1 and A3.2 respectively. Demultiplexed fastq files were trimmed for low quality bases using TRIMMOMATIC v0.36 (Bolger et al. 2014). We removed residual adapters, leading bases with quality less than 20 and trimmed low quality bases within reads using a sliding window (window size 5 bp : mean base quality Q20). Short (<36 bp) and low-quality reads (average base quality < 30) were dropped. Paired reads were mapped against the reference genome using BWA MEM 0.7.17 (Li & Durbin 2009; Li 2013). In order to make the appropriate comparisons in our analyses downstream, reads from the southern right whales and the bowhead whale were mapped to both North Atlantic and southern right whale reference genomes. Bam files for each sample and lane were created and sorted with SortSam in GATK v4.1.0.0 (McKenna et al. 2010). Bam files across the three lanes for each sample were merged and duplicate reads were marked with GATK4's MergeSamFiles and MarkDuplicates tools respectively. We performed joint variant (SNPs and INDELs) calling in GATK4 using HaplotypeCaller in GVCF mode, GenomicsDBImport and GenotypeGVCFs (Figure A3.1). Using genomic feature repeat files available from DNAZoo, we removed variant sites from repetitive regions in the genome using BEDTOOLS INTERSECT v2.30.0 (Quinlan & Hall 2010). We filtered variants to require individual genotype calls to have a minimum sequencing depth of

10X and genotype quality of 30 using VCFTOOLS v0.1.14 (Danecek et al. 2011). We used BCFTOOLS v1.11 (Danecek et al. 2021) to remove sites missing more than 25% of genotype calls, sites with mapping quality of less than 30 and applied a maximum depth threshold for each dataset that was calculated as the two times the median depth of variant sites summed across all individuals (INFO field "DP" of the VCF file) after repeat masking (NARW alone: 790X, SRW alone: 558X, All on NARW: 1370X, SRW with bowhead: 636X) (Figure A3.2). We identified the sex chromosomes using the D-Genies pipeline (Cabanettes & Klopp 2018) to map our reference genomes to the well annotated blue whale genome assembly (GCA_009873245.3) using minimap2 (Li 2018) and visualized the similarities with a dot-plot. This allowed us to identify the scaffold representing the X-chromosome and confirm the Y-chromosome was excluded when we only retained long scaffolds for our reference genomes. We also confirmed the identity of the X-chromosomes by visualizing read depth across scaffolds for individuals of known sexes where read depth on the X-chromosome was lower than on autosomal scaffolds for males. We then removed variants from the scaffold representing the X-chromosome to include only autosomal variants in our dataset. Subsequent analyses were based on 21 scaffolds in the North Atlantic right whale genome and 20 scaffolds in the southern right whale.

3.2.3 Population genetic analyses

A flowchart of the different analyses conducted in this study with the key options is provided in Figure A3.3. We used KING v2.2.8 (Manichaikul et al. 2010) to identify pairwise kinship coefficients for all samples. As the presence of related individuals violates the assumptions of the analyses we used in this study, we removed from subsequent analyses three North Atlantic right whale and one southern right whale samples displaying a kinship coefficient (φ) larger than 0.177, corresponding to first-degree relatives. Our subsequent analyses were thus performed using nine samples each from North Atlantic and southern right whales.

Variant sites in close physical proximity are often inherited together as haplotypes. Linkage-disequilibrium (LD) decay can be informative for inferring historical population size variation (Nordborg & Tavaré 2002; Park 2012), yet analyses of population structure require unlinked SNPs (Liu et al. 2020). In order to understand the extent of LD in the right whale genomes to identify the best strategy for pruning linked sites, LD decay was calculated with PopLDdecay v3.41 (Zhang et al. 2019) (Figure A3.4). Using PLINK v1.9 (Purcell et al. 2007), we pruned the datasets to remove linked SNPs with a mean r^2 threshold determined for each population (NARW: $r^2 = 0.2$, SRW: $r^2 = 0.1$). Coalescent-based models of demographic history can also be influenced by population structure. While there is only one extant population of North Atlantic right

whale, there are known subpopulations in the southern hemisphere. Most of our southern right whale samples were collected from a known, distinct population near Peninsula Valdés, Argentina, with the exception of one sample that was collected on the feeding grounds in South Georgia where whales aggregate from Argentina and less frequently South Africa (Patenaude et al. 2007; Carroll et al. 2020). To ensure the inclusion of this sample would not influence our results, we estimated population structure and individual genetic ancestry proportions using ADMIXTURE v1.3.0 (Alexander et al. 2009), varying the number of putative genetics clusters (K) from one to five, with ten replicates for each K. We plotted the cross-validation error from our ten independent runs to estimate the most appropriate K value and we used PONG v1.5 (Behr et al. 2016) to assess convergence across ADMIXTURE runs within our samples from each species. We also visualized population structure across species using both species mapped to North Atlantic right whales with a principal component analysis using PLINK v1.9. Finally, we calculated the level of genetic differentiation between species using Weir & Cockerham weighted FST in VFCTOOLS v0.1.14.

To assess levels of genomic diversity, we estimated observed heterozygosity, nucleotide diversity (π), and Watterson's theta (θ) in non-overlapping 10kb windows using SCIKIT-ALLEL v1.3.5 (Miles et al. 2021). Genome-wide measures were calculated as the mean of these statistics across all windows. Observed heterozygosity was

calculated with both right whale populations being mapped to the North Atlantic reference assembly so that the same sites would be considered across both species. We calculate the length of all runs of homozygosity over 10kb and estimated inbreeding coefficients by calculating the proportion of the autosomal genome covered by runs of homozygosity (ROH) of different lengths (FROH) using BCFTOOLS ROH v1.16 (with the option -G 30). We used SCIKIT-ALLEL v1.3.5 and VCFTOOLS v0.1.14 to repeat our estimates of FROH with different restrictions on how ROHs are identified. We used strict parameters in a multinomial HMM model implemented by SCIKIT-ALLEL to identify ROH with the probability of observing a heterozygote in a ROH of 0 (Phet_roh = 0) due to the deep sequencing at hand delivering high confidence in our identified variants. We used default parameters in VCFTOOLS with the --LROH option and report the maximum ROH length estimates with less than two mismatches. We plotted the length distribution of ROHs generated by BCFTOOLS to compare the time frame of potential inbreeding events as long ROHs from recent inbreeding will be broken up by recombination over time (McQuillan et al. 2008).

The lengths of ROHs are a factor of time, but also a factor of biological events such as inbreeding or bottleneck events that can results in long stretches of the genome being identical by descent (Kardos et al. 2017). From Thompson (2013), the time to most common recent ancestor (TMCRA) of ROHs can be estimated as:

$$Time(generations) = \frac{50}{ROHLength(Mb) x recombinationrate(cM/Mb)}$$

We estimate the age of the longest ROH tracts and for the median length of ROH tracts over 10kb for each species using a constant recombination rate of 1 cM/Mb to help understand the timing of historical inbreeding events.

3.2.4 Demographic history

We estimated the divergence dynamics of the species using MSMC-IM (Wang et al. 2020). First, we ran MSMC2 on four North Atlantic right whale haplotypes and four southern right whale haplotypes that were mapped to the North Atlantic right whale genome and on the 16 pairwise comparisons between the two species. We combined these within and between species outputs from MSMC2 and ran MSMC-IM to generate cumulative migration probabilities over time. We repeated these steps to run 10 MSMC-IM replicates using different combinations of samples/haplotypes. We also used MSMC2 v2.1.3 (Wang et al. 2020) to estimate changes in effective population size after the species diverged. We generated a list of accessible genomic regions for variant calling for each species and we removed indels and sites with missing data and phased biallelic SNPs using SHAPEIT2 (Delaneau et al. 2013). We ran 10 iterations of MSMC2 for each species, including a different combination of three samples (six haplotypes) in

each run (therefore the sample from South Georgia was only included in three iterations). We plotted the results using different estimates of mutation rate to assess the robustness of our results to small changes in this parameter (Figure A3.8).

We generated estimates of effective population size over time based on the unfolded SFS in STAIRWAY PLOT v2.1.1 (Liu & Fu 2020). First, we polarized the ancestral vs derived allelic states of the SNP calls for North Atlantic right whale and southern right whales setting the bowhead whale sample as the ancestral allele with a custom python script. We calculated the unfolded SFS of biallelic variant sites without missing data using SCIKIT-ALLEL v1.3.5. We estimated the demographic history of each species using STAIRWAY PLOT v2.1.1 using a mutation rate of 0.9664×10^{-8} mutations/site/generation (estimate for mysticetes reported in Dornburg et al. 2012; and within the range of estimates from Suárez-Menéndez et al. 2022) and generation time of 32 years (based on mean pre-whaling estimates from Taylor et al. 2007).

We estimated recent demographic changes from around four to around 200 generations ago using IBDNe (Browning & Browning 2015). To that aim, we first used IBDSeq (Browning & Browning 2013) to calculate blocks of the genome that are identical by descent (IBD) using default parameters and a maximum LD coefficient between SNPs (r^2 max) of 1.0 to include all variants. The inclusion of all variants was

based on previous studies suggesting that increasing this parameter can improve the quality of the results (Tataru et al. 2014), especially in datasets such as ours where lower values of r^2 max base the analyses on very few variants. We present the results for r^2 max = 1.0 in the main text and include a range of values in the supplementary material (Figure A3.9). We used filtered variant call files for each population alone, pruned to remove sites with missing data. We then used IBDNe (Browning & Browning 2015) to generate population size estimates using a constant recombination rate of 1.0 cM/Mbp and a minimum IBD block length of 2cM. As no recombination map is currently available, we also ran the program with a range of constant recombination rates (0.8-1.2cM/Mbp) to reflect a range of recombination rates reported for mammals (Dumont & Payseur 2008) (Figure A3.9). We also estimated Tajima's D values genome wide in SCIKIT-ALLEL as means to understand how demography is shaping diversity.

3.3 Results

Nearly 7.8 billion raw reads from right whale samples mapped to their reference genomes with a mean sequencing depth coverage of 39.48 ± 15.89 (more information on individual samples can be found in Table A3.1). The number of variant sites called

depended on the filtering parameters required for each test and are broadly summarized in Table 3.1.

Table 3.1. Number of variant sites retained after different filtering regimes. The inclusion of the bowhead whale in the same data set allows for polarization of the SNPs, and analyses of all three species when the same reference genome was required. See Figure A3.1 and A3.2 for filtering regimes used.

	RAW NUMBER OF VARIANTS	AFTER REPEAT MASKING AND FILTERING	ONLY AUTOSOMES AND UNRELATED INDIVIDUALS AFTER FILTERING
NARW alone	5 620 253	1 249 378	1 213 686
SRW alone	21 094 633	6 389 186	6 252 485
NARW, SRW and Bowhead mapped to NARW	39 148 338	13 124 077	12 793 691
SRW WITH BOWHEAD	34 368 321	11 094 247	10 841 437

3.3.1 Population genetic analyses

Both right whale species showed distinct clustering on the PCA (Figure 3.1d) (Weir & Cockerham weighted FST = 0.575). One southern right whale individual showed separation from the others along the second principal component axis, suggesting it may come from a genetically differentiated population, as may have been predicted as it

was sampled from a different location (South Georgia). This subdivision was not detected in the results of ADMIXTURE which has limited power to identify a single genetically distinct sample. The cross-validation error across 10 independent ADMIXTURE runs showed that the lowest values were obtained at K=1. These results suggest that little or no population sub-structure occurs within either set (NARW or SRW) of our samples (Figure A3.7). This was also suggested by the ancestry plots estimated using ADMIXTURE on both right whale species independently which displayed no convergence for K values greater than one (Figure A3.5 & A3.6).

When mapped to their own reference genomes, there were over five times more autosomal variant sites identified in southern right whales (6,252,485) compared to North Atlantic right whales after filtering (1,213,686). North Atlantic right whales exhibited lower genome-wide diversity than southern right whales (mean value from non-overlapping 10kb windows ± SD values – NARW: nucleotide diversity $\pi = 1.85 \times 10^{-4} \pm 3.57 \times 10^{-4}$, Watterson's $\theta = 1.62 \times 10^{-4} \pm 2.57 \times 10^{-4}$, observed heterozygosity = 0.0480 ± 0.0489; SRW: nucleotide diversity $\pi = 8.01 \times 10^{-4} \pm 6.96 \times 10^{-4}$, Watterson's $\theta = 8.12 \times 10^{-4} \pm 6.35 \times 10^{-4}$, observed heterozygosity = 0.257 ± 0.0535) (Figure 3.1a-c). North Atlantic right whales had a larger proportion of their autosomal genome found in ROHs than southern right whales (BCFTOOLS estimates of mean FROH ≥ 1Mb ± SD – NARW: 0.014 ± 0.0078, SRW: 0.0045 ± 0.0078; Figure 3.1e), indicating that North Atlantic right

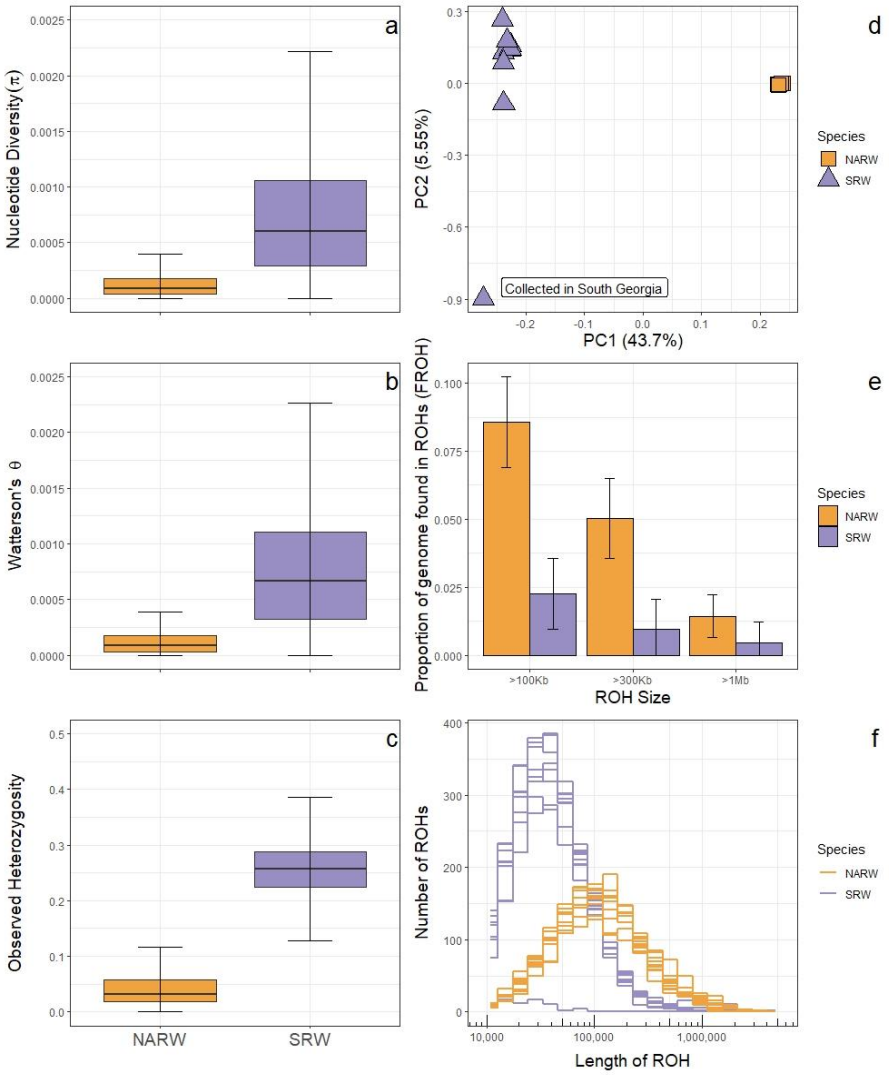

Figure 3.1 Genomic diversity of right whales. a) Nucleotide diversity (π), b) Watterson's θ and c) Observed heterozygosity calculated in 10kb non-overlapping windows across the genome of North Atlantic right whales (left, orange) and Southern right whales (right, purple). Outliers were not plotted beyond 1.5 x IQR (interquartile range) indicated by the bars. For (c) Observed heterozygosity, both species were mapped to the North Atlantic right whale genome to consider the same sites between species. d) Principle component analysis of both right whale species: North Atlantic right whale (orange squares) and Southern right whale (purple triangle). e) Proportion of the autosomal genome (F_{ROH}) of North Atlantic right whale (orange) and southern right whale (purple) found in different ROH size classes estimated with BCFTOOLS. F) The number of ROHs of different lengths in each species for each individual North Atlantic right whale (orange lines) and southern right whale (purple lines) estimated with BCFTOOLS.

whales have higher inbreeding coefficients than southern right whales from the Southwest Atlantic. These results were relatively consistent across different methods (Table A3.3). The length of ROHs also differed between populations with our southern right whales having shorter ROHs compared to the greater abundance of long ROHs found in North Atlantic right whales (Figure 3.1f).

Using the median value of ROH >10kb (NARW: 105,976 bp, SRW: 36,883 bp), we estimate that most ROHs were formed in North Atlantic right whales approximately 472 generations ago (~15,100 years ago based on a generation time of 32 years as estimated by Taylor (2007)) and 1355 generations (~43,360 years) ago in Southwest Atlantic southern right whales. The longest ROHs in both species were approximately 2-4Mb long and were likely formed 12.5-25 generations (400-800 years) ago.

3.3.2 Demographic History

The cumulative migration probabilities (M(t)) between North Atlantic and southern right whales estimated with MSMC-IM identified the putative split between the two species as occurring approximately between 124 and 480kya, when the M(t) dropped below 0.5 (Figure 3.2a). Visually, we confirmed this putative split time with the effective population size (Ne) estimates from STAIRWAY PLOT2 converging around 200-300kya (Figure 3.3).

Following divergence, MSMC2 reported a decrease in Ne in North Atlantic right whales that plateaued around 2000ya and remained relatively constant until this method loses resolution a few thousand years ago (Figure 3.2b). STAIRWAY PLOT2 identified a similar pattern of post-divergence decline, followed by a plateau of Ne values ca. 2000 between 100kya and 500ya (Figure 3.3). The initial post divergence decline identified by these methods could also be indicative of emerging population structure between the eastern and western North Atlantic. In the most recent times (<500ya), Ne kept decreasing in two steps: a first one occurring between 500ya and 200ya with a Ne contraction from approximately 2000 to 500, and a second one, though associated with more uncertainty as shown by the large confidence intervals, indicating an additional five-fold contraction ~50ya. In southern right whales, after the species diverged both methods estimate that effective population size remained much higher than in the North Atlantic right whale. MSMC2 estimated a slight increase in effective population size between 100,000 and 10,000 years ago whereas STAIRWAY PLOT2 identified a steadier decline in the population in that time frame (Figures 3.2b & 3.3). The broad patterns in these results were found across a range of mutation rates used in MSMC2 (Figure A3.8).

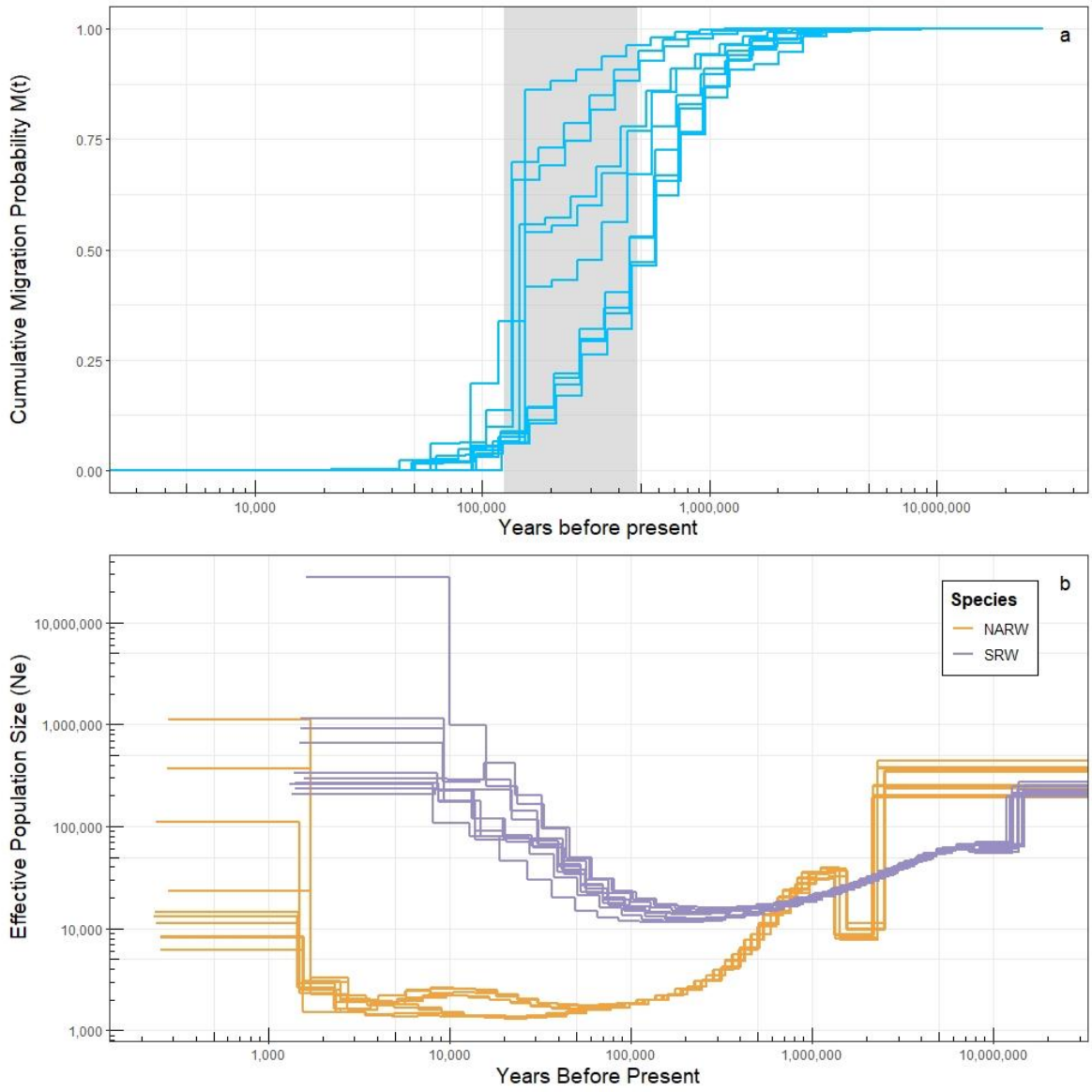

Figure 3.2 Demographic reconstruction from MSMC2. a) Ten iterations of MSMC-IM estimating the cumulative migration probability M(t) between North Atlantic and southern right whale over time. Estimated divergence time identified by the shaded area as M(t) dropping below 0.5 as suggested by Wang et al. (2020) and Schiffels & Wang (2020). b) Ten estimates of effective population size through time for each North Atlantic (orange) and southern right whale (purple) generated with MSMC2.

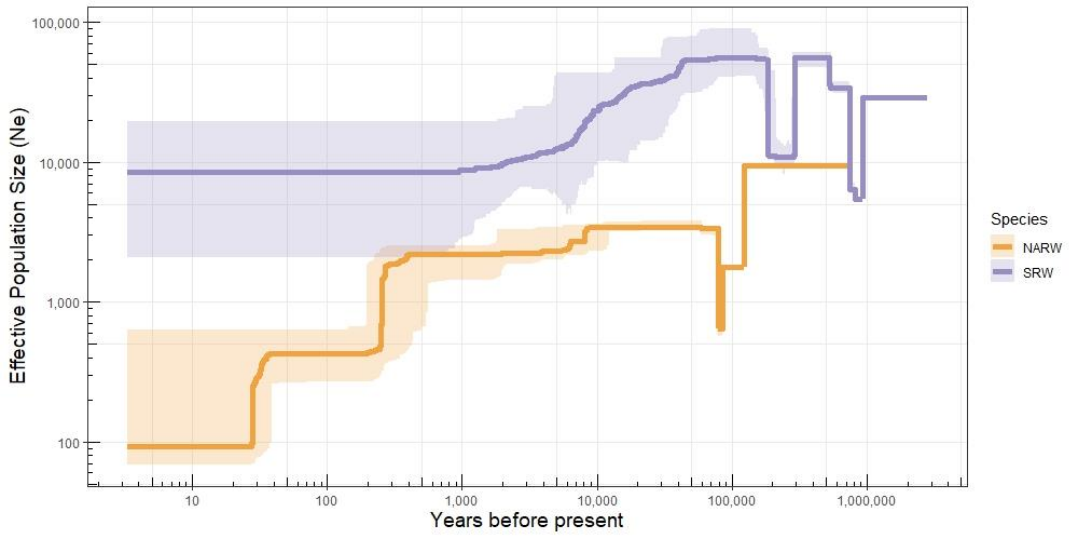

Figure 3.3 Demographic reconstruction from STAIRWAY PLOT2. Estimates of effective population size of North Atlantic (orange) and southern right whales (purple) through time based on the site frequency spectrum in STAIRWAY PLOT2. The shaded regions represent the 2.5%-97.5% confidence limits for 200 estimates.

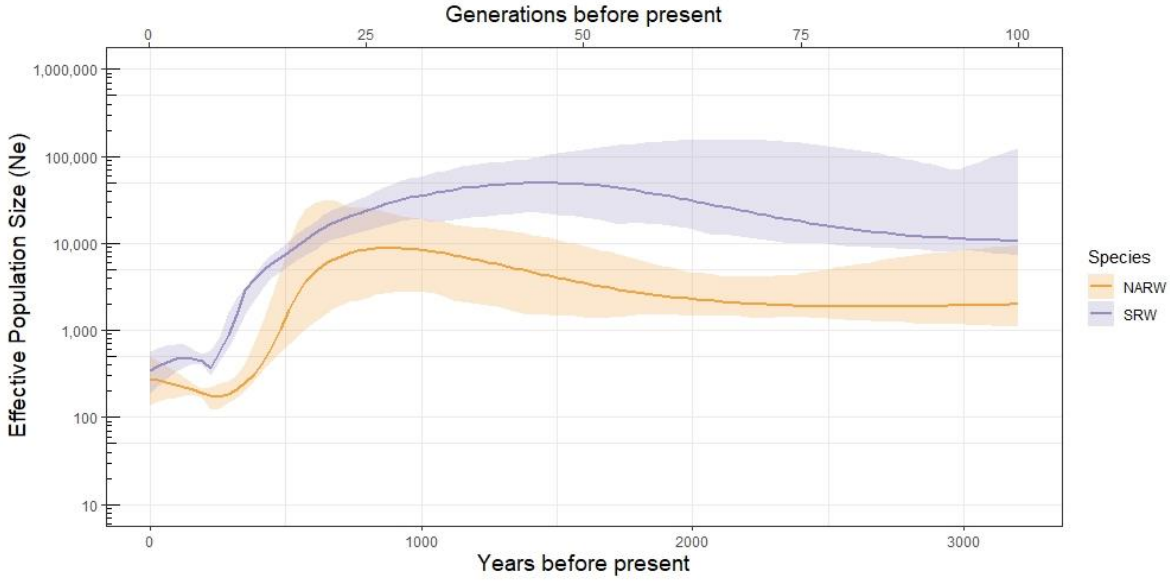

Figure 3.4 Demographic reconstruction from IBDNe. Effective population size in recent history as estimated by IBDNe for North Atlantic (orange) and southern right whale (purple) using a constant recombination rate of 1.0 cm/Mb and a generation time of 32 years. 95% confidence intervals are depicted by the shaded areas.

The most contemporary estimates of effective population size were best captured by IBDNe (Figure 3.4). From 3kya to roughly 1kya, IBDNe identified relatively stable effective population size in North Atlantic right whales of ~1000 and of Southwest Atlantic southern right whales at >10,000. Within the last 1000 years, a steep decline was detected in both species lasting until approximately 225 years ago, when the effective population size of North Atlantic right whales was below 200 and southern right whales was around 350. Both populations began to show growth in Ne following this bottleneck event. A slight decline was detected within the last 100 years (3 generations) in Southwest Atlantic southern right whales; however, this should be interpreted with caution as the methods used have much lower resolution in this recent time frame (Browning & Browning 2015). These general patterns with a decline in Ne until ~225 years ago, followed by increasing Ne in more recent years were consistent across different r^2 max values and recombination rates (Figure A3.9). The mean Tajima's D value was slightly positive in North Atlantic right whales (0.298 ± 1.209) and slightly negative in our southern right whales (-0.150 ± 0.816) potentially reflecting these differences in their demographic histories.

3.4 Discussion

Using whole genome sequence data from nine unrelated individuals from both North Atlantic and southern right whales from the Southwest Atlantic, we found lower levels of genetic diversity and higher inbreeding coefficients in North Atlantic right whales compared to southern right whales. We also described the changes in effective population size over time in each of these species since their divergence. We estimate a divergence time approximately 124-480kya, which is more recent than has been described in other studies (McGowen et al. 2009: 95% HPD 0.34-1.28 MYA; McGowen et al., 2020: 6-partition AR model 95% CI 3.65-5.11 MYA, 6-partition IR model 95% CI 1.67-2.11 MYA; Slater et al. 2017: 95% HPD 0.45-1.43 MYA). This slightly more recent split estimate is most likely related to the population genetic analytical framework we have used in this study using MSMC-IM and to the different mutation rates used here compared to McGowen et al. (2020). Our analytical approach contrasts with more 'classic' phylogenetic approaches by modelling patterns of lineage sorting under the cross coalescent rate (CCR) among species from which the rate of cessation of gene-flow can be estimated. The recent split time estimated by the CCR decay of MSMC-IM could reflect potential secondary contact that may have occurred between North and southern Atlantic right whales during the Quaternary glaciations. Furthermore, while our study used pedigree-based mutation rate estimates, the phylogenetic approaches of McGowen

et al. (2020) used a relaxed molecular clock calibrated with fossil records to estimate divergence times. While both approaches are valid in their respective statistical framework, they usually lead to one or two order magnitude differences in split time estimates (Howell et al. 2003; Ho, Shapiro, et al. 2007; Henn et al. 2008). This discrepancy between the CCR decay approach implemented in MSMC-IM and more traditional phylogenetic approaches comes from the fact that phylogenetic approaches as used in McGowen et al. (2020) rely on estimating substitution rate (complete nucleotide replacement), ignoring within species polymorphic variation, and the inter-loci gene-tree discordance due to incomplete lineage sorting and gene flow. Recently developed phylogenetic approaches based on the multi-species coalescent now allow incorporating these properties to better estimate species diversification processes (Solís-Lemus et al. 2016; Rannala & Yang 2017; Müller et al. 2017, 2018; Flouri et al. 2018; Jiao et al. 2020, 2021). While not the primary focus of this study, understanding the timing and rate associated with how the species diverged with dedicated approaches warrants further investigation specific to the *Eubalaena* and could benefit from the inclusion of additional samples through the range of the southern right whale.

After diverging, southern right whales had effective population sizes nearly ten times greater than North Atlantic right whales for thousands of years. We also detected a rapid and recent decline in both populations over the past thousand years, as expected

based on the history of whaling (Aguilar 1986; Reeves et al. 1999; Rastogi et al. 2004). The recent decline in Southwest Atlantic southern right whales was only detected by the most sensitive analysis of IBDNe relying on IBD tract lengths and was not detected based on site frequency spectrum data used by STAIRWAY PLOT2. This analysis suggests the bottleneck occurred more recently than in the North Atlantic, as predicted by the onset of industrial whaling beginning a couple centuries later into the 1700s (International Whaling Commission 2013).

The history of whaling over the past thousand years is clearly imprinted in the genomic diversity of both North Atlantic and southern right whales. The longest ROHs identified for both species were likely formed 400 to 800 years ago as effective population sizes drastically declined (see discussion below how parameter selection can lead to a small discrepancy with the timing of historical events). While most European and American whaling ended in the mid 1900s (Reeves et al. 2007; Reeves & Smith 2007), illegal Soviet whaling continued in the Southern Oceans and over 3300 southern right whales were estimated to have been killed between 1951-1971 (Tormosov et al. 1998). This spike in whaling effort might explain a small dip in the effective population size of southern right whales from the Southwest Atlantic in the past few generations. However, some caution is warranted with this interpretation as the power to detect

changes in effective population size during this time frame is low (Browning & Browning 2015).

Our results showed slightly different patterns of effective population size over time generated by MSMC2 and STAIRWAY PLOT2 between 10-100 KYA in southern right whales. Discrepancies produced by these methods could be explained by gene flow from other parts of their range and/or could provide some evidence of selection (Johri et al. 2021; Boitard et al. 2022). Historical population size estimates can be biased by population structure as it violates the panmictic assumptions of these models (Ho & Shapiro 2011; Heller et al. 2013). In our population structure analysis, one southern right whale sample was slightly differentiated from the others and was collected from a different location (on the feeding grounds of South Georgia) where individuals from different breeding populations aggregate (Patenaude et al. 2007; Carroll et al. 2020). Therefore, we repeated our analyses of IBDNe and STAIRWAY PLOT2 excluding this individual (Figs. A3.10 & A3.11) and confirmed that its inclusion had little impact on our inferences of demographic history. The timing of the discrepancy between MSMC2 and STAIRWAY PLOT2 also corresponds with the age of median ROH length in southern right whales (~43,360 years ago) adding additional evidence for interesting dynamics during this time frame. The extent of selection, both historical and contemporary in southern right whales from the Southwest Atlantic remains unknown.

North Atlantic right whales have been extirpated from the Northeast Atlantic and only a single extant population remains in the Northwest Atlantic. Estimates of historical effective population size of the North Atlantic right whale population could therefore be inflated due to past gene flow with the unsampled population from the Northeast Atlantic – or in contrast underestimate past effective population size of the species as a whole. Likewise, gene flow from an unsampled population of southern right whales (Carroll et al. 2019, 2020) could inflate our estimates of diversity and effective population size and therefore a precautionary approach should be taken in using these estimates for conservation planning. The potential for current and future gene flow, the relatively short nature of the bottleneck event, recent increasing population sizes (Crespo et al. 2019), the lower inbreeding coefficient, and the excess of rare variants in the Southwest Atlantic right whale population are all promising signs for the resiliency and recovery of this population.

Small population sizes have a number of consequences: 1) low population sizes can erode genetic diversity (Wright 1931; Allendorf 1986), 2) reduced mating opportunities increase the inbreeding coefficient as average pairwise relatedness between individuals increases (Wright 1922; Charlesworth & Charlesworth 1987; Charlesworth & Willis 2009) and 3) long-term small population size can lead to long tracks of homozygosity, exposing deleterious alleles and thus reducing the strength of

purifying selection thereby altering the genetic load in a population (Frankham 1995a, 1999; Keller & Waller 2002; Grossen et al. 2020). North Atlantic right whales possess extremely low genomic diversity – lower than the cheetah (Dobrynin et al. 2015), on par with the Iberian lynx (Westbury et al. 2018) and one of the lowest nucleotide diversity values recorded for cetaceans (Morin et al. 2021) (Figure 3.5, Table A3.4). We also see evidence of both long-term and recent inbreeding through the high proportion of the genome found in ROHs (FROH) and a higher proportion of long ROHs indicative of recent inbreeding events where not enough time has passed to allow recombination to break them up. The extent of historical and/or contemporary purifying or balancing selection in the North Atlantic right whale is unknown and remains an important question for future study. Therefore, the cumulative effects of long-term small effective population size of North Atlantic right whales are yet to be determined.

The estimates of demographic history are based on coalescent models using site frequency spectrum or runs of homozygosity and are influenced not only by the data themselves, but also by factors such as generation time, mutation rate, recombination rate, and selection. An inherent assumption of many biological models is non-overlapping generations, which is violated for most mammalian species. In right whales, a female can be reproductively active for decades (Hamilton et al. 1998) and

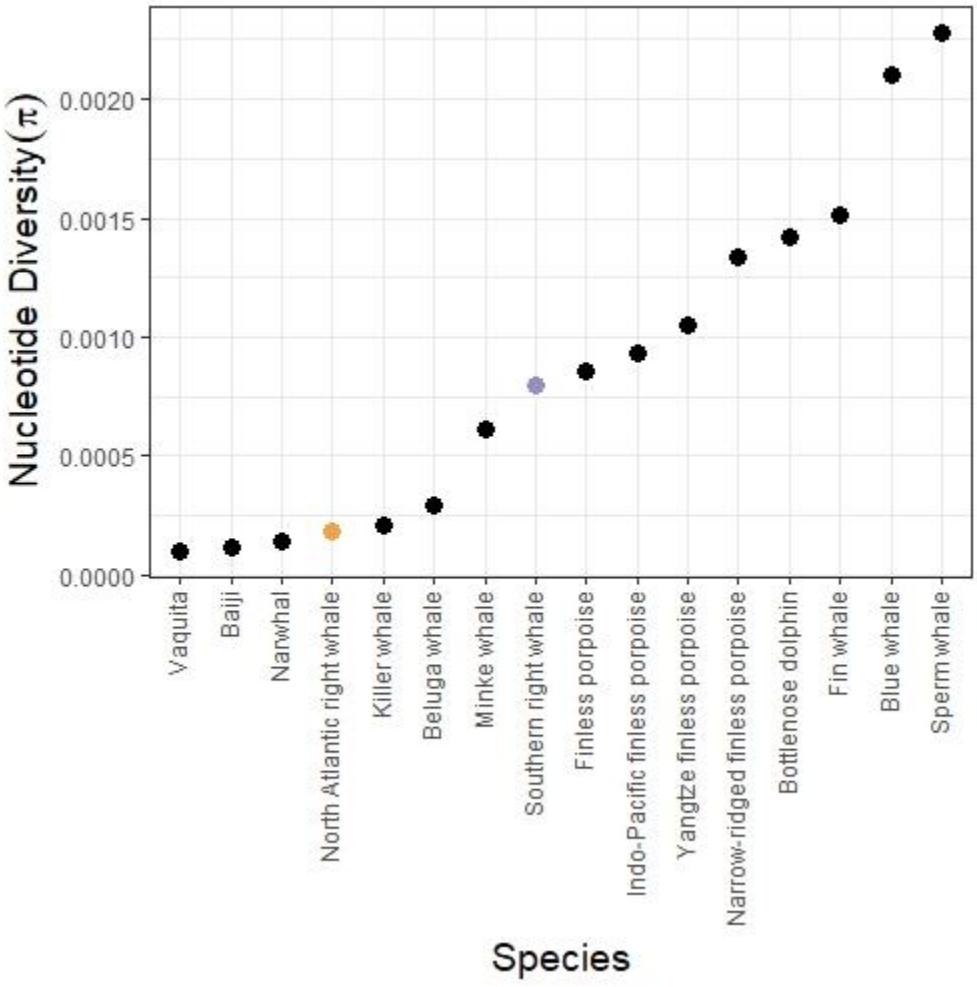

Figure 3.5 Nucleotide diversity reported for other cetacean species based on genomic datasets. Raw data and sources for the non-right whale species can be found in Table A3.4.

therefore be calving at the same time as her granddaughters. This assumption is largely unavoidable with long-lived mammalian studies but should be recognized as a potential source of bias where Ne is likely underestimated and bottlenecks may be detected earlier than they occurred (Waples et al. 2014; Charbonnel et al. 2022).

Mutation rate is used in estimating changes in effective population sizes over time. Estimates of mutation rate based on phylogenies have been generated for many species including baleen whales (Jackson et al. 2009; Dornburg et al. 2012), but these don't account for within-species variation and therefore likely underestimate the true mutation rate. Pedigree based estimates of mutation rate, or those based on ancient DNA should therefore be more accurate (Ho, Kolokotronis, et al. 2007; Ho et al. 2008), but these are still lacking for many species. The mutation rate used in this study is similar to those reported for other mammals based on pedigree estimates, including closely related whale species (Suárez-Menéndez et al. 2022) and the patterns in our results are fairly robust to small changes in mutation rate (Figure A3.8). Recombination rate will help dictate the break-up of IBD blocks over time. In this study, we assumed a constant recombination rate across the genome, which is inherently incorrect. Our results, however, are quite robust to a range of recombination rates expected for mammals (Figure A3.9) and therefore, while a variable recombination rate map would be more accurate, no recombination maps for right whales or related species were available and we do not feel it would alter the broad patterns found in our results. Finally, accounting for potential effects of selection is difficult, and while we may have some evidence of historical selection in southern right whales as indicated by a discrepancy of population trends from different methods, the patterns in our results –

especially in more recent timescales, remain the same. Parameter choice is extremely important in model-based studies and can create many biases in results, however we are fairly confident we mitigated these biases as best we could and we caution against over interpretation of exact timing and population sizes, and rather encourage focusing on the broader trends presented by the data.

In addition to mitigating biases, our study design provides a rich and highly informative genomic dataset capable of tackling questions regarding the impact of whaling that were unable to be addressed in previous work (see Taylor et al. 2021). We used high coverage, genome wide data, polarized SFSs, identity-by-descent tracts, and runs of homozygosity to both describe diversity and investigate changes in effective population size, with fine scale resolution thanks to the use of dedicated methods to analyze demographic changes from the times of divergence to the near present (~4 generations ago). Whole genome datasets and the increasing availability of reference genomes are opening doors to help answer new and long-standing questions important for conservation (Formenti et al. 2022; Theissinger et al. 2023).

When tackling real-world conservation issues and trying to save a species from the brink of extinction, there are often obvious (while extremely difficult to achieve) goals including stopping habitat destruction, preventing anthropogenic mortalities, and

protecting critical habitat, among others. What is less clear however, is when do we consider a species or population to be "recovered". The identification of clear and specific goals in conservation is often a challenge. Genomic data is proving to be an invaluable tool for conservation management providing more accurate measures of genetic diversity, a better understanding of the genetics of small populations and estimates of demographic history of wildlife populations (Taylor et al. 2021; Kardos et al. 2021; Formenti et al. 2022; Grueber & Sunnucks 2022; Paez et al. 2022; Theissinger et al. 2023). Understanding historical effective population sizes and how they have changed over different time scales can identify onset of anthropogenic impacts and can help determine what a stable effective population size for a species might be. We can use genomic data to identify long-term small population sizes, and subsequently predict the impact of inbreeding and/or genetic load on fitness and population viability (i.e. Dussex et al. 2021; Robinson et al. 2022). When comparing populations or species, such as between right whales in this study, these historical effective population sizes should not be the target for census size themselves but can help in setting different expectations of long-term recovery for different populations or species. The vastly different effective population sizes between extant North Atlantic and Southwest Atlantic southern right whales throughout much of their history identified in our results, paired with the differences in reproductive output and inbreeding coefficients

suggest the recovery potential of these two populations may be very different and therefore using this southern right whale population as a benchmark for recovery in North Atlantic right whales is likely not appropriate.

Limiting anthropogenic mortality is essential to the conservation of North Atlantic right whales, but we can also now begin to understand how the genomic landscape will affect the subsequent population recovery. Although the combination of long-term small Ne, low levels of diversity, and recent inbreeding put North Atlantic right whales in a situation where genetic factors are likely to be impacting individual fitness and species recovery; the situation may not be as bleak as it initially sounds. The field of conservation biology has numerous success stories of species brought to very small numbers and low levels of genetic diversity, but which were still able to recover. Some examples include the northern elephant seal (Hoelzel 1999) and the Channel Island foxes (Robinson et al. 2018), the Chatham Island black robin (von Seth et al. 2022) and the kākāpō (Dussex et al. 2021). Therefore, although quantifying the impacts of genetic characteristics (including genetic load) on individual fitness and species recovery are important next steps we are carrying out, the overall forecast for the species should still be positive, particularly given the fact that the population has increased in the past – although at a modest rate – during periods of lower mortality (e.g. Pace et al. 2017).

3.5 Data Accessibility Statement

Aligned sequence reads have been deposited to the NCBI Sequence Read Archive under accession numbers SRR22863734-SRR22863755 (BioProject: PRJNA914998). Code used in these analyses has been made publicly available: https://github.com/carlacrossman/RightWhale_WGS

3.6 Benefit Sharing Statement

These long-term sample collections have been made possible by numerous collaborations including participation from a number of regional organizations. Specifically, we draw attention to those individuals and organizations acknowledged in Payne et al (1986), and Schaeff et al (1997) for their contributions to southern right whale sample collection and recognize the number of organizations that contribute information from their region to the North Atlantic Right Whale Consortium (https://www.narwc.org). All samples were collected with the required local permits and imported/exported with CITES permits as required. The hope is that these data improve our understanding of the biology of these species, and will be used to improve conservation actions in both locations.

Just as our study species disperse across international borders, we strive to make all aspects of our project available to all stakeholders. All code, scripts and data used in this study have been made publicly available. We particularly hope that our colleagues currently studying southern right whales in Argentina (and elsewhere in the southern hemisphere) benefit from these data and can incorporate them into their ongoing analyses.

3.7 References

Aguilar A (1986) A review of old Basque whaling and its effect on the right whales (*Eubalaena glacialis*) of the North Atlantic. Reports of the International Whaling Commission, Special Issue, 191–199.

Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. Genome Research, 19, 1655–1664.

Allendorf FW (1986) Genetic drift and the loss of alleles versus heterozygosity. Zoo Biology, 5, 181–190.

Allendorf FW, Hohenlohe PA, Luikart G (2010) Genomics and the future of conservation genetics. Nature Reviews Genetics, 11, 697–709.

Baker CS, Patenaude NJ, Bannister JL, Robins J, Kato H (1999) Distribution and diversity of mtDNA lineages among southern right whales (*Eubalaena australis*) from Australia and New Zealand. Marine Biology, 134, 1–7.

Behr AA, Liu KZ, Liu-Fang G, Nakka P, Ramachandran S (2016) Pong: Fast analysis and visualization of latent clusters in population genetic data. Bioinformatics, 32, 2817–2823.

Best PB, Brandao A, Butterworth DS (2001) Demographic parameters of southern right whales off South Africa. Journal of Cetacean Research & Management Special Issue., 2, 161–169.

Boitard S, Arredondo A, Chikhi L, Mazet O (2022) Heterogeneity in effective size across the genome: effects on the inverse instantaneous coalescence rate (IICR) and implications for demographic inference under linked selection. Genetics, 220, iyac008.

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics, 30, 2114–2120.

Braham HW, Rice DW (1984) The Right whale, *Balaena glacialis*. Marine Fisheries Review, 46, 38–44.

Brown MW, Kraus SD, Gaskin DE (1991) Reaction of North Atlantic right whales (*Eubalaena glacialis*) to skin biopsy sampling for genetic and pollutant analysis. Reports of the International Whaling Commission, Special Is, 81–89.

Browning BL, Browning SR (2013) Detecting identity by descent and estimating genotype error rates in sequence data. American Journal of Human Genetics, 93, 840–851.

Browning SR, Browning BL (2015) Accurate Non-parametric Estimation of Recent Effective Population Size from Segments of Identity by Descent. American Journal of Human Genetics, 97, 404–418.

Cabanettes F, Klopp C (2018) D-GENIES: Dot plot large genomes in an interactive, efficient and simple way. PeerJ, 2018.

Cabrera AA, Schall E, Bérubé M, et al. (2022) Strong and lasting impacts of past global warming on baleen whales and their prey. Global Change Biology, 28, 2657–2677.

Carroll EL, Alderman R, Bannister JL, et al. (2019) Incorporating non-equilibrium dynamics into demographic history inferences of a migratory marine species. Heredity, 122, 53–68.

Carroll EL, Ott PH, McMillan LF, et al. (2020) Genetic Diversity and Connectivity of Southern Right Whales (*Eubalaena australis*) Found in the Brazil and Chile-Peru

Wintering Grounds and the South Georgia (Islas Georgias del Sur) Feeding Ground. Journal of Heredity, 111, 263–276.

Charbonnel E, Daguin-Thiébaut C, Caradec L, et al. (2022) Searching for genetic evidence of demographic decline in an arctic seabird: beware of overlapping generations. Heredity, 128, 364–376.

Charlesworth D, Charlesworth B (1987) Inbreeding Depression and its Evolutionary Consequences. Annual Review of Ecology and Systematics, 18, 237–268.

Charlesworth D, Willis JH (2009) The genetics of inbreeding depression. Nature Reviews Genetics, 10, 783–796.

Cooke JG (2018) *Eubalaena australis* (Chile-Peru subpopulation). The IUCN Red List of Threatened Species, e.T133704A.

Cooke JG, Zerbini AN (2018) *Eubalaena australis*. The IUCN Red List of Threatened Species, e.T8153A50354147.

Crespo EA, Pedraza SN, Dans SL, et al. (2019) The southwestern Atlantic southern right whale, *Eubalaena australis*, population is growing but at a decelerated rate. Marine Mammal Science, 35, 93–107.

Cumbaa SL (1986) Archaeological evidence of the 16th century Basque right whale fishery in Labrador. Report of the International Whaling Commission, SI 10, 371–373.

Danecek P, Auton A, Abecasis G, et al. (2011) The variant call format and VCFtools. Bioinformatics, 27, 2156–2158.

Danecek P, Bonfield JK, Liddle J, et al. (2021) Twelve years of SAMtools and BCFtools. GigaScience, 10, 1–4.

Delaneau O, Zagury JF, Marchini J (2013) Improved whole-chromosome phasing for disease and population genetic studies. Nature Methods, 10, 5–6.

Dobrynin P, Liu S, Tamazian G, et al. (2015) Genomic legacy of the African cheetah, *Acinonyx jubatus*. Genome Biology, 16, 277.

Dornburg A, Brandley MC, McGowen MR, Near TJ (2012) Relaxed clocks and inferences of heterogeneous patterns of nucleotide substitution and divergence time

estimates across whales and dolphins (Mammalia: Cetacea). Molecular Biology and Evolution, 29, 721–736.

Dudchenko O, Batra SS, Omer AD, et al. (2017) De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. Science, 356, 92–95.

Dudchenko O, Shamim MS, Batra SS, et al. (2018) The Juicebox Assembly Tools module facilitates de novo assembly of mammalian genomes with chromosome-length scaffolds for under $1000. bioRxiv, 254797.

Dumont BL, Payseur BA (2008) Evolution of the genomic rate of recombination in mammals. Evolution, 62, 276–294.

Duntsch L, Whibley A, Brekke P, Ewen JG, Santure AW (2021) Genomic data of different resolutions reveal consistent inbreeding estimates but contrasting homozygosity landscapes for the threatened Aotearoa New Zealand hihi. Molecular Ecology, 30, 6006–6020.

Dussex N, van der Valk T, Morales HE, et al. (2021) Population genomics of the critically endangered kākāpō. Cell Genomics, 1, 100002.

Flouri T, Jiao X, Rannala B, Yang Z (2018) Species Tree Inference with BPP Using Genomic Sequences and the Multispecies Coalescent (AD Yoder, Ed,). Molecular Biology and Evolution, 35, 2585–2593.

Foote AD, Hooper R, Alexander A, et al. (2021) Runs of homozygosity in killer whale genomes provide a global record of demographic histories. Molecular Ecology, 30, 6162–6177.

Formenti G, Theissinger K, Fernandes C, et al. (2022) The era of reference genomes in conservation genomics. Trends in Ecology and Evolution, 37, 197–202.

Frankel OH (1974) Genetic conservation: our evolutionary responsibility. Genetics, 78, 53–65.

Frankham R (1995a) Inbreeding and Extinction: A Threshold Effect. Conservation Biology, 9, 792–799.

Frankham R (1995b) Conservation Genetics. Annual Review of Genetics, 29, 305–327.

Frankham R (1999) Quantitative genetics in conservation biology. Genetical Research, 74, 237–244.

Frasier TR, Gillett RM, Hamilton PK, et al. (2013) Postcopulatory selection for dissimilar gametes maintains heterozygosity in the endangered North Atlantic right whale. Ecology and Evolution, 3, 3483–3494.

Frasier TR, McLeod BA, Gillett RM, Brown MW, White BN (2007) Right whales past and present as revealed by their genes. In: The urban whale: North Atlantic right whales at the crossroads, pp. 200–231. Harvard University Press, Cambridge, MA.

Gaskin DE (1991) An Update on the Status of the Right Whale, *Eubalaena Glacialis*, in Canada. Canadian Field-Naturalist, 105, 198–205.

Grossen C, Guillaume F, Keller LF, Croll D (2020) Purging of highly deleterious mutations through severe bottlenecks in Alpine ibex. Nature Communications, 11, 1001.

Grueber CE, Sunnucks P (2022) Using genomics to fight extinction. Science, 376, 574–575.

Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD (2009) Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. PLoS Genetics, 5, e1000695.

Hakamada T, Matsuoka K (2016) The number of blue, fin, humpback, and North Pacific right whales in the western North Pacific in the JARPNII Offshore survey area. Paper SC/F16/JR13 presented to the Expert Panel Workshop of the Final Review on the Western North Pacific Japanese Special Permit Programme (JARPN II), 22-26 February 2016, Tokyo, Japan (unpublished)., 14pp.

Hamilton P, Knowlton A, Marx M, Kraus S (1998) Age structure and longevity in North Atlantic right whales *Eubalaena glacialis* and their relation to reproduction. Marine Ecology Progress Series, 171, 285–292.

Harcourt R, van der Hoop J, Kraus S, Carroll EL (2019) Future directions in *Eubalaena spp.*: Comparative research to inform conservation. Frontiers in Marine Science, 6, 1–25.

Heller R, Chikhi L, Siegismund HR (2013) The Confounding Effect of Population Structure on Bayesian Skyline Plot Inferences of Demographic History. PLoS ONE, 8, e62992.

Henn BM, Gignoux CR, Feldman MW, Mountain JL (2008) Characterizing the Time Dependency of Human Mitochondrial DNA Mutation Rate Estimates. Molecular Biology and Evolution, 26, 217–230.

Ho SY., Kolokotronis S-O, Allaby RG (2007) Elevated substitution rates estimated from ancient DNA sequences. Biology Letters, 3, 702–705.

Ho SYW, Saarma U, Barnett R, Haile J, Shapiro B (2008) The effect of inappropriate calibration: three case studies in molecular ecology (P Bennett, Ed,). PLoS ONE, 3, e1615.

Ho SYW, Shapiro B (2011) Skyline-plot methods for estimating demographic history from nucleotide sequences. Molecular Ecology Resources, 11, 423–434.

Ho SYW, Shapiro B, Phillips MJ, Cooper A, Drummond AJ (2007) Evidence for Time Dependency of Molecular Rate Estimates (J Sullivan, Ed,). Systematic Biology, 56, 515–522.

Hoelzel AR (1999) Impact of population bottlenecks on genetic variation and the importance of life-history; a case study of the northern elephant seal. Biological Journal of the Linnean Society., 68, 23–39.

Howell N, Smejkal CB, Mackey DA, et al. (2003) The Pedigree Rate of Sequence Divergence in the Human Mitochondrial Genome: There Is a Difference Between Phylogenetic and Pedigree Rates. The American Journal of Human Genetics, 72, 659–670.

International Whaling Commission (2013) Report of the IWC Workshop on the Assessment of Southern Right Whales. Journal of Cetacean Research & Management, 14, 439–462.

IUCN (2022) The IUCN Red List of Threatened Species. Version 2022-1.

IWC (2001) Report of the Workshop on the Comprehensive Assessment of Right Whales: A Worldwide Comparison. Journal of Cetacean Research and Management (Special Issue), 1–35.

Jackson JA, Baker CS, Vant M, et al. (2009) Big and slow: Phylogenetic estimates of molecular evolution in Baleen whales (Suborder Mysticeti). Molecular Biology and Evolution, 26, 2427–2440.

Jiao X, Flouri T, Rannala B, Yang Z (2020) The Impact of Cross-Species Gene Flow on Species Tree Estimation (P Foster, Ed,). Systematic Biology, 69, 830–847.

Jiao X, Flouri T, Yang Z (2021) Multispecies coalescent and its applications to infer species phylogenies and cross-species gene flow. National Science Review, 8, 1–16.

Johri P, Riall K, Becher H, et al. (2021) The Impact of Purifying and Background Selection on the Inference of Population History: Problems and Prospects. Molecular Biology and Evolution, 38, 2986–3003.

Kardos M, Armstrong EE, Fitzpatrick SW, et al. (2021) The crucial role of genome-wide genetic variation in conservation. Proceedings of the National Academy of Sciences, 118, e2104642118.

Kardos M, Qvarnström A, Ellegren H (2017) Inferring individual inbreeding and demographic history from segments of identity by descent in *Ficedula* flycatcher genome sequences. Genetics, 205, 1319–1334.

Keane M, Semeiks J, Webb AE, et al. (2015) Insights into the evolution of longevity from the bowhead whale genome. Cell Reports, 10, 112–122.

Keller LF, Waller DM (2002) Inbreeding effects in wild populations. Trends in Ecology and Evolution, 17, 230–241.

Knowlton AR, Kraus SD, Kenney RD (1994) Reproduction in North Atlantic right whales (*Eubalaena glacialis*). Canadian Journal of Zoology, 72, 1297–1305.

Kraus SD, Pace III RM, Frasier TR (2007) High investment, low return: The strange case of reproduction in *Eubalaena glacialis*. In: The urban whale: North Atlantic right whales at the crossroads (eds Kraus SD, Rolland RM), pp. 172–199. Harvard University Press, Cambridge, MA.

Li H (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv, 00, 1–3.

Li H (2018) Minimap2: Pairwise alignment for nucleotide sequences. Bioinformatics, 34, 3094–3100.

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics, 25, 1754–1760.

Liu X, Fu YX (2020) Stairway Plot 2: demographic history inference with folded SNP frequency spectra. Genome Biology, 21, 1–9.

Liu C-C, Shringarpure S, Lange K, Novembre J (2020) Exploring Population Structure with Admixture Models and Principal Component Analysis. In: Statistical Population Genomics, pp. 67–86.

Lockyer C (1976) Body weights of some species of large whales. ICES Journal of Marine Science, 36, 259–273.

Malik S, Brown MW, Kraus SD, White BN (2000) Analysis of mitochondrial DNA diversity within and between North and South Atlantic right whales. Marine Mammal Science, 16, 545–558.

Manichaikul A, Mychaleckyj JC, Rich SS, et al. (2010) Robust relationship inference in genome-wide association studies. Bioinformatics, 26, 2867–2873.

McGowen MR, Spaulding M, Gatesy J (2009) Divergence date estimation and a comprehensive molecular tree of extant cetaceans. Molecular Phylogenetics and Evolution, 53, 891–906.

McGowen MR, Tsagkogeorga G, Álvarez-Carretero S, et al. (2020) Phylogenomic Resolution of the Cetacean Tree of Life Using Target Sequence Capture. Systematic Biology, 69, 479–501.

McKenna A, Hanna M, Banks E, et al. (2010) The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Research, 20, 1297–1303.

McLeod BA, Brown MW, Frasier TR, White BN (2010) DNA profile of a sixteenth century western North Atlantic right whale (*Eubalaena glacialis*). Conservation Genetics, 11, 339–345.

McLeod BA, Brown MW, Moore MJ, et al. (2008) Bowhead whales, and not right whales, were the primary target of 16th- to 17th-century Basque whalers in the western North Atlantic. Arctic, 61, 61–75.

McQuillan R, Leutenegger AL, Abdel-Rahman R, et al. (2008) Runs of Homozygosity in European Populations. American Journal of Human Genetics, 83, 359–372.

Miles A, Pyup.io bot, Murillo R, et al. (2021) Cggh/Scikit-Allel: v1. 3.3., DOI: 10.5281/zenodo.4759368.

Mona S, Benazzo A, Delrieu-trottin E, Lesturgie P (2023) Population genetics using low coverage RADseq data in non-model organisms: biases and solutions. Preprint in Authorea, DOI: 10.22541/au.168252801.19878064/v1.

Morin PA, Archer FI, Avila CD, et al. (2021) Reference genome and demographic history of the most endangered marine mammal, the vaquita. Molecular Ecology Resources, 21, 1008–1020.

Müller N, Ogilvie H, Zhang C, et al. (2018) Joint inference of species histories and gene flow. bioRxiv, 348391.

Müller NF, Rasmussen DA, Stadler T (2017) The Structured Coalescent and Its Approximations. Molecular Biology and Evolution, 34, 2970–2981.

NOAA (2008) Endangered and Threatened Species; Endangered Status for North Pacific and North Atlantic Right Whales. Federal Register, 75, Docket No. 080225302–8314–01.

Nordborg M, Tavaré S (2002) Linkage disequilibrium: What does history have to tell us. Trends in Environmental Analytical Chemistry, 18, 83–90.

North HL, McGaughran A, Jiggins CD (2021) Insights into invasive species from whole-genome resequencing. Molecular Ecology, 30, 6289–6308.

Ouborg NJ, Pertoldi C, Loeschcke V, Bijlsma RK, Hedrick PW (2010) Conservation genetics in transition to conservation genomics. Trends in Genetics, 26, 177–187.

Pace RM, Corkeron PJ, Kraus SD (2017) State-space mark-recapture estimates reveal a recent decline in abundance of North Atlantic right whales. Ecology and Evolution, 7, 8730–8741.

Paez S, Kraus RHS, Shapiro B, et al. (2022) Reference genomes for conservation. Science, 377, 364–366.

Palsbøll PJ, Larsen F, Hansen ES (1991) Sampling of skin biopsies from free-ranging large cetaceans in West Greenland: Development of new biopsy tips and bolt designs. Report of the International Whaling Commission, Special Is, 71–79.

Park L (2012) Linkage Disequilibrium Decay and Past Population History in the Human Genome. PLoS ONE, 7.

Pastene LA, Taguchi M, Lang A, Goto M, Matsuoka K (2022) Population genetic structure of North Pacific right whales. Marine Mammal Science, 38, 1249–1261.

Patenaude NJ, Portway VA, Schaeff CM, et al. (2007) Mitochondrial DNA diversity and population structure among southern right whales (*Eubalaena australis*). Journal of Heredity, 98, 147–157.

Payne R (1986) Long term behavioral studies of the southern right whale (*Eubalaena australis*). Reports of the International Whaling Commission Special Issue, 10, 161–167.

Pettis HM, Pace III RM, Hamilton PK (2022) North Atlantic Right Whale Consortium 2021 Annual Report Card. Report to the North Atlantic Right Whale Consortium, 25 pgs.

Purcell S, Neale B, Todd-Brown K, et al. (2007) PLINK: A tool set for whole-genome association and population-based linkage analyses. American Journal of Human Genetics, 81, 559–575.

Quinlan AR, Hall IM (2010) BEDTools: A flexible suite of utilities for comparing genomic features. Bioinformatics, 26, 841–842.

Rannala B, Yang Z (2017) Efficient Bayesian Species Tree Inference under the Multispecies Coalescent. Systematic Biology, 66, 823–842.

Rastogi T, Brown MW, McLeod BA, et al. (2004) Genetic analysis of 16th-century whale bones prompts a revision of the impact of Basque whaling on right and bowhead whales in the western North Atlantic. Canadian Journal of Zoology, 82, 1647–1654.

Reeves RR (2001) Overview of catch history, historic abundance and distribution of right whales in the western North Atlantic and in Cintra Bay, West Africa. J. Cetacean Res. Manage., Special Is, 187–192.

Reeves RR, Breiwick JM, Mitchell E (1992) Pre-Exploitation Abundance of Right Whales Off the Eastern United States. In: The Right Whale in the Western North Atlantic : A Science and Management Workshop (ed Hain J), pp. 5–7.

Reeves RR, Breiwick JM, Mitchell ED (1999) History of whaling and estimated kill of right whales, *Balaena glacialis*, in the Northeastern United States, 1620-1924. Marine Fisheries Review, 61, 1–36.

Reeves RR, Smith TD (2007) A Taxonomy of World Whaling. In: Whales, Whaling, and Ocean Ecosystems, pp. 82–101. University of California Press.

Reeves RR, Smith TD, Josephson EA (2007) Near-Annihilation of a Species: Right Whaling in the North Atlantic. In: The urban whale: North Atlantic right whales at the crossroads (eds Kraus SD, Rolland RM), pp. 39–74. Harvard University Press, Cambridge, MA.

Robinson JA, Brown C, Kim BY, Lohmueller KE, Wayne RK (2018) Purging of Strongly Deleterious Mutations Explains Long-Term Persistence and Absence of Inbreeding Depression in Island Foxes. Current Biology, 28, 3487-3494.e4.

Robinson JA, Kyriazis CC, Nigenda-Morales SF, et al. (2022) The critically endangered vaquita is not doomed to extinction by inbreeding depression. Science, 376, 635–639.

Schaeff CM, Kraus SD, Brown MW, et al. (1997) Comparison of genetic variability of North and South Atlantic right whales (*Eubalaena*), using DNA fingerprinting. Canadian Journal of Zoology, 75, 1073–1080.

Schiffels S, Durbin R (2014) Inferring human population size and separation history from multiple genome sequences. Nature Genetics, 46, 919–925.

Schiffels S, Wang K (2020) MSMC and MSMC2: The Multiple Sequentially Markovian Coalescent. In: Statistical Population Genomics, pp. 147–166.

von Seth J, van der Valk T, Lord E, et al. (2022) Genomic trajectories of a near-extinction event in the Chatham Island black robin. BMC Genomics, 23, 1–17.

Slater GJ, Goldbogen JA, Pyenson ND (2017) Independent evolution of baleen whale gigantism linked to Plio-Pleistocene ocean dynamics. Proceedings of the Royal Society B: Biological Sciences, 284.

Solís-Lemus C, Yang M, Ané C (2016) Inconsistency of Species Tree Methods under Gene Flow. Systematic Biology, 65, 843–851.

Suárez-Menéndez M, Bérubé M, Furni F, et al. (2022) Direct estimation of genome mutation rates from pedigrees in free-ranging baleen whales (pre-print). biorXiv.

Tataru P, Nirody JA, Song YS (2014) DiCal-IBD: Demography-aware inference of identity-by-descent tracts in unrelated individuals. Bioinformatics, 30, 3430–3431.

Taylor BL, Chivers SJ, Larese J, Perrin WF (2007) Generation length and percent mature estimates for IUCN assessments of cetaceans. Administrative Report LJ-07-01 National Marine Fisheries, 24.

Taylor RS, Jensen EL, Coltman DW, Foote AD, Lamichhaney S (2021) Seeing the whole picture: What molecular ecology is gaining from whole genomes. Molecular Ecology, 30, 5917–5922.

Theissinger K, Fernandes C, Formenti G, et al. (2023) How genomics can help biodiversity conservation. Trends in Genetics, 39, 545-559.

Thompson EA (2013) Identity by descent: Variation in meiosis, across genomes, and in populations. Genetics, 194, 301–326.

Tormosov DD, Mikhaliev YA, Best PB, et al. (1998) Soviet catches of southern right whales *Eubalaena australis*, 1951 ± 1971. Biological data and conservation implications. Biological Conservation, 86, 185–197.

Wade PR, Kennedy A, LeDuc R, et al. (2011) The world's smallest whale population? Biology Letters, 7, 83–85.

Waldick RC, Kraus S, Brown M, White BN (2002) Evaluating the effects of historic bottleneck events: An assessment of microsatellite variability in the endangered, North Atlantic right whale. Molecular Ecology, 11, 2241–2250.

Wang JY, Frasier TR, Yang SC, White BN (2008) Detecting recent speciation events: The case of the finless porpoise (genus *Neophocaena*). Heredity, 101, 145–155.

Wang K, Mathieson I, O'Connell J, Schiffels S (2020) Tracking human population structure through time from whole genome sequences. PLoS Genetics, 16, 1–24.

Waples RS, Antao T, Luikart G (2014) Effects of overlapping generations on linkage disequilibrium estimates of effective population size. Genetics, 197, 769–780.

Westbury M V., Hartmann S, Barlow A, et al. (2018) Extended and continuous decline in effective population size results in low genomic diversity in the world's rarest hyena species, the brown Hyena. Molecular Biology and Evolution, 35, 1225–1237.

Wolf M, de Jong M, Halldórsson SD, Árnason Ú, Janke A (2022) Genomic Impact of Whaling in North Atlantic Fin Whales. Molecular Biology and Evolution, 39, 1–14.

Wright S (1922) Coefficients of Inbreeding and Relationship. The American Naturalist, 56, 330–338.

Wright S (1931) Evolution in Mendelian Populations. Genetics, 16, 97–159.

Zhang C, Dong SS, Xu JY, He WM, Yang TL (2019) PopLDdecay: A fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. Bioinformatics, 35, 1786–1788.

4 Effects of inbreeding on reproductive success in endangered North Atlantic right whales

4.1 Introduction

Globally, the largest threats to biodiversity are anthropogenic in nature, but as population sizes dwindle, further risks associated with small populations remain, and can be much harder to mitigate. Genetic diversity erodes faster in small populations due to genetic drift (Wright 1931), and limited options for mates leads to mating between related individuals, which can compromise the health and fitness of offspring known as inbreeding depression (Wright 1922; Charlesworth and Charlesworth 1987; Keller and Waller 2002; Charlesworth and Willis 2009). Spielman et al. (2004) found that genetic factors impact species long before they go extinct but genetic effects are inherently difficult to manage in most species. Introducing individuals from other populations to increase genetic diversity, known as genetic rescue, has been effective in some cases such as an isolated population of grey wolves (*Canis lupus*) that benefitted from the arrival of an immigrant (Vilà et al. 2003) or in an endangered population of adders (*Vipera berus*) where new males were introduced (Madsen et al. 1999) which both resulted in rapid population growth. However, when an entire species is endangered,

or when the size or general biology of a species limits the practical application of genetic interventions, little can be done to mitigate the impacts of genetic diversity directly. In these cases, recognizing the impacts reduced genetic diversity and inbreeding are having on the species is still essential to conservation management to understand genetic limitations on population recovery and/or resiliency.

Inbreeding depression is being increasingly reported in wildlife populations with effects on many different measures of fitness such as parasite infection (Hoffman et al. 2014), maternal effects on offspring growth (Bérénos et al. 2016), lifetime breeding success (Huisman et al. 2016; Duntsch et al. 2023), and survival (Bérénos et al. 2016; Stoffel et al. 2021; Duntsch et al. 2023), but barriers to its detection still exist. Inbreeding depression is more likely to affect life-history traits than morphological traits (Crnokrak and Roff 1999; Coltman and Slate 2003), however there can be practical limitations to collecting these data in wildlife populations. For example, lifetime reproductive success may not be available for long-lived species, and limited resources for field work may result in missed detection of offspring. Furthermore, the true strength of inbreeding depression may be masked if the survival of highly inbred individuals is lower, leading to them being under-sampled. This was demonstrated in a study on harbour seals (*Phoca vitulina*) where inbreeding was associated with lungworm burden in young animals but not in older individuals presumably because more inbred animals

succumbed to infections early in life and at higher rates than non-inbred individuals (Rijks et al. 2008); therefore assessing inbreeding depression in harbour seals by studying only adults would have failed to detect evidence of inbreeding depression entirely. Thankfully, methods to quantify individual inbreeding in individuals have improved in recent years. Genomic estimates of inbreeding are now easily accessible and far outperform in their ability to capture realized inbreeding across many markers, and this has also opened the doors for the number of ways we can measure or quantify an individual's inbreeding coefficient. Inbreeding can be assessed with measures of heterozygosity such as standardized multi-locus heterozygosity (sMLH; Coltman et al. 1999) or homozygosity by loci (HL) which was developed specifically to assess heterozygosity fitness correlations (Aparicio et al. 2006). Inbreeding ancestry can also be measured by looking at runs of homozygosity (ROH) or stretches of the genome that are homozygous by descent (HBD). Recent inbreeding leads to longer ROH, whereas the ROH from older inbreeding events tend to break up into shorter segments by recombination over time (McQuillan et al. 2008). Evaluating inbreeding depression with different inbreeding coefficients may help disentangle its drivers and better understand how inbreeding is affecting, and will continue to affect, the species (Caballero et al. 2021).

North Atlantic right whales (*Eubalaena glacialis*) are an endangered species of baleen whale that inhabit the near coastal waters of eastern North America. An estimated 356 individuals remain and the population has been declining for the past decade due to a combination of high mortality and low reproduction (Pettis and Hamilton 2024). Vessel strikes and entanglement in commercial fishing gear are the leading causes of mortality (Daoust et al. 2017; Sharp et al. 2019; Pace et al. 2021). In addition, the sublethal effects of changes in food resources linked to climate change (Meyer-Gutbrod et al. 2023) and of anthropogenic activities (Knowlton et al. 2022) are having widespread impacts. Population growth in the species is facing even more challenges with a heavily male-biased sex-ratio (Pace et al. 2017) and females taking longer to transition from immature to breeding adults (Reed et al. 2022). Furthermore, females are not reproducing as frequently as expected with calving success only 27% of what would be expected (Frasier et al. 2023) and 36 reproductive age females have not produced a known calf (Moore et al. 2021). Based on the congeneric southern right whale (*E. australis*) and on the reproductive histories of some North Atlantic right whales, females should be capable of reproducing every three years (Knowlton et al. 1994; Best et al. 2001), yet few females are living up to that expectation and there is large variance in calving output with some females calving reliably every few years, while

others only produced one calf and some never having produced a calf (Moore et al. 2021).

Previous research has shown that North Atlantic right whales have extremely low levels of genetic diversity (Malik et al. 2000; Waldick et al. 2002; Frasier et al. 2007; Crossman et al. 2023), have had a relatively small effective population size throughout much of their history (Crossman et al. 2023), and show signs of relatively high rates of recent inbreeding (Crossman et al. 2023). These factors suggest that the poor reproductive success may be a result of inbreeding depression. North Atlantic right whales have been the focus of long-term field studies for over 40 years – with a significant increase in field effort beginning in the mid 1980s (Brown et al. 2007). Individual whales can likely live for over 70 years (Hamilton et al. 1998) and are uniquely recognizable after their first year based on external markings (Kraus et al. 1986). Also, calves stay in close proximity to their mothers while they are nursing for their first year of life (Hamilton et al. 1995), which has allowed many females' reproductive success to be tracked through time. Combined, these decades of individual-based data provide a wealth of information that forms the basis for our understanding of this species and for testing hypotheses regarding factors impacting health, mortality, and reproduction; as well as for better understanding general aspects of biology and evolution (e.g. Clutton-Brock and Sheldon 2010; Sheldon et al. 2022).

To estimate individual reproductive fitness, a metric is needed that is not biased by observer effort or the underlying biology of the species. Comparing the age at first calving as a measure of fitness could be biased toward younger females, for which their year of birth is known (whereas older females were born before studies began and are therefore of unknown age). Lifetime reproductive output require confidence in her first calving attempt and would underestimate the reproductive capacity of living females or females killed by anthropogenic activities who still have (or would have had) many calving years ahead of them. Estimating survivorship of calves beyond their first year would be interesting, but calves cannot be individually identified until the latter half of their first year, and therefore the fate of some calves is not known, and such analyses would have an inherent bias in detecting calves that survived their first year. Ideally, North Atlantic right whale female reproduction should follow a three year cycle where a female is pregnant for a year, suckles a calf for a year and then has a year of recovery before becoming pregnant again (Knowlton et al. 1994). Quantifying the inter-calf interval for a female as the mean number of years between calves could be an indicator of fitness if it accurately represents her ability to recover from pregnancy and lactation; however, if a calf dies very young, some females reproduce next with a two year interval which can be interpreted in different ways. It could either be biologically 'impressive' for a female's ability to reproduce, or it could signify she produces unfit

calves. Moreover, because our data are time-limited, there will always be a bias towards observing more shorter intervals than longer intervals, because there are more opportunities to observe shorter intervals. Such a metric would also ignore those females who have never given birth or have had only one calf.

Coulson et al. (2006) proposed a measure to estimate relative fitness called the 'de-lifing method' that overcomes many of the limitations described above. It combines a survivorship and a fecundity component to quantify an individual's relative lifetime contribution to population growth. One of the strengths of this method is that it scales an individual's contribution based on what is happening with others in the population – for example reproducing in a year when everyone else reproduced is weighted less heavily than reproducing in a year when very few offspring are born. Likewise, surviving a year with high mortality is weighted more heavily than surviving a year in which few individuals die. The fecundity component of the de-lifing method may be well suited to measure reproductive fitness in female right whales as it allows for quantifying an individual's relative reproductive success within the context of population-wide trends and fluctuations. In right whales, our fitness measurements will still suffer from an observation bias, however by using a de-lifing approach to measure fecundity, we can overcome some of the temporal fluctuations in the environmental,

such as changes to food resources which we know are having an increasing effect on the species (Meyer-Gutbrod et al. 2023).

North Atlantic right whales are critically endangered (IUCN 2022), have a low effective population size (Crossman et al. 2023) and estimated abundance (Pettis and Hamilton 2024), and few females are reaching their maximum reproductive potential (Moore et al. 2021; Reed et al. 2022). If inbreeding is having a strong influence on reproductive success, understanding this relationship could be extremely important for understanding genetic limitations on population growth. Weaker patterns could be indicative of external confounding factors affecting reproductive performance and could suggest that inbreeding may not be a main limitation to population growth – yet. If there is no correlation between inbreeding and reproductive success, inbreeding depression may still help explain poor reproductive success in the species, just not through reduced female fecundity. We used reduced representation genome sequencing to estimate inbreeding coefficients for 105 female North Atlantic right whales and used known calving history data to investigate the effects of inbreeding on reproductive success.

4.2 Materials and Methods

4.2.1 Sample Selection

The North Atlantic Right Whale Consortium (www.narwc.org) collaboratively maintains a database of sighting histories and life history data, as well as a DNA/tissue archive, for North Atlantic right whales over the past 40 years. For this study, we selected 105 North Atlantic right whale females over 10 years of age with known calving histories and for whom we had tissue or DNA archived. Briefly, skin samples have been collected since the late 1980s via biopsy using specially designed biopsy tips attached to crossbow bolts (see Brown et al. 1991 for more details). Sample collection is carried out in conjunction with photo-identification, to ensure the identity of the sampled whale is known. Samples are stored at Saint Mary's University (Halifax, Nova Scotia, Canada) in a 20% DMSO solution with 0.5M EDTA and saturated with NaCl (Seutin et al. 1991). If needed, DNA was extracted from skin biopsy samples in the same manner as the archived samples using standard phenol : chloroform methods (see Wang et al. 2008 for more details). We included eight duplicate samples in our library to verify the consistency of our genotyping methods. These duplicates represent DNA from the same extraction, run through the library preparation steps in parallel.

4.2.2 ddRAD Library Preparation

We prepared our libraries for double digest restriction site associated DNA sequencing (ddRADseq) (Peterson et al. 2012). For each sample, 400ng of DNA was digested with NlaIII and EcoRI-HF (New England Biolabs, NEB). Double-stranded adapters were annealed to the cut sites with T4 DNA ligase (NEB). We cleaned each reaction with Ampure XP beads prior to attaching a unique combination of Nextera xt indexes (Illumina) to each sample and performed another bead clean up. We pooled pairs of samples and ran each pool in a separate lane of a Pippin Prep (2% agarose cassettes with ethidium bromide) selecting fragments 440bp-540bp in size. Eluted product from two lanes of the Pippin Prep was pooled for a final bead clean-up. Four pools (each representing eight individually barcoded samples) were combined prior to normalizing. Libraries were sent to the McGill Genome Centre (Montreal, Canada) for normalizing and sequencing on one lane of a NovaSeq 6000 s4 150bp x 150bp run. The final library included 15% PhiX to increase library diversity. The supplementary information includes more details on library preparation (including Figure A4.1).

4.2.3 Read Mapping and Variant Calling

Additional details, including schematics of the bioinformatics pipelines from raw reads through variant calling and filtering, are provided in the supplemental methods and in Figures A4.2 and A4.3.

Briefly, a near chromosome length assembly of the North Atlantic Right Whale genome is available from DNAZoo (DNAZoo.org). We used the ShortRead package (Morgan et al. 2009) in R v3.6.0 to extract the reference sequences for the 21 longest scaffolds (2n = 42 in North Atlantic right whales; Pause et al. 2006). These 21 scaffolds account for ~91.4% of the entire assembly.

We used Trimmomatic v0.39 (Bolger et al. 2014) on demultiplexed sequence reads to remove Illumina adapters, drop leading bases with base quality scores < 20, drop reads with an average quality score < 30 and to trim reads when the mean base quality in 5bp sliding window fell below 20. We mapped reads from all samples using bwa-mem and used SAMtools (Li et al. 2009) to generate indexed bam files.

4.2.4 Variant Calling

We wanted to compare the performance to two prominent variant calling approaches, and therefore performed variant calling using Stacks v2.64 (Rochette et al. 2019) and Freebayes (Garrison and Marth 2012) both with a reference genome. Using both algorithms we also called invariant sites to better quantify the coverage of the genome. In the Stacks pipeline, we ran *gstacks* twice with different stringency thresholds setting both variant (--var-alpha) and genotype discovery (--gt-alpha) to 0.01 or 0.001. We ran the *populations* module for each *gstacks* data set requiring a locus to be present in 80% of

individuals (-r 0.8) and generated a vcf file of all sites called (variant and invariant). In Freebayes, we called variants, including monomorphic sites, in consecutive 6Mbp regions of the genome. In complex regions, we included additional constraints with the --use-best-n-alleles 4 flag and in some cases the --skip-coverage 10000 flag to allow Freebayes to complete variant calling.

We undertook a series of filtering steps on each dataset outlined in Figure A4.3 resulting in four different datasets representing two different stringency thresholds produced by each of the calling algorithms. Briefly, all datasets were filtered for missingness, repetitive regions were removed based on the RepeatMasker files that accompanied the reference genome on DNAZoo, the datasets were limited to biallelic SNPs and filtered to only retain sites with a minor allele frequency >0.01 and a minor allele count of at least 3. The Freebayes datasets were also filtered on mapping quality and depth (producing two Freebayes datasets based on different minimum genotype depth thresholds: DP5 and DP10). Stacks provides fewer annotations on the produced VCF files limiting the types of filters that can be applied - especially to invariant sites, and therefore instead of filtering on depth, we used lower alpha thresholds to increase the evidence needed to call a site or genotype as suggested by Rivera-Colón and Catchen (2022).

Prior to filtering to biallelic SNPs, Stacks called nearly twice as many sites as Freebayes and nearly all of the sites called by Freebayes were also called by Stacks. In the final datasets, the Stacks pipelines resulted in nearly 50% more biallelic SNPs called than the datasets generated by the Freebayes pipeline (Table A4.4). We compared the genotype calls between two of the datasets: Freebayes DP10 and Stacks 0.001 representing the strictest filtering regimes for each calling algorithm and of the sites that were identified as variant sites by both methods, genotypes across individuals were called with high concordance having only 0.10-0.34% discordant genotypes (i.e. 7-27 discordant genotypes / 4834-8089 assessed). True discordance rates are likely slightly higher however, as sites called as a variant by one pipeline and invariant by another were not compared.

We proceed with presenting results for Stacks called with the stricter alpha threshold (0.001) to retain a balance for strict filtering as well as retaining a larger number of SNPs. We present the main results from the other datasets in the supplemental information.

Our sample set included duplicate samples which we can use to assess the consistency of our variant calling pipelines. We used BCFtools gtcheck (Danecek et al. 2021) to compare genotypes of each pair of duplicates.

4.2.5 Calculating Reproductive Fitness

Following the de-lifing method put forth by Coulson et al. (2006), we calculated the fecundity contribution of individual female right whales to overall population fecundity. Long-term studies on North Atlantic right whales, conducted by a range of entities, but with data organization and management being led by the New England Aquarium and the North Atlantic Right Whale Consortium (www.narwc.org), has amassed decades of sightings and calving data, providing detailed life history information for individual whales. We limit our fecundity measures to between 1990, when field work became relatively consistent (Brown et al. 2007; Pace et al. 2017), and 2020, when survey effort was high, and sighting records in the database were considered complete. Sighting reports and field data take time to compile and verify, therefore database records for more recent years are not considered comprehensive.

Females were considered an adult at nine years of age if their birth year was known or eight years after initially being sighted if their birth year was unknown. A female was also considered an adult the year in which she had her first calf if this occurred before the nine-year threshold.

Population fecundity for each year was calculated as the number of calves that were observed divided by the number of adult estimated to be females alive in that year

based on updates to the Pace et al. (2017) model (R.M. Pace III *pers. comm.*). For example, given the optimal three-year reproductive cycle (Knowlton et al. 1994), in a 'perfect year' one third of all adult females would calve and the population fecundity would be calculated as 0.33.

We calculated the individual fecundity contribution for each of the 105 unique females in our genetic dataset as the relative contribution of a female to the population fecundity in a given year. In this way, a birth event is weighted more heavily in a year when few calves are born and given less weight in a year with a greater number of calving events. As the optimal reproductive cycle of a right whale female would be three years, we did not want to penalize a female for not having a calf in a year where many other calves were born because she was in a recovery or pregnancy year; therefore, we considered whether or not a female gave birth in a year with a sliding window approach meaning that her calving contribution for $year_i$ would consider whether or not she had a calf in $year_{i-1}$, $year_i$ or $year_{i+1}$. We calculated the mean individual fecundity contributions for each female over the years she was alive and adult (Equation 4.1), and omitted two females for which we had fewer than six years of fecundity values.

$$mean\left(\frac{reproduced_{year(i-1|i|i+1)}(1 \mid 0) - popfecundity_{year}}{NumAdultFemales_{yeari} - 1}\right)$$

Equation 4.1 Mean annual fecundity contributions calculated for each female.

4.2.6 Calculating Inbreeding Coefficients

Individual inbreeding coefficients can be calculated with a number of different approaches. We used five different methods to calculate individual inbreeding coefficients for each female North Atlantic right whale in our genetic dataset to assess potential nuances in the way inbreeding may be presenting itself the species. First, we calculated the inbreeding coefficient F for each individual using the --het flag in VCFtools v0.1.16 (Danecek et al. 2011) to estimate the deviation between observed and expected heterozygosity within an individual. We used the *genhet* package (Coulon 2010) in R v4.2.2 (R Core Team 2022) to calculate internal relatedness (IR: Amos et al. 2001) which incorporates allele frequency into its calculations and homozygosity by locus (HL: Aparicio et al. 2006) which also uses allele frequency by weighting the contribution of each locus to overall homozygosity. We used the package *InbreedR* (Stoffel et al. 2016) in R v3.6.0 to calculate standardized multi-locus heterozygosity (sMLH: Coltman et al. 1999; Szulkin et al. 2010) which assesses relative heterozygosity across the genome of individuals. Finally, we used the package RZooRoH to identify

the lengths of HBD segments using K=11 (Representing 10 HBD classes, and 1 non-HBD class). We calculated the proportion of the genome found in HBD tracts (F_{HBD}) as the total length of HBD segments longer than 100Kbp as a fraction of the total length of the 21-scaffold reference.

4.2.7 Estimating effect of inbreeding on fitness

We assessed the relationship between each inbreeding coefficient (F, sMLH, IR, HL and F_{HBD}), and fecundity with a Bayesian regression model (Eq. 4.2 & 4.3). As continuous variables, both individual fecundity and inbreeding coefficients were standardized (z-transformation) prior to being included in the model. The model was run in R v4.3.0 with the RStan package (Stan Development Team 2020), with 2000 warm-up steps and 10,000 subsequent iterations.

$$mu = \beta 0 + (\beta 1 * InbreedingCoefficient)$$

$$Individual\ Fecundity \sim normal(mu, sigma)$$

Equations 4.2 & 4.3 Bayesian models to estimate the effect of inbreeding coefficient on individual fecundity.

We gathered the posteriors of each model and calculated the mean of the posterior distribution and the 95% highest density interval (HDI) for the slope associated with each inbreeding coefficient and each dataset. To test whether specific genetic regions may be having a greater affect on fecundity, we conducted an association test using PLINK v1.9 (Purcell et al. 2007) to look for correlations between genotypes at given sites and individual fecundity. We identified the genes associated with the most significant SNPs by intersecting them with the genome feature file that accompanied the reference assembly on DNAZoo, looking for putative genes located 100Kb up- or down-stream from the positions of interest. We identified the potential function of these genes using the NCBI Gene database (National Center for Biotechnology Information).

4.3 Results

After sequencing, we obtained a mean of 19.5M paired-end reads per sample (NCBI BioProject: PRJNA1027072), with nearly 90% of reads passing filters and a mean mapping quality of 54.69 after mapping to 21 scaffolds of the North Atlantic right whale reference assembly (Table 4.1). For each variant calling program (Freebayes and Stacks), we produced two datasets with the stricter thresholds from each program (Freebayes

DP10 and Stacks 0.001) yielding fewer SNPs after filtering, as expected (Table A4.4). The number of sites (both variant and invariant) called after removing repetitive regions was between 34.3Mbp and 56.5Mbp depending on the dataset, indicating that our results are based on approximately 1.59 – 2.61% of the 21-scaffold reference assembly. In order to balance a trade off between including a greater number of SNPs and individuals, as well as maintaining high confidence in variant calls, we opted to proceed with presenting results for the Stacks 0.001 dataset which called 26,011 biallelic SNPs across 93 individuals. The main results for the other datasets are presented in Appendix A4.

Table 4.1 Sequencing summary statistics

Number of samples sequenced	113
Raw paired-end reads per sample (mean ± SD)	19 565 803 ± 10 884 334
Percent of reads passing filters (mean ± SD)	89.87 ± 3.92
Length of 21 scaffold assembly (bp)	2 166 782 980
Mean mapping quality	54.69 ± 0.89

4.3.1 Fitness

Individual fecundity estimates were able to be calculated for 103 females (Figure 4.1a). There was clear variation in reproductive performance with individual fecundity values being shaped like a normal distribution across all samples.

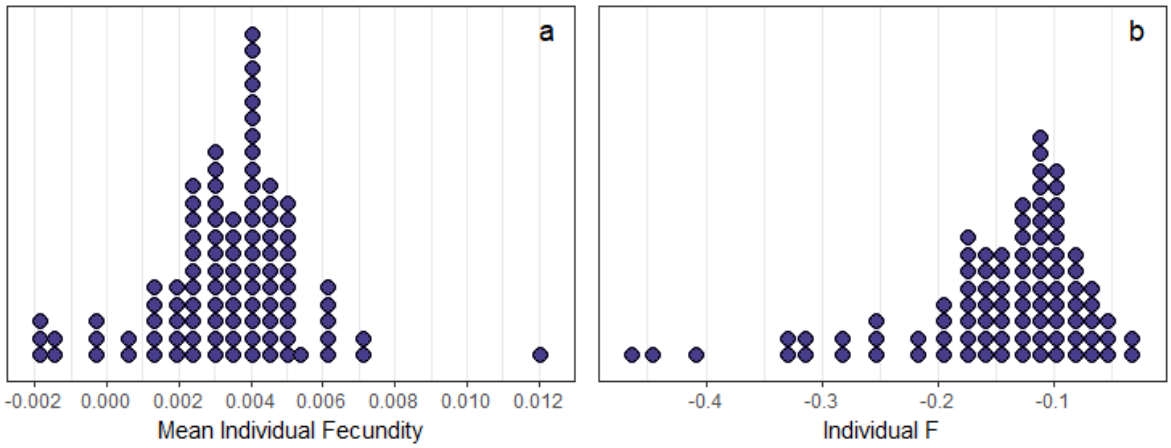

Figure 4.1 The distribution of a) individual mean fecundity scores representing reproductive fitness for individual females (n = 103) and b) individual inbreeding coefficient (F) calculated for each individual female (n = 93).

4.3.2 Inbreeding coefficients

Each inbreeding coefficient showed quite a bit of variation across samples. F values for all individuals were negative (mean ± SD = -0.148 ± 0.083) indicating individuals were all more heterozygous than expected (Figure 4.1b). All individuals also had negative IR values (mean ± SD = -0.151 ± 0.064) indicative of outbreeding or disassortative mating. The proportion of an individual's genome found in HBD segments ($F_{HBD>100Kb}$) ranged

from 0.06% to over 10% (mean ± SD = 0.027 ± 0.020). sMLH ranged from 0.910-1.258 (mean ± SD = 1.008 ± 0.068) and HL ranged from 0.466-0.621 (mean ± SD = 0.577 ± 0.030).

4.3.3 Inbreeding and fitness

The relationship between all inbreeding coefficients and fitness was very weak. F, HL, IR and $F_{HBD>100Kb}$ all had a very slight negative relationship with individual fecundity, and the opposite pattern was exhibited for sMLH, suggesting that as inbreeding levels increase, an individual's fecundity only slightly decreases (Figure 4.2). These trends were echoed by the posteriors of the Bayesian models, which all showed a peak probability of a negative effect, but for which the 95% HDI overlapped a slope of zero (Figure 4.3). The results from all four datasets are presented in the appendix (Figures A4.4 - A4.9). There are slight differences in the patterns found between each dataset, suggesting the specific sites included may be important for identifying inbreeding depression – therefore the effects of inbreeding on individual fecundity, while small, may be driven by effects at specific loci – rather than global patterns.

Our genome-wide association test revealed a number of SNPs with genotypes that were highly correlated with individual fecundity (Figure 4.4). Table 4.2 lists the 20 SNPs with the lowest P-values and the putative genes identified by the genome feature

file accompanying the reference assembly within ±100Kb. Interestingly, seven of the most significant SNP associations were found on HiC_scaffold_20 and a few have

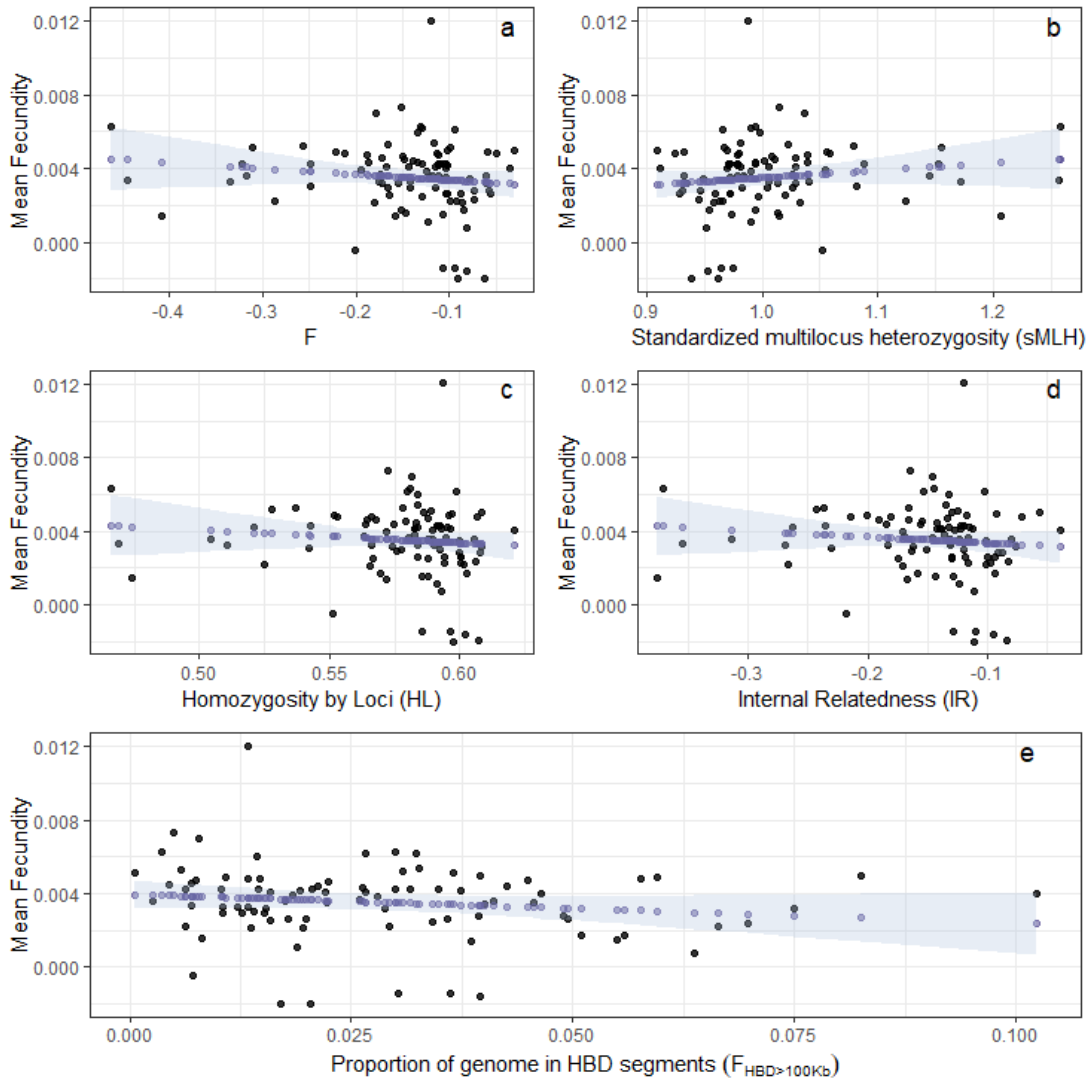

Figure 4.2 Relationship between individual fecundity and inbreeding coefficients (a. F, b. sMLH, c. HL, d. IR, e. $F_{HBD>100Kb}$) using the Stacks 0.001 dataset. The posterior predicted values are indicated by the blue points and the 95% HDI is represented by the shaded area. Note that sMLH is a measure of heterozygosity, whereas the others are measures of homozygosity. So even though the slopes go in the opposite direction, they mean the same thing: as homozygosity increases, fecundity slightly decreases.

putative involvement in reproduction: C2ORF80 may be associated with gonad development, PRMT7 may be involved in genomic imprinting and INO80E may be involved in DNA recombination (National Center for Biotechnology Information). The nature of ddRADseq means that coverage across the genome is not complete and therefore these sites identify possible genomic regions where haplotypes of specific genes may be having a direct effect on fitness.

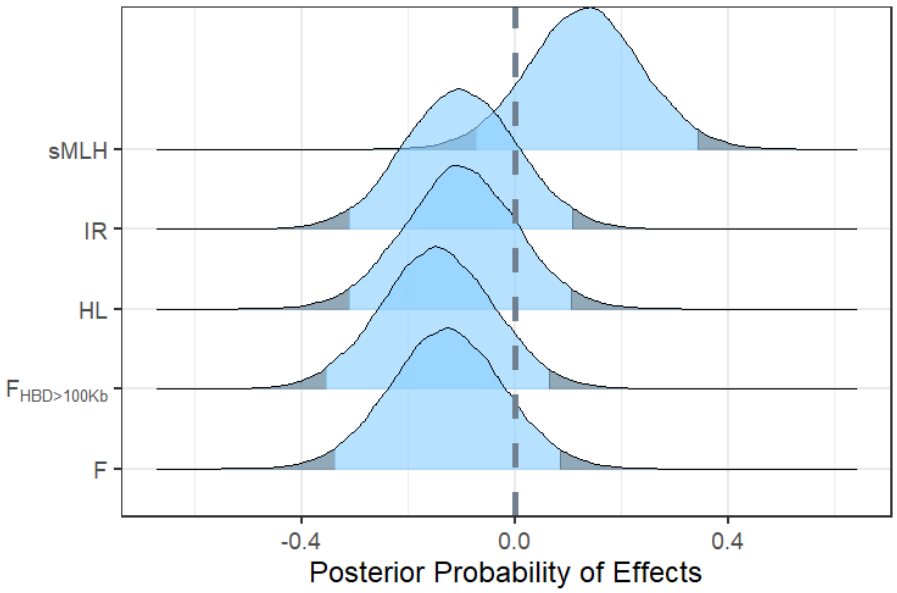

Figure 4.3 The effect of each inbreeding coefficient on individual mean fecundity for the Stacks 0.001 dataset. Tails of the posterior distributions falling in the highest or lowest 2.5% are shaded. Note that sMLH is a measure of heterozygosity, whereas the others are measures of homozygosity. So even though the effects go in the opposite direction, the magnitude of the effect size is very similar across all inbreeding coefficients.

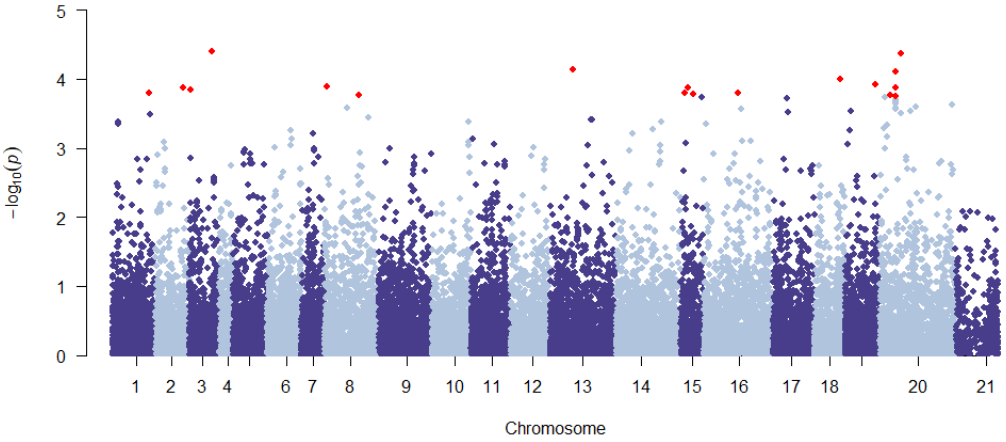

Figure 4.4 Association between genotype and reproductive fecundity across the genome. The twenty most significant SNPs are coloured in red and listed in Table 4.2.

4.4 Discussion

We identified a small, negative correlation between inbreeding coefficients and individual fecundity in female North Atlantic right whales, suggesting that although inbreeding may have a slight impact on female reproductive success, it is likely not the primary driver of reproductive variance found across individuals. This result was surprising given the extreme low levels of genetic diversity found in this species, its history of long-term small population size, and reproductive histories that varied highly across females, but which showed much smaller variation within each female (suggesting an intrinsic trait was likely influencing reproductive success). These patterns were consistently identified across a suite of inbreeding coefficients, and across datasets generated with different variant calling pipelines.

Table 4.2 List of the top twenty SNPs with the lowest P-values identified in our GWAS and the gene annotations that were identified within ±100Kb per the annotations associated with the reference assembly. Genes listed in bold have putative associations with reproduction as described in the footnotes.

Site	*Similar Genes within 100Kb*	*P Value*
HiC_scaffold_20 : 39101094	-	1.75E-04
HiC_scaffold_8 : 83910123	CBLB	1.72E-04
HiC_scaffold_20 : 26579680	PLEKHM3, IDH1, **C2ORF80**[1], CRYGA, PIKFYVE, Protein of unknown function	1.72E-04
HiC_scaffold_15 : 29455338	AKTIP, RBL2, CHD9	1.68E-04
HiC_scaffold_1 : 88360774	-	1.60E-04
HiC_scaffold_15 : 9924377	NUDT7, VAT1L	1.58E-04
HiC_scaffold_16 : 83116836	SMARCD2, S1PR1, CTNNAL1	1.57E-04
HiC_scaffold_3 : 2991346	-	1.44E-04
HiC_scaffold_15 : 17526456	SMPD3, SLC7A6OS, **PRMT7**[2], ESRP2, SLC7A6, PLA2G15, Protein of unknown function	1.34E-04
HiC_scaffold_20 : 39597891	-	1.33E-04
HiC_scaffold_20 : 39597903	-	1.33E-04
HiC_scaffold_2 : 67039621	HIRIP3, KCTD13, SEZ6L2, CDIPT, **INO80E**[3], TAOK2, TMEM219, ASPHD1, T-ENOL, RAB2A, MVP, PRRT2, MAZ, KIF22, C16ORF54, COX5A, Protein of unknown function (x4)	1.33E-04
HiC_scaffold_8 : 5989881	-	1.30E-04
HiC_scaffold_19 : 72789327	FAM49A	1.20E-04
HiC_scaffold_18 : 62196626	PREX2	1.02E-04
HiC_scaffold_20 : 39573960	-	7.71E-05
HiC_scaffold_20 : 39573966	-	7.71E-05
HiC_scaffold_13 : 56211802	-	7.36E-05
HiC_scaffold_20 : 50834179	TTN, Protein of unknown function	4.33E-05
HiC_scaffold_3 : 53787103	ZC3H13, CPB2, LCP1	3.95E-05

[1] associated with gonad development
[2] involved in genomic imprinting
[3] involved in DNA recombination

Reduced representation sequencing methods have become a cost-effective way of greatly increasing the number of markers able to be sequenced across a large number of samples where whole genome sequencing is still not feasible (Catchen et al. 2017). The efficacy, accuracy, and efficiency of different variant calling pipelines employed for RADseq data have often been compared (e.g. Puritz et al. 2014; Torkamaneh et al. 2016; Shafer et al. 2017; Rochette et al. 2019; Wright et al. 2019; Casanova et al. 2021) and with no clear winner emerging, the best pipeline for each dataset may ultimately depend on the expectations of the allele frequencies, sample size, coverage, availability of a reference genome, and likely many other factors. In this study, variant calling with Stacks2 (Rochette et al. 2019) and Freebayes (Garrison and Marth 2012) yielded a different number of SNPs, and while we found high concordance in genotypes at variant sites, it is likely that rare alleles were handled differently by the two approaches and were therefore called differently by the two algorithms (e.g. some variable sites that were retained by Stacks2, may have been called by Freebayes as homozygous across all individuals). Casanova et al. (2021) compared outputs produced by different variant calling pipelines for five different RADseq datasets, and as in our study, the different programs resulted in different SNP panels, but had little downstream effects on estimates of genetic differentiation. Conversely, Shafer et al. (2017) suggest variant calling pipeline and downstream filtering can greatly influence the final outcomes from

a study. Our approach of using two different pipelines, both detecting very minimal influence of inbreeding on fecundity should lend further support to our findings.

This study suggests that adult female North Atlantic right whales are largely avoiding the negative consequences often associated with inbreeding. This could be a result of genetic purging whereby deleterious mutations have been selectively removed from the species' gene pool, but confirmation of this will require subsequent dedicated studies investigating genetic load. The inbreeding coefficients estimated for these females are also informative on their own with respect to understanding population dynamics within the species. For the past several decades, there have been fewer than 500 North Atlantic right whales (Pace et al. 2017; Pettis and Hamilton 2024), clearly limiting the opportunities for mates. While female reproduction may not be experiencing the effects of inbreeding depression, the viability of inbred calves may be affected. Observed heterozygosity was much greater than expected heterozygosity for all individuals in our dataset. This reinforces previous work based on microsatellites that right whales were more heterozygous than would be expected given the genotypes of their parents (Frasier et al. 2013). If female heterozygosity is not impacting her own reproductive fitness, then genotypes of the calves are likely dictating their survival. Frasier et al (2013) suggested post-copulatory selection for dissimilar gametes may be driving this pattern, however these findings may also be explained by fetal mortality

being biased to fetuses with high inbreeding coefficients. Inbreeding depression acting on fetal survival or viability could help explain the decrease in successful calving events and the higher than expected heterozygosity rates seen in all of our samples. Excess heterozygosity was recently identified in an inbred population of Soay sheep (*Ovis aries*; Stoffel et al. 2023). As in this study, Stoeffl et al. (2023) suggest this is likely due to increased embryonic or fetal mortality of inbred individuals. Future analyses comparing genome-wide data from known mother-father-calf triads would be able to estimate the degree to which inbreeding is impacting fetal survival, and therefore better quantify the impact of inbreeding on reproductive success (or reproductive failure), and better identify the specific regions of the genome involved (sensu VanRaden et al. 2011; Pausch et al. 2015; Derks et al. 2017).

Excess genome-wide heterozygosity can be an artifact of sampling in small populations following a genetic bottleneck (Robertson 1965; Kirby 1975; Pudovkin et al. 1996; Balloux 2004). In this study, while we did only sample females which can increase this bias, these females represent overlapping generations and therefore are better representative of the allele frequencies in the entire population and not just in a single sex. Furthermore, in models that demonstrate excess heterozygosity, the modelled F statistics are much smaller than those we identified in this study – even for much lower effective population sizes (Robertson 1965; Balloux 2004). For both of these reasons, we

believe that while the small population of size of North Atlantic right whales could be contributing to the excess heterozygosity, the magnitude of this effect is likely driven by a loss of inbred fetuses.

It may also be possible that inbreeding depression on female reproductive fecundity is occurring, but due to effects at particular loci rather than at a genome-wide scale. Our GWAS suggests that there are SNPs throughout the genome where genotypes are highly associated with fecundity – especially on HiC_scaffold_20. Three of these sites were in close proximity to genes with potential involvement in reproduction (including C2ORF80). If mutations at C2ORF80 (also known as GONDA1 - gonad development associated 1) affect gonad development (National Center for Biotechnology Information), this could in turn, interfere with a female's reproductive potential, and while we don't have direct evidence for this involvement, it could help explain the presence of nulliparous females. Here genotypes at a few SNPs near genes involved in reproduction and at a handful of sites on HiC_scaffold_20 had significant association with female fecundity, and therefore the potential involvement of genes from these regions (including those not captured by our RADseq panel) should be investigated further for heterozygosity fitness correlations to better understand the impact of inbreeding on reproductive fecundity.

Our results suggest that inbreeding coefficients may only explain a small part of a female's reproductive fecundity, but the question remains as to what is driving variance in reproductive success? A recent study suggests variation in body size may play a role in reproductive success (Stewart et al. 2022). North Atlantic right whales are heavily exposed to anthropogenic stressors, but this is variable across females. Over 80% of individual right whales have been entangled at least once in their lifetime (Knowlton et al. 2012) and all whales are likely exposed to vessel disturbance (in various ways and to varying degrees), throughout their annual migratory routes, critical habitat areas and their lifetime. These sublethal stressors affect all right whales and can influence their fecundity through acute interactions (Knowlton et al. 2022), or through lifetime cumulative stress leaving epigenetic or other signatures of stress which could in turn affect fecundity. Models are being developed to quantify individual fitness and recovery throughout the lifetime of individual right whales incorporating visual health assessments (e.g. Pirotta et al. 2018, 2023; Knowlton et al. 2022). While inbreeding is having only a very small effect on reproductive fecundity, it could be affecting other fitness traits, or could be used as a correction factor for future individual based models.

This study is part of a larger effort to understand genetic limitations to population recovery in North Atlantic right whales and the scope of this study was to use genome-wide markers to investigate the effects of inbreeding on reproduction. We

found that inbreeding is only having a slight impact on the reproductive performance of female North Atlantic right whales. This is good news for the conservation and the long-term viability of the species, suggesting that the variation in female reproductive success is likely due to anthropogenic or other external factors such as food availability, many of which, unlike inbreeding, can be addressed through protection measures. However, our genomic data support results from earlier analyses suggesting that fetal mortality is biased towards inbred individuals, and therefore that inbreeding is impacting reproduction, just in a different manner than explicitly examined here. This finding is a double-edged sword, where such a pattern would have the negative short-term effect of lowering reproductive performance of the species (due to fetal loss), but the positive long-term effect (observed here) of maintaining heterozygosity at much higher levels than would be expected for a species with a small effective population size and this demographic history. Genomic analyses focused on addressing this pattern in more detail should be a high priority, as should teasing apart the relationship between stressors and female fecundity.

4.5 Data and Code Availability Statement

Raw sequence data have been archived in NCBI's Sequence Read Archive as BioProject PRJNA1027072. Code and data used for our analyses can be found at https://github.com/carlacrossman/NARW_ddRAD_InbreedingDepression.

4.6 References

Amos W, Worthington Wilmer J, Fullard K, et al (2001) The influence of parental relatedness on reproductive success. Proc R Soc B Biol Sci 268:2021–2027. https://doi.org/10.1098/rspb.2001.1751

Aparicio JM, Ortego J, Cordero PJ (2006) What should we weigh to estimate heterozygosity, alleles or loci? Mol Ecol 15:4659–4665. https://doi.org/10.1111/j.1365-294X.2006.03111.x

Balloux F (2004) Heterozygote excess in small populations and the heterozygote-excess effective population size. Evolution 58:1891–1900. https://doi.org/10.1111/j.0014-3820.2004.tb00477.x

Bérénos C, Ellis PA, Pilkington JG, Pemberton JM (2016) Genomic analysis reveals depression due to both individual and maternal inbreeding in a free-living mammal population. Mol Ecol 25:3152–3168. https://doi.org/10.1111/mec.13681

Best PB, Brandão A, Butterworth DS (2001) Demographic parameters of southern right whales off South Africa. J Cetacean Res Manag Spec Issue 2:161–169. https://doi.org/10.47536/jcrm.vi.296

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Brown MW, Kraus SD, Gaskin DE (1991) Reaction of North Atlantic right whales (*Eubalaena glacialis*) to skin biopsy sampling for genetic and pollutant analysis. Reports Int Whal Comm Special Is:81–89

Brown MW, Kraus SD, Slay CK, Garrison LP (2007) Surveying for discovery, science, and management. In: The urban whale: North Atlantic right whales at the crossroads. Harvard University Press, Cambridge, MA, pp 105–137

Caballero A, Villanueva B, Druet T (2021) On the estimation of inbreeding depression using different measures of inbreeding from molecular markers. Evol Appl 14:416–428. https://doi.org/10.1111/eva.13126

Casanova A, Maroso F, Blanco A, et al (2021) Low impact of different SNP panels from two building-loci pipelines on RAD-Seq population genomic metrics: case study on five diverse aquatic species. BMC Genomics 22:150. https://doi.org/10.1186/s12864-021-07465-w

Catchen JM, Hohenlohe PA, Bernatchez L, et al (2017) Unbroken: RADseq remains a powerful tool for understanding the genetics of adaptation in natural populations. Mol Ecol Resour 17:362–365. https://doi.org/10.1111/1755-0998.12669

Charlesworth D, Charlesworth B (1987) Inbreeding depression and its evolutionary consequences. Annu Rev Ecol Syst 18:237–268. https://doi.org/10.1146/annurev.es.18.110187.001321

Charlesworth D, Willis JH (2009) The genetics of inbreeding depression. Nat Rev Genet 10:783–796. https://doi.org/10.1038/nrg2664

Clutton-Brock T, Sheldon BC (2010) Individuals and populations: The role of long-term, individual-based studies of animals in ecology and evolutionary biology. Trends Ecol Evol 25:562–573. https://doi.org/10.1016/j.tree.2010.08.002

Coltman DW, Pilkington JG, Smith JA, Pemberton JM (1999) Parasite-mediated selection against inbred Soay sheep in a free-living island population. Evolution 53:1259–1267. https://doi.org/10.1111/j.1558-5646.1999.tb04538.x

Coltman DW, Slate J (2003) Microsatellite measures of inbreeding: A meta-analysis. Evolution 57:971–983. https://doi.org/10.1111/j.0014-3820.2003.tb00309.x

Coulon A (2010) Genhet: An easy-to-use R function to estimate individual heterozygosity. Mol Ecol Resour 10:167–169. https://doi.org/10.1111/j.1755-0998.2009.02731.x

Coulson T, Benton TG, Lundberg P, et al (2006) Estimating individual contributions to population growth: Evolutionary fitness in ecological time. Proc R Soc B Biol Sci 273:547–555. https://doi.org/10.1098/rspb.2005.3357

Crnokrak P, Roff DA (1999) Inbreeding depression in the wild. Heredity 83:260–270. https://doi.org/10.1038/sj.hdy.6885530

Crossman CA, Fontaine MC, Frasier TR (2023) A comparison of genomic diversity and demographic history of the North Atlantic and Southwest Atlantic southern right whales. Mol Ecol 00:1–17. https://doi.org/10.1111/mec.17099

Danecek P, Auton A, Abecasis G, et al (2011) The variant call format and VCFtools. Bioinformatics 27:2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Danecek P, Bonfield JK, Liddle J, et al (2021) Twelve years of SAMtools and BCFtools. Gigascience 10:1–4. https://doi.org/10.1093/gigascience/giab008

Daoust PY, Couture EL, Wimmer T, Bourque L (2017) Incident Report: North Atlantic right whale mortality event in the Gulf of St. Lawerence, 2017. Collaborative Report Produced by: Canadian Wildlife Health Cooperative, Marine Animal Response Society, and Fisheries and Oceans Canada. 256 pp. 256 pp.

Derks MFL, Megens HJ, Bosse M, et al (2017) A systematic survey to identify lethal recessive variation in highly managed pig populations. BMC Genomics 18:1–12. https://doi.org/10.1186/s12864-017-4278-1

Duntsch L, Whibley A, de Villemereuil P, et al (2023) Genomic signatures of inbreeding depression for a threatened Aotearoa New Zealand passerine. Mol Ecol 1–15. https://doi.org/10.1111/mec.16855

Frasier TR, Gillett RM, Hamilton PK, et al (2013) Postcopulatory selection for dissimilar gametes maintains heterozygosity in the endangered North Atlantic right whale. Ecol Evol 3:3483–3494. https://doi.org/10.1002/ece3.738

Frasier TR, Hamilton PK, Brown MW, et al (2007) Patterns of male reproductive success in a highly promiscuous whale species: The endangered North Atlantic right whale. Mol Ecol 16:5277–5293. https://doi.org/10.1111/j.1365-294X.2007.03570.x

Frasier TR, Hamilton PK, Pace RM (2023) How compromised is reproductive performance in the endangered North Atlantic right whale? A proposed method for quantification and monitoring. bioRxiv 2011:2023.11.21.568115

Garrison E, Marth G (2012) Haplotype-based variant detection from short-read sequencing. arXiv preprint arXiv:1207.3907 [q-bio.GN]

Hamilton P, Knowlton A, Marx M, Kraus S (1998) Age structure and longevity in North Atlantic right whales *Eubalaena glacialis* and their relation to reproduction. Mar Ecol Prog Ser 171:285–292. https://doi.org/10.3354/meps171285

Hamilton PK, Marx MK, Kraus SD (1995) Weaning in North Atlantic Right Whales. Mar Mammal Sci 11:386–390. https://doi.org/10.1111/j.1748-7692.1995.tb00293.x

Hoffman JI, Simpson F, David P, et al (2014) High-throughput sequencing reveals inbreeding depression in a natural population. Proc Natl Acad Sci 111:3775–3780. https://doi.org/10.1073/pnas.1318945111

Huisman J, Kruuk LEB, Ellis P a., et al (2016) Inbreeding depression across the lifespan in a wild mammal population. Proc Natl Acad Sci 113:3585–3590. https://doi.org/10.1073/pnas.1518046113

IUCN (2022) The IUCN Red List of Threatened Species. In: Version 2022-1. https://www.iucnredlist.org. Accessed 19 Oct 2022

Keller LF, Waller DM (2002) Inbreeding effects in wild populations. Trends Ecol Evol 17:230–241. https://doi.org/10.1016/S0169-5347(02)02489-8

Kirby GC (1975) Heterozygote frequencies in small subpopulations. Theor Popul Biol 8:31–48. https://doi.org/10.1016/0040-5809(75)90037-4

Knowlton AR, Clark JS, Hamilton PK, et al (2022) Fishing gear entanglement threatens recovery of critically endangered North Atlantic right whales. Conserv Sci Pract 4:1–14. https://doi.org/10.1111/csp2.12736

Knowlton AR, Hamilton PK, Marx MK, et al (2012) Monitoring North Atlantic right whale *Eubalaena glacialis* entanglement rates: A 30 yr retrospective. Mar Ecol Prog Ser 466:293–302. https://doi.org/10.3354/meps09923

Knowlton AR, Kraus SD, Kenney RD (1994) Reproduction in North Atlantic right whales (*Eubalaena glacialis*). Can J Zool 72:1297–1305. https://doi.org/10.1139/z94-173

Kraus SD, Moore KE, Price CA, et al (1986) The use of photographs to identify individual North Atlantic right whales (*Eubalaena glacialis*). Rep Int Whal Comm Special Is:145–151

Li H, Handsaker B, Wysoker A, et al (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Madsen T, Shine R, Olsson M, Wittzell H (1999) Restoration of an inbred adder population. Nature 402:34–35. https://doi.org/10.1038/46941

Malik S, Brown MW, Kraus SD, White BN (2000) Analysis of mitochondrial DNA diversity within and between North and South Atlantic right whales. Mar Mammal Sci 16:545–558. https://doi.org/10.1111/j.1748-7692.2000.tb00950.x

McQuillan R, Leutenegger AL, Abdel-Rahman R, et al (2008) Runs of Homozygosity in European Populations. Am J Hum Genet 83:359–372. https://doi.org/10.1016/j.ajhg.2008.08.007

Meyer-Gutbrod EL, Davies KTA, Johnson CL, et al (2023) Redefining North Atlantic right whale habitat-use patterns under climate change. Limnol Oceanogr 68:S71–S86. https://doi.org/10.1002/lno.12242

Moore M, Rowles T, Fauquier D, et al (2021) Assessing North Atlantic right whale health: threats, and development of tools critical for conservation of the species. Dis Aquat Organ 143:205–226. https://doi.org/10.3354/dao03578

Morgan M, Anders S, Lawrence M, et al (2009) ShortRead: A bioconductor package for input, quality assessment and exploration of high-throughput sequence data. Bioinformatics 25:2607–2608. https://doi.org/10.1093/bioinformatics/btp450

National Center for Biotechnology Information Gene [Internet]. Bethesda Natl Libr Med 2004-2023 Accessed:2023-09-21 [www.ncbi.nlm.nih.gov/gene/]

Pace RM, Corkeron PJ, Kraus SD (2017) State-space mark-recapture estimates reveal a recent decline in abundance of North Atlantic right whales. Ecol Evol 7:8730–8741. https://doi.org/10.1002/ece3.3406

Pace RM, Williams R, Kraus SD, et al (2021) Cryptic mortality of North Atlantic right whales. Conserv Sci Pract 1–8. https://doi.org/10.1111/csp2.346

Pausch H, Schwarzenbacher H, Burgstaller J, et al (2015) Homozygous haplotype deficiency reveals deleterious mutations compromising reproductive and rearing success in cattle. BMC Genomics 16:1–13. https://doi.org/10.1186/s12864-015-1483-7

Pause KC, Bonde RK, McGuire PM, et al (2006) G-banded karyotype and ideogram for the North Atlantic right whale (*Eubalaena glacialis*). J Hered 97:303–306. https://doi.org/10.1093/jhered/esj033

Peterson BK, Weber JN, Kay EH, et al (2012) Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PLoS One 7:e37135. https://doi.org/10.1371/journal.pone.0037135

Pettis HM, Hamilton PK (2024) North Atlantic Right Whale Consortium 2023 annual report card

Pirotta E, Booth CG, Costa DP, et al (2018) Understanding the population consequences of disturbance. Ecol Evol 8:9934–9946. https://doi.org/10.1002/ece3.4458

Pirotta E, Schick RS, Hamilton PK, et al (2023) Estimating the effects of stressors on the health, survival and reproduction of a critically endangered, long-lived species. Oikos 1–15. https://doi.org/10.1111/oik.09801

Pudovkin AI, Zaykin D V., Hedgecock D (1996) On the potential for estimating the effective number of breeders from heterozygote-excess in progeny. Genetics 144:383–387

Purcell S, Neale B, Todd-Brown K, et al (2007) PLINK: A tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81:559–575. https://doi.org/10.1086/519795

Puritz JB, Hollenbeck CM, Gold JR (2014) dDocent: A RADseq, variant-calling pipeline designed for population genomics of non-model organisms. PeerJ 2014:1–14. https://doi.org/10.7717/peerj.431

R Core Team (2022) R: A language and environment for statistical computing.

Reed J, New L, Corkeron P, Harcourt R (2022) Multi-event modeling of true reproductive states of individual female right whales provides new insights into their decline. Front Mar Sci 9:1–12. https://doi.org/10.3389/fmars.2022.994481

Rijks JM, Hoffman JI, Kuiken T, et al (2008) Heterozygosity and lungworm burden in harbour seals (*Phoca vitulina*). Heredity (Edinb) 100:587–593. https://doi.org/10.1038/hdy.2008.18

Rivera-Colón AG, Catchen J (2022) Population Genomics Analysis with RAD, Reprised: Stacks 2. In: Methods in Molecular Biology. pp 99–149

Robertson A (1965) The interpretation of genotypic ratios in domestic animal populations. Anim Sci 7:319–324

Rochette NC, Rivera-Colón AG, Catchen JM (2019) Stacks 2: Analytical methods for paired-end sequencing improve RADseq-based population genomics. Mol Ecol 28:4737–4754. https://doi.org/10.1111/mec.15253

Seutin G, White BN, Boag PT (1991) Preservation of avian blood and tissue samples for DNA analyses. Can J Zool 69:82–90. https://doi.org/10.1139/z91-013

Shafer ABA, Peart CR, Tusso S, et al (2017) Bioinformatic processing of RAD-seq data dramatically impacts downstream population genetic inference. Methods Ecol Evol 8:907–917. https://doi.org/10.1111/2041-210X.12700

Sharp SM, McLellan WA, Rotstein DS, et al (2019) Gross and histopathologic diagnoses from North Atlantic right whale *Eubalaena glacialis* mortalities between 2003 and 2018. Dis Aquat Organ 135:1–31. https://doi.org/10.3354/dao03376

Sheldon BC, Kruuk LEB, Alberts SC (2022) The expanding value of long-term studies of individuals in the wild. Nat Ecol Evol 6:1799–1801. https://doi.org/10.1038/s41559-022-01940-7

Spielman D, Brook BW, Frankham R (2004) Most species are not driven to extinction before genetic factors impact them. Proc Natl Acad Sci 101:15261–15264. https://doi.org/10.1073/pnas.0403809101

Stan Development Team (2020) RStan: the R interface to Stan. Version 2.21.2

Stewart JD, Durban JW, Europe H, et al (2022) Larger females have more calves: influence of maternal body length on fecundity in North Atlantic right whales. Mar Ecol Prog Ser 689:179–189. https://doi.org/10.3354/MEPS14040

Stoffel MA, Esser M, Kardos M, et al (2016) inbreedR: an R package for the analysis of inbreeding based on genetic markers. Methods Ecol Evol 7:1331–1339. https://doi.org/10.1111/2041-210X.12588

Stoffel MA, Johnston SE, Pilkington JG, Pemberton JM (2021) Genetic architecture and lifetime dynamics of inbreeding depression in a wild mammal. Nat Commun 12:1–10. https://doi.org/10.1038/s41467-021-23222-9

Stoffel MA, Johnston SE, Pilkington JG, Pemberton JM (2023) Purifying and balancing selection on embryonic semi-lethal haplotypes in a wild mammal. Evol Lett 2022.12.02.518882. https://doi.org/10.1093/evlett/qrad053

Szulkin M, Bierne N, David P (2010) Heterozygosity-fitness correlations: A time for reappraisal. Evolution 64:1202–1217. https://doi.org/10.1111/j.1558-5646.2010.00966.x

Torkamaneh D, Laroche J, Belzile F (2016) Genome-wide SNP calling from genotyping by sequencing (GBS) data: A comparison of seven pipelines and two sequencing technologies. PLoS One 11:1–14. https://doi.org/10.1371/journal.pone.0161333

VanRaden PM, Olson KM, Null DJ, Hutchison JL (2011) Harmful recessive effects on fertility detected by absence of homozygous haplotypes. J Dairy Sci 94:6153–6161. https://doi.org/10.3168/jds.2011-4624

Vilà C, Sundqvist AK, Flagstad Ø, et al (2003) Rescue of a severely bottlenecked wolf (*Canis lupus*) population by a single immigrant. Proc R Soc B Biol Sci 270:91–97. https://doi.org/10.1098/rspb.2002.2184

Waldick RC, Kraus S, Brown M, White BN (2002) Evaluating the effects of historic bottleneck events: An assessment of microsatellite variability in the endangered, North Atlantic right whale. Mol Ecol 11:2241–2250. https://doi.org/10.1046/j.1365-294X.2002.01605.x

Wang JY, Frasier TR, Yang SC, White BN (2008) Detecting recent speciation events: The case of the finless porpoise (genus *Neophocaena*). Heredity 101:145–155. https://doi.org/10.1038/hdy.2008.40

Wright B, Farquharson KA, McLennan EA, et al (2019) From reference genomes to population genomics: Comparing three reference-aligned reduced-representation sequencing pipelines in two wildlife species. BMC Genomics 20:1–10. https://doi.org/10.1186/s12864-019-5806-y

Wright S (1931) Evolution in Mendelian Populations. Genetics 16:97–159. https://doi.org/10.1093/genetics/16.2.97

Wright S (1922) Coefficients of Inbreeding and Relationship. Am Nat 56:330–338

5 A SNP panel designed for monitoring North Atlantic right whales

5.1 Introduction

Long-term studies provide incredibly powerful datasets to better understand the behaviour, ecology, evolution, population dynamics and more of wildlife populations (Clutton-Brock and Sheldon 2010; Grant and Grant 2014; Sheldon et al. 2022; Packer 2023). By monitoring a species or population over multiple generations, researchers are able to better detect changes that occur over time, and ideally identify what factors and forces resulted in those changes (Sheldon et al. 2022). Individual-based studies that monitor individuals throughout their lifetime and span multiple generations have the added benefit of quantifying individual variation overall, as well as in response to changes or pressures in the environment; measuring lifetime fitness; and identifying age-based life history traits (Clutton-Brock and Sheldon 2010; Pemberton et al. 2022; Sheldon et al. 2022). Long-term studies are also disproportionately cited in conservation reports, suggesting they provide valuable insight for management actions (Hughes et al. 2017).

In these study systems, it is not always clear what questions will be of greatest interest in the future, so many types of data are often collected and vary from study to

study. Many research teams will routinely collect samples for population genetic monitoring including tissue, blood, fur, scat, etc., because population genetic data can be valuable for ongoing population monitoring (Schwartz et al. 2007). For example, the combination of long-term individual-based field and ecological data with genetic data has resulted in a fine-scale understanding of many of the factors shaping differences in individual fitness, and how these patterns impact overall population trends in a population of Soay Sheep (*Ovis aries*) (e.g. Clutton-Brock and Pemberton 2004). As another example, the long-term consequences of natural genetic rescue to an isolated population of wolves (*Canis lupus*) on Isle Royale was only possible because of the long-term genetic monitoring (Hedrick et al. 2014).

Routine population genetic monitoring of endangered species usually entails relatively low throughput of samples, with genetic profiles of individuals stored in a database, with periodic reassessment of population level diversity statistics. Long-term databases (containing genetic profiles of individuals often dating back decades) therefore are often not using the most cutting-edge methods for routine monitoring as the contemporary tools/methods were not necessarily available when the study began, and with low sample throughput, the cost effectiveness of updating the entire database using more recent tools need to be considered as funds are often limited for long-term studies. Moreover, in almost all cases, sources of DNA are limited, and likely

irreplaceable (due to the death of individuals). Therefore, project managers also have to consider what is the best use of their limited stock of DNA. Many such managers may be reluctant to use valuable and limited DNA to update their database using the "latest techniques" if these techniques themselves may be replaced relatively soon.

For ongoing population genetic monitoring, Meek and Larson (2019) highlighted that despite the new genomic techniques being available for quite some time, many studies have been hesitant to make the switch away from traditional markers such as microsatellites to genomics. A recent study even advocated for the continued use for microsatellites being more than sufficient for assessing paternity in baleen whales where long-term datasets are well established (Suárez-Menéndez et al. 2024) (though our personal experience, at least with North Atlantic right whales, *Eubalaena glacialis,* indicate that this is untrue). Despite the slow progress, many projects are now making the transition to genomics (e.g. Van Bers et al. 2012; Roques et al. 2019; Stoffel et al. 2021). Panels of 100s or 1000s of single nucleotide polymorphism (SNPs) can now be developed with little reliance on specialized equipment (Campbell et al. 2015; Ali et al. 2016; Hoffberg et al. 2016). While there is a large upfront cost to converting the markers used for genetic monitoring studies, these SNP panels have demonstrated increased power over microsatellites (Zimmerman et al. 2020; Osborne et al. 2023; Pérez-González et al. 2023). SNP data can also help address a wider range of questions than

microsatellites as panel sizes are larger and the mutation rate of SNPs is better understood than the evolution microsatellites (Morin et al. 2004). Finally, genotyping SNPs requires much less discretion and subjectivity than scoring microsatellites and datasets can be easily compared across labs further improving a SNP panel's long-term utility (Moran et al. 2006; Ellis et al. 2011).

Regular population monitoring of North Atlantic right whales began in the late 1980s and is now overseen by member organizations of the North Atlantic Right Whale Consortium (NARWC). Individual right whales are uniquely identifiable by their callosity patterns which allows data collected at different time points to be assigned to the same individual (Kraus et al. 1986). Routine collection of skin samples via biopsy began in the late 1980s for the initial genetic studies on the species using restriction length polymorphisms (RLFP) of mitochondrial DNA (Schaeff et al. 1993), banding patterns following restriction enzyme digest of DNA captured by a human Y-chromosome probe (Brown et al. 1994) and DNA fingerprinting with mini-satellites (Schaeff et al. 1997). Presently, biopsy samples are collected from all new individuals in the population both as calves (when their mothers are known) and again as juveniles or adults when their callosity patterns develop, and they can be more easily identified. Each individual is profiled at 36 microsatellite loci (Frasier et al. 2006) and the mitochondrial haplotype is identified for a 500bp sequence of the control region (Malik

et al. 1999). These data have been routinely used to monitor genetic diversity, assign paternity, confirm the identity of dead whales and to match samples from calves with known mothers, to juveniles who are now identifiable, but no longer travel with their mother (e.g. Hamilton et al. 2022).

North Atlantic right whales have extremely low levels of genetic diversity (Frasier et al. 2007; Crossman et al. 2023). The panel of microsatellite loci currently used for genetic profiling are not extremely variable with almost half of them only having two or three alleles in the population (Waldick et al. 1999; Frasier et al. 2006) and therefore, it has a low power to assign some familial relationships. For example, although paternity can be assigned with relatively high confidence (when the mother is known); the resolution to assign parentage to whales when neither parent is known is very low, limiting the utility of these data for population monitoring. Furthermore, there are some inherent issues with using microsatellites to infer population genetic statistics. There is some subjectivity required in microsatellite scoring that we try to minimize with rigorous procedures for scoring including often running samples in triplicate, and having each individual scored by more than one observer (T. Frasier *pers comm*.).

We are now transitioning the marker panel for routine genotyping of North Atlantic right whales to a larger more diverse set of markers. Based on previous whole genome sequence data (Crossman et al. 2023) and a ddRADseq study (Crossman et al. 2024), we had a set of known variant sites and opted for a GT-Seq approach. Genotyping in thousands by sequencing (GT-Seq; Campbell et al. 2015) uses next-generation sequencing to sequence hundreds of multiplexed amplicons prepared in two simple reactions. GT-Seq panels have been developed for many other species with great success (e.g. McKinney et al. 2020; Morin et al. 2021; Hayward et al. 2022). We considered a RAPTURE (Ali et al. 2016) based approach as well, but the simplicity of GT-Seq both in the library preparation, which requires no specialized equipment, and in the simple analysis that requires little bioinformatic expertise once the panel has been developed and optimized, convinced us that GT-Seq was the most appropriate option for the North Atlantic right whale genetic database. One additional advantage of the GT-Seq panel over other sequencing approaches is that additional loci can be added to the panel with relative ease compared to other methods and therefore, as interest in other specific loci arise, they could easily be added into existing genotyping efforts. Here we present our development of a SNP panel designed for North Atlantic right whales and demonstrate that it will improve the resolution of ongoing genetic monitoring and increase confidence in identifying relationships between individuals.

5.2 Methods

5.2.1 Variant Site Selection

A previous study on North Atlantic right whales used ddRAD sequencing of over 100 females and identified 7,544 SNP loci that were called by two different variant calling pipelines (Crossman et al. 2024). We filtered these loci to remove sites with a minor allele frequency less than 10%, resulting in 2,872 potential SNP loci. In order to reduce the presence of null alleles as a result of variants within priming sites, we used datasets of unfiltered variant calls from both the same ddRADseq study noted above (Crossman et al. 2024) and from whole genome sequence data (Crossman et al. 2023) to exclude SNP loci that contained potential variant sites within possible priming regions 200 bases both up- and down-stream of the SNP of interest.

5.2.2 SNP Panel Primer Development

Using bedtools getfasta v.2.30 (Quinlan and Hall 2010), we extracted fasta sequences from the reference genome 200bp up- and down-stream of each SNP of interest. We proceeded with designing primers for 873 of 885 sites using Primer3 v.2.5.0 (Koressaar and Remm 2007; Untergasser et al. 2012) that would have an optimal melting temperature of 60°C, an optimum length of 23bp and generate a product 100-150bp in length. Using mfeprimer v.3.3.0 (Wang et al. 2019) we estimated the possible formation of hairpins and dimers within our set of primers. We removed one site where there was

a high possibility of hairpin formation by one of the primers and excluded a further 503 sites where at least one of the primers could potentially form five or more dimers. Finally, we used mfeprimer to test for specificity of binding across the genome using the reference genome (Genbank Assembly: GCA_028564815.2). We excluded 62 additional sites where at least one of the primers could create more than five potential amplicons.

We developed this SNP panel to replace ongoing genetic monitoring and profiling of individual North Atlantic right whales and as a result, we wanted to include loci for sex determination and mitochondrial haplotype identification. Two sets of primers designed for sex determination in cetaceans in previous studies (SRY: Einfeldt, Orbach, and Feyrer, 2019, Richard, McCarrey, and Wright, 1994; ZFX/ZFY: Konrad, Dupuis, Gero, and Frasier, 2017) were included in our panel. Finally, we designed two sets of primers that capture the known variant sites in the mitochondrial control region that are used to assign mitochondrial haplotypes (Malik et al. 2000). We proceeded with optimizing our panel for these 311 loci. Additional details on the library preparation including primer sequences are available in Appendix A5 and Table A5.1.

5.2.3 SNP Panel Optimization

Two round of panel optimization were performed using eighteen or nineteen North Atlantic right whale samples each. The DNA had been previously extracted following

Wang et al. (2008) and archived as part of a long-term genetic monitoring program led by the North Atlantic Right Whale Consortium (www.NARWC.org) and Saint Mary's University (Halifax, NS, Canada). We used the general methods described in Campbell et al. (2015) for library preparation with a few modifications. The first PCR consisted of 95 °C for 15 min; 5 cycles of: 95 °C for 30 sec, 5% ramp down to 57 °C for 30 sec, 72 °C for 2 min; 15 cycles of: 95 °C for 30 sec, 65 °C for 30 sec, 72 °C for 30 sec; hold at 4 °C and then visualized the performance of the reaction on a 1.5% agarose gel. We made a 10x dilution of the product and performed a bead clean-up with Ampure XP Beads with a 1.8:1 bead to diluted sample ratio. We indexed samples with a unique combination of Nextera XT indexes (Illumina, San Diego, USA), adding more product from PCR1 as template for samples with poor amplification. We cleaned the indexed product with a 1:1 ratio of Ampure XP beads to product and quantified each sample on a Qubit fluorometer (ThermoFisher, Waltham, USA). Completed libraries were sent to the McGill Genome Centre (Montreal, Canada) for pooling and sequencing with an Illumina MiSeq Nano 2x150bp run.

Demultiplexed sequences were processed through fastp v.0.23.4 (Chen et al. 2018) to remove paired end adapters, filter reads with an average quality below 30, trim poly-X tails and merge paired-end reads. Merged reads for each sample were analyzed through the GT-Seq pipeline (https://github.com/GT-Seq/GT-Seq-Pipeline) to compile

genotypes and read counts for each sample at each locus. Loci that could not be genotyped in any samples were excluded. Loci that had extremely high amplification were added at lower concentrations to the primer pool. This was repeated in a second round of optimization, lowering concentrations, or dropping additional primers if needed.

After these initial trials, the final round of sequencing we performed to assess panel performance included primers for 278 loci using the library preparation described above. We tested the panel on 47 unique North Atlantic right whale samples which included 11 mother-offspring pairs and six fathers that complete previously known mother-father-calf triads (based on previous work: Frasier et al. 2013, T. Frasier *pers. comm.*) and an additional four technical replicates to assess reproducibility in genotyping. Finally, we included five Southern right whale samples (*E. australis*) to assess the performance of the panel in that species. The 56 libraries were prepared as above and sequenced with an Illumina MiSeq Micro 2x150bp run.

5.2.4 Assessing Panel Performance

After our final sequencing run, demultiplexed samples were processed and merged with fastp as described above and run through the GT-Seq pipeline (https://github.com/GT-Seq/GT-Seq-Pipeline) to call genotypes. Mitochondrial

haplotypes were called by mapping merged reads produced by fastp v.0.23.4 (Chen et al. 2018) to a 500bp sequence of the mitochondrial control region (Malik et al. 1999) with bwa v.0.7.17 (Li and Durbin 2009). We generated consensus fasta sequences with samtools consensus v.0.17 that we used to compare to known haplotypes. We identified the haplotype of each individual and identified males based on heterozygous genotypes at the ZFX/ZFY locus and the presence of a genotype call at the SRY locus. We cross-referenced the sex and mtDNA haplotypes with those previously identified for each sample (following methods of: Malik et al. 1999; Shaw et al. 2003; Konrad et al. 2017). We assessed repeatability of the method by comparing genotypes called across four pairs of samples run in duplicate. Finally, we calculated mean heterozygosity across autosomal SNPs in sites missing less than 25% of genotype calls using the hierfstat package (Goudet 2005) in R v.4.3.1 and compared this to heterozygosity values estimated from a previous ddRAD study (Crossman et al. 2024). General panel performance was measured for the Southern right whale samples, but the sex and mitochondrial haplotypes were not assessed.

5.2.5 Paternity Assignment

We used Cervus v.3.0.7 (Kalinowski et al. 2007) to assign paternities for 11 individuals with known mothers. We used allele frequency data for the 231 autosomal sites in our panel that had less than 25% missing data as measured across ~100 individuals in a

previous study (Crossman et al. 2024). We simulated genotypes for 100,000 offspring with 150 candidate fathers, and assumed we sampled 10% of the population. Our simulations typed 85% of loci, the proportion of loci mistyped allowed was set to 0.005 and required a minimum of 75% of loci to be typed in a single individual. Between nine and thirteen candidate fathers were proposed for each of the individuals with known mothers based on males in our dataset that would have been alive and adults at the time the offspring was conceived.

5.2.6 Comparison with microsatellite dataset

Routine paternity assignments using the current North Atlantic right whale genetic database are conducted with a panel of 36 microsatellites (Frasier et al. 2006, 2013). We compared the probability of identity (PID) as calculated from both that panel (as a mean of PID calculated annually from 1980-2023) and from our SNP panel as calculated based on allele frequency data estimated in Cervus v.3.0.7 (Kalinowski et al. 2007). For the GT-Seq panel, PID was calculated given the allele frequencies identified from the 47 individuals in our dataset. We also compared the non-exclusion probability of fathers identified for calves with known mothers in the paternity analyses in Cervus v. 3.0.7 (Kalinowski et al. 2007). Existing paternity data from the microsatellites considered a much larger number of samples and candidate fathers for each offspring.

5.3 Results

5.3.1 Panel Performance

Our final round of optimization yielded 95,821 ± 21,949 (mean ± standard deviation) raw reads across all samples with a mean of 44,692 merged reads per individual passing filters (Table 5.1). The North Atlantic right whale samples had a high percentage of on target reads (95.77% ± 0.71%) and a mean depth per locus of 152.86 ± 32.71. These samples were successfully genotyped at 85.74% ± 2.97% of loci. Sex was correctly assigned for all North Atlantic right whale samples. After removing duplicates, sex chromosomes and poorly performing loci, the mean observed heterozygosity across all North Atlantic right whales was 0.2811 (based on 231 loci), which is comparable to results from the ddRAD dataset (0.2777; Crossman et al. 2024) (note that the higher heterozygosity we identified with GT-Seq can be explained at least in part by the filtering out of sites with low minor allele frequencies in our site selection). Across the four samples run in duplicate, there no mismatched genotype calls. However, there were 10-20 sites that were called in one sample but not called in the duplicate, but these do not represent a mismatch.

Table 5.1 Summary of sequencing results and genotyping success for both North Atlantic and southern right whales

	North Atlantic right whale (mean ± standard deviation)	***Southern right whale (mean ± standard deviation)***
Total Raw Reads	94,744.27 ± 20,233.53	106,804.80 ± 36,620.50
Merged Reads Passing Filters	44,516.37 ± 9,410.46	46,486.80 ± 16,065.92
% On-Target Reads	95.77 ± 0.70	92.89 ± 1.006
Mean Depth per Locus	151.86 ± 32.71	155.91 ± 53.91
mtDNA Reads Mapped	733.10 ± 578.06	135.00 ± 37.46
% Loci Genotyped	85.74 ± 2.97	81.89 ± 3.45

Mapping to the mitochondrial region for the North Atlantic right whale samples was quite variable in performance across samples (mean reads: 733.10, standard deviation: 578.06; Table 5.1). The mitochondrial haplotype was determined by combining two amplicons – one of which performed better than the other in the multiplex reaction. Of the 47 unique North Atlantic right whale samples, the mitochondrial haplotype was successfully matched to the previously known haplotype in 33 samples. The haplotype for a further five samples could be determined by only reads mapping from one amplicon. The haplotypes for the remaining nine samples could not be determined – largely due to poor amplification of the first amplicon of the mitochondrial haplotype (which contains most of the variant sites, including those responsible for identifying a heteroplasmic haplotype).

The five Southern right whale samples were successfully genotyped at 77.5% – 85.8% of loci (Table 5.1). However, the primers were not as specific for this species and

resulted in a lower percent of on-target reads (92.88% ± 1.01%) including fewer reads mapping to the mtDNA fragments (135 ± 37.46 reads, none of which came from the first mitochondrial amplicon).

5.3.2 Paternity assignment

Of the 11 individuals with known mothers, Cervus v3.0.7 identified eight fathers with strict confidence – seven of which had high confidence across the entire mother-offspring-father triad (Table 5.2).

5.3.3 Comparison with microsatellite dataset

The paternities assigned from our GT-Seq panel, were the same as those identified from the microsatellite panel for six of the eight offspring. In both mismatched pairs, the father identified with microsatellites was not included in the GT-Seq panel. The probability of non-exclusion from the microsatellite panel was nearly an order of magnitude higher for all individuals than was calculated for the GT-Seq panel. The mean PID for the GT-Seq panel (5.354×10^{-57}) was much high than PID calculated for the microsatellites (mean PID = 2.60×10^{-13}).

Table 5.2 Paternity and non-exclusion probabilities of fathers identified to offspring of known mothers with both the GT-Seq SNP panel and the existing microsatellite data. All paternities were assigned strict confidence unless otherwise listed.

		GT-Seq SNPs		Microsatellites	
Offspring	**Known Mother**	**Identified Father**	**Probability of non-exclusion**	**Identified Father**	**Probability of non-exclusion**
1123	1142	1050	2.353×10^{-19}	1050[+]	1.401×10^{-4}
2042	1142	1514	3.143×10^{-20}	1514[!]	4.734×10^{-5}
1703	1157	1516	1.086×10^{-19}	1516[+]	2.006×10^{-5}
1241	1240	1131[‡]	5.409×10^{-18}	1218*[!]	2.937×10^{-4}
3648	1248	1113	3.088×10^{-21}	1113	3.651×10^{-10}
1610	1509	1033	1.933×10^{-17}	1033[!]	6.287×10^{-6}
4353	2753	1803	2.073×10^{-20}	1803	8.605×10^{-7}
2557	1157	1113	9.378×10^{-18}	1238*[#]	2.748×10^{-9}

[+] Identified with relaxed confidence.
[‡] No confidence identified in triad and five mismatched loci between father and offspring.
[!] Most likely father, but paternity but not assigned.
* Not included in the GT-Seq test panel.
[#] Individual 1113 was included in microsatellite panel but not identified.

5.4 Discussion

The GT-Seq panel was able to successfully amplify over 234 loci. Genetic diversity as measured by our panel of 231 autosomal SNPs was comparable to previous measures of heterozygosity from ddRAD sequence data. Sexes could accurately be identified based on two sex determination loci. This panel was designed to convert the existing North Atlantic right whale genetic database of microsatellite genotypes to a larger, more informative panel of SNPs, and simultaneously resolve mtDNA haplotypes and sexes

for each sample. When applying this panel to the broader dataset (or in applying it to Southern right whales), we suggest that the two mitochondrial amplicons be prepared in their own multiplexed library for each sample. These can still use the same barcode/index but should be prepared independently of the larger multiplex. The same bioinformatic pipeline can be used, but this separation should reduce the problems with amplification of the first mtDNA primer pair and increase read depth to bolster confidence in haplotype assignment. The increased read depth will also help in identifying a known heteroplasmic haplotype that has previously been identified in North Atlantic right whales (McLeod and White 2010) that we were unable to resolve with our data.

Despite high confidence in paternity assignments generated from our new GT-Seq panel, not all of them matched the findings from the microsatellite data. These discrepancies may be resolved by increasing the number of candidate fathers in the paternity analyses, but some concern remains over the high confidence in these assignments from the GT-Seq panel. The GT-Seq dataset has been filtered for quality and includes far more loci than the microsatellite panel. At present, paternity assignments allows for some mismatches genotypes between father and offspring and strict assignments required 95% confidence. While it is important to allow some room for erroneous calls, null alleles etc., the low diversity in the species suggests a

confidence level above 95% may be required for better paternity assignments in North Atlantic right whales. Alternatively, methods implementing a Bayesian approach to paternity assignment have been gaining popularity and may reduce errors associated with incorrect paternity assignments (Hadfield et al. 2006; Hadfield 2009; Christie 2013; Christie et al. 2013; Anderson and Ng 2014). A Bayesian approach to paternity assignment may therefore be more suitable for this dataset moving forward.

The GT-Seq panel provides far greater power to identify individuals and paternities than the existing microsatellite panel as measured by both probability of identity and the probability of non-exclusion. While small tweaks to the library preparation and analytical pipeline will further strengthen its utility, it is clear that this will provide a much more robust dataset for future ongoing genetic monitoring of North Atlantic right whales. Further testing of this panel on more degraded or poor-quality samples – such as from dead carcasses, blow or fecal samples, will be another important step to understand if we can broaden its utility for less invasive sampling methods.

Long-term studies afford us with invaluable datasets to understand more about the ecology and evolution of a species (Grant and Grant 2024; Hughes et al. 2017; Pemberton et al. 2022; Sheldon et al. 2022), especially with the inclusion of genetic

monitoring which can be of additional value to species with long generation times and/or aid in understanding how species/populations are changing over time (Schwartz et al. 2007). Changing over an existing genetic monitoring program from any marker that has been used for decades to another is a daunting task that requires time and an upfront financial investment. Despite these barriers, the datasets produced can be much more powerful and enable a wider array of questions to be asked of the data. In this study we described the development of a panel of SNPs that can be used for simultaneous genotyping, sexing and mitochondrial haplotype identification for endangered North Atlantic right whales. We demonstrate the increased power of this new panel and make recommendations for changes to optimize it further. We also provide clear documentation of our development strategy – including our *in-silico* primer selection and development to increase the utility of these methods and/or breakdown some of the bioinformatic barriers to other studies that are seeking to convert from an existing monitoring program based on microsatellites to a panel of SNPs.

5.5 References

Ali OA, O'Rourke SM, Amish SJ, et al (2016) Rad capture (Rapture): Flexible and efficient sequence-based genotyping. Genetics 202:389–400. https://doi.org/10.1534/genetics.115.183665

Anderson EC, Ng TC (2014) Comment on "Bayesian parentage analysis with systematic accountability of genotyping error, missing data and false matching." Bioinformatics 30:743–745. https://doi.org/10.1093/bioinformatics/btt588

Brown MW, Kraus SD, Gaskin DE, White BN (1994) Sexual composition and analysis of reproductive females in the North Atlantic right whale, *Eubalaena glacialis*, population. Mar. Mammal Sci. 10:253–265

Campbell NR, Harmon SA, Narum SR (2015) Genotyping-in-Thousands by sequencing (GT-seq): A cost effective SNP genotyping method based on custom amplicon sequencing. Mol Ecol Resour 15:855–867. https://doi.org/10.1111/1755-0998.12357

Chen S, Zhou Y, Chen Y, Gu J (2018) Fastp: An ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34:i884–i890. https://doi.org/10.1093/bioinformatics/bty560

Christie MR (2013) Bayesian parentage analysis reliably controls the number of false assignments in natural populations. Mol Ecol 22:5731–5737. https://doi.org/10.1111/mec.12528

Christie MR, Tennessen JA, Blouin MS (2013) Bayesian parentage analysis with systematic accountability of genotyping error, missing data and false matching. Bioinformatics 29:725–732. https://doi.org/10.1093/bioinformatics/btt039

Clutton-Brock T, Sheldon BC (2010) Individuals and populations: The role of long-term, individual-based studies of animals in ecology and evolutionary biology. Trends Ecol Evol 25:562–573. https://doi.org/10.1016/j.tree.2010.08.002

Clutton-Brock TH, Pemberton JM (2004) Soay Sheep: Dynamics and Selection in an Island Population. Cambridge University Press, Cambridge, UK

Crossman CA, Fontaine MC, Frasier TR (2023) A comparison of genomic diversity and demographic history of the North Atlantic and Southwest Atlantic southern right whales. Mol Ecol 00:1–17. https://doi.org/10.1111/mec.17099

Crossman CA, Hamilton PK, Brown MW, et al (2024) Effects of inbreeding on reproductive success in endangered North Atlantic right whales. R Soc Open Sci Accepted:

Einfeldt AL, Orbach DN, Feyrer LJ (2019) A method for determining sex and chromosome copy number: Sex-by-sequencing reveals the first two species of marine mammals with XXY chromosome condition. J Mammal 100:1671–1677. https://doi.org/10.1093/jmammal/gyz131

Ellis JS, Gilbey J, Armstrong A, et al (2011) Microsatellite standardization and evaluation of genotyping error in a large multi-partner research programme for conservation of Atlantic salmon (*Salmo salar L.*). Genetica 139:353–367. https://doi.org/10.1007/s10709-011-9554-4

Frasier TR, Gillett RM, Hamilton PK, et al (2013) Postcopulatory selection for dissimilar gametes maintains heterozygosity in the endangered North Atlantic right whale. Ecol Evol 3:3483–3494. https://doi.org/10.1002/ece3.738

Frasier TR, McLeod BA, Gillett RM, et al (2007) Right whales past and present as revealed by their genes. In: The urban whale: North Atlantic right whales at the crossroads. Harvard University Press, Cambridge, MA, pp 200–231

Frasier TR, Rastogi T, Brown MW, et al (2006) Characterization of tetranucleotide microsatellite loci and development and validation of multiplex reactions for the study of right whale species (genus *Eubalaena*). Mol Ecol Notes 6:1025–1029. https://doi.org/10.1111/j.1471-8286.2006.01417.x

Goudet J (2005) HIERFSTAT, a package for R to compute and test hierarchical F-statistics. Mol Ecol Notes 5:184–186. https://doi.org/10.1111/j.1471-8286.2004.00828.x

Grant PR, Grant BR (2014) 40 Years of Evolution: Darwin's Finches on Daphne Major Island. Princeton University Press, Princeton, NJ

Hadfield JD (2009) MasterBayes: Maximum Likelihood and Markov chain Monte Carlo methods for pedigree reconstruction, analysis and simulation. R web 1–60

Hadfield JD, Richardson DS, Burke T (2006) Towards unbiased parentage assignment: Combining genetic, behavioural and spatial data in a Bayesian framework. Mol Ecol 15:3715–3730. https://doi.org/10.1111/j.1365-294X.2006.03050.x

Hamilton PK, Frasier BA, Conger LA, et al (2022) Genetic identifications challenge our assumptions of physical development and mother–calf associations and separation

times: a case study of the North Atlantic right whale (*Eubalaena glacialis*). Mamm Biol 102:1389–1408. https://doi.org/10.1007/s42991-021-00177-4

Hayward KM, Clemente-Carvalho RBG, Jensen EL, et al (2022) Genotyping-in-thousands by sequencing (GT-seq) of noninvasive faecal and degraded samples: A new panel to enable ongoing monitoring of Canadian polar bear populations. Mol Ecol Resour 22:1906–1918. https://doi.org/10.1111/1755-0998.13583

Hedrick PW, Peterson RO, Vucetich LM, et al (2014) Genetic rescue in Isle Royale wolves: genetic analysis and the collapse of the population. Conserv Genet 15:1111–1121. https://doi.org/10.1007/s10592-014-0604-1

Hoffberg SL, Kieran TJ, Catchen JM, et al (2016) RADcap: sequence capture of dual-digest RADseq libraries with identifiable duplicates and reduced missing data. Mol Ecol Resour 16:1264–1278. https://doi.org/10.1111/1755-0998.12566

Hughes BB, Beas-Luna R, Barner AK, et al (2017) Long-Term studies contribute disproportionately to ecology and policy. Bioscience 67:271–278. https://doi.org/10.1093/biosci/biw185

Kalinowski ST, Taper ML, Marshall TC (2007) Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. Mol Ecol 16:1099–1106. https://doi.org/10.1111/j.1365-294X.2007.03089.x

Konrad CM, Dupuis A, Gero S, Frasier T (2017) A sexing technique for highly degraded cetacean DNA. Aquat Mamm 43:655–660. https://doi.org/10.1578/AM.43.6.2017.655

Koressaar T, Remm M (2007) Enhancements and modifications of primer design program Primer3. Bioinformatics 23:1289–1291. https://doi.org/10.1093/bioinformatics/btm091

Kraus SD, Moore KE, Price CA, et al (1986) The use of photographs to identify individual North Atlantic right whales (*Eubalaena glacialis*). Rep Int Whal Comm Special Is:145–151

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Malik S, Brown MW, Kraus SD, et al (1999) Assessment of mitochondrial DNA structuring and nursery use in the North Atlantic right whale (*Eubalaena glacialis*). Can J Zool 77:1217–1222. https://doi.org/10.1139/z99-073

Malik S, Brown MW, Kraus SD, White BN (2000) Analysis of mitochondrial DNA diversity within and between North and South Atlantic right whales. Mar Mammal Sci 16:545–558. https://doi.org/10.1111/j.1748-7692.2000.tb00950.x

McKinney GJ, Pascal CE, Templin WD, et al (2020) Dense SNP panels resolve closely related chinook salmon populations. Can J Fish Aquat Sci 77:451–461. https://doi.org/10.1139/cjfas-2019-0067

McLeod BA, White BN (2010) Tracking mtDNA heteroplasmy through multiple generations in the North Atlantic right whale (*Eubalaena glacialis*). J Hered 101:235–239. https://doi.org/10.1093/jhered/esp098

Meek MH, Larson WA (2019) The future is now: Amplicon sequencing and sequence capture usher in the conservation genomics era. Mol Ecol Resour 795–803. https://doi.org/10.1111/1755-0998.12998

Moran P, Teel DJ, Lahood ES, et al (2006) Standardising multi-laboratory microsatellite data in Pacific salmon: An historical view of the future. Ecol Freshw Fish 15:597–605. https://doi.org/10.1111/j.1600-0633.2006.00201.x

Morin PA, Forester BR, Forney KA, et al (2021) Population structure in a continuously distributed coastal marine species, the harbor porpoise, based on microhaplotypes derived from poor quality samples. Mol Ecol 1457–1476. https://doi.org/10.1111/mec.15827

Morin PA, Luikart G, Wayne RK (2004) SNPs in ecology, evolution and conservation. Trends Ecol Evol 19:208–216. https://doi.org/10.1016/j.tree.2004.01.009

Osborne MJ, Caeiro-Dias G, Turner TF (2023) Transitioning from microsatellites to SNP-based microhaplotypes in genetic monitoring programmes: Lessons from paired data spanning 20 years. Mol Ecol 32:316–334. https://doi.org/10.1111/mec.16760

Packer C (2023) The Lion: Behavior, Ecology, and Conservation of an Iconic Species. Princeton University Press, Princeton, NJ

Pemberton JM, Kruuk LEB, Clutton-Brock T. 2022. The Unusual Value of Long-Term Studies of Individuals: The Example of the Isle of Rum Red Deer Project. Annu Rev Ecol Evol Syst 53:327–351.

Pérez-González J, Carranza J, Anaya G, et al (2023) Comparative Analysis of Microsatellite and SNP Markers for Genetic Management of Red Deer. Animals 13:3374. https://doi.org/10.3390/ani13213374

Quinlan AR, Hall IM (2010) BEDTools: A flexible suite of utilities for comparing genomic features. Bioinformatics 26:841–842. https://doi.org/10.1093/bioinformatics/btq033

Richard KR, McCarrey SW, Wright JM (1994) DNA sequence from the SRY gene of the sperm whale (*Physeter macrocephalus*) for use in molecular sexing. Can J Zool 72:873–877. https://doi.org/10.1139/z94-118

Roques S, Chancerel E, Boury C, et al (2019) From microsatellites to single nucleotide polymorphisms for the genetic monitoring of a critically endangered sturgeon. Ecol Evol 9:7017–7029. https://doi.org/10.1002/ece3.5268

Schaeff CM, Kraus SD, Brown MW, et al (1997) Comparison of genetic variability of North and South Atlantic right whales (*Eubalaena*), using DNA fingerprinting. Can J Zool 75:1073–1080. https://doi.org/10.1139/z97-129

Schaeff CM, Kraus SD, Brown MW, White BN (1993) Assessment of the population structure of western North Atlantic right whales (*Eubalaena glacialis*) based on sighting and mtDNA data. Can J Zool 71:339–345. https://doi.org/10.1139/z93-047

Schwartz MK, Luikart G, Waples RS (2007) Genetic monitoring as a promising tool for conservation and management. Trends Ecol Evol 22:25–33. https://doi.org/10.1016/j.tree.2006.08.009

Shaw CN, Wilson PJ, White BN (2003) A reliable molecular method of gender determination for mammals. J Mammal 84:123–128. https://doi.org/10.1644/1545-1542(2003)084<0123:ARMMOG>2.0.CO;2

Sheldon BC, Kruuk LEB, Alberts SC (2022) The expanding value of long-term studies of individuals in the wild. Nat Ecol Evol 6:1799–1801. https://doi.org/10.1038/s41559-022-01940-7

Stoffel MA, Johnston SE, Pilkington JG, Pemberton JM (2021) Genetic architecture and lifetime dynamics of inbreeding depression in a wild mammal. Nat Commun 12:1–10. https://doi.org/10.1038/s41467-021-23222-9

Suárez-Menéndez M, Bérubé M, Bachmann L, et al (2024) Baleen whale microsatellite panel for individual identification and parentage assignment in Mysticeti. Glob Ecol Conserv 51:e02947. https://doi.org/10.1016/j.gecco.2024.e02947

Untergasser A, Cutcutache I, Koressaar T, et al (2012) Primer3-new capabilities and interfaces. Nucleic Acids Res 40:1–12. https://doi.org/10.1093/nar/gks596

Van Bers NEM, Santure AW, Van Oers K, et al (2012) The design and cross-population application of a genome-wide SNP chip for the great tit *Parus major*. Mol Ecol Resour 12:753–770. https://doi.org/10.1111/j.1755-0998.2012.03141.x

Waldick RC, Brown MW, White BN (1999) Characterization and isolation of microsatellite loci from the endangered North Atlantic right whale. Mol Ecol 8:1763–1765. https://doi.org/10.1046/j.1365-294x.1999.00723-6.x

Wang JY, Frasier TR, Yang SC, White BN (2008) Detecting recent speciation events: The case of the finless porpoise (genus *Neophocaena*). Heredity (Edinb) 101:145–155. https://doi.org/10.1038/hdy.2008.40

Wang K, Li H, Xu Y, et al (2019) MFEprimer-3.0: Quality control for PCR primers. Nucleic Acids Res 47:W610–W613. https://doi.org/10.1093/nar/gkz351

Zimmerman SJ, Aldridge CL, Oyler-Mccance SJ (2020) An empirical comparison of population genetic analyses using microsatellite and SNP data for a species of conservation concern. BMC Genomics 21:1–16. https://doi.org/10.1186/s12864-020-06783-9

6 Conclusions

6.1 Summary of findings and application to policy

Genomic data and analyses are powerful tools that are more accessible than ever and enable questions to be answered that were not possible before – particularly as they pertain to small populations and their persistence in face of non-lethal stressors (Ouborg et al. 2010; Hohenlohe et al. 2021; Formenti et al. 2022; Theissinger et al. 2023). My thesis sought to provide four examples of how genomics can better inform conservation policy relating to marine mammal conservation in Canada.

In Chapter 2, I demonstrated that methylation patterns can be used to quantify differential stressors between populations using resident killer whales as an example. Assessing the cumulative effects of anthropogenic stressors and the long-term effects of chronic disturbance are listed as specific approaches and recovery measures in the Action Plan for the Northern and Southern Resident Killer Whale in Canada (Fisheries and Oceans Canada 2017). My work is a clear demonstration of how genomic tools can help inform these management measures.

In Chapter 3, I identified that North Atlantic right whales have existed with smaller effective population sizes than southern right whales from the Southwest

Atlantic for thousands of years, and confirmed they have extremely low diversity genome-wide and are suffering from more recent inbreeding than their southern counterparts. This work will help better manage expectations for how successful recovery for the species will look and helps explain how population growth may be limited via the demographic effects of long-term small effective population sizes as suggested in their most recent COSEWIC Status Report (COSEWIC 2013).

In Chapter 4, I identified that inbreeding depression was likely impacting North Atlantic right whale reproductive success, not by explaining variation in female reproductive success as we had originally hypothesized, but rather via selection against inbred calves. This work directly addresses a recovery objective from the Action Plan for the North Atlantic Right Whale in that it "increase[s our] understanding of life history characteristics, low reproductive rate […] and threats to recovery through research" (Fisheries and Oceans Canada 2021).

In Chapter 5, I developed a SNP panel that will greatly improve ongoing genetic monitoring of the North Atlantic right whale and will improve our ability to access genetic data from poor quality samples, and to assign paternity/maternity to individual whales. I also demonstrated its utility to be used for related southern right whale genotyping. The North Atlantic right whale Action Plan specifically lists supporting

genetic studies as an action item to address monitoring of the population and this new tool will improve upon ongoing monitoring efforts.

6.2 Knowledge transfer

> *"Science isn't finished until it's communicated."*
> -Sir Mark Walport, Chief Scientific Advisor to the UK government

My research not only demonstrates the applicability of genomic tools to address important questions in conservation, but in doing so, addresses specific knowledge gaps for two endangered species/populations of marine mammals in Canada as detailed above. Throughout my PhD research, I have worked to disseminate the findings of my research to a number of different audiences beyond the typical academic community that will encounter this research in the primary literature (three of the four data chapters from my thesis have been accepted for publication at the time of its submission).

My work on epigenetics and stress was part of a larger project investigating effects of stress over different timescales in resident killer whales. We reported these

findings at the end of the project directly to the funding agency (Fisheries and Oceans Canada – one of the primary policy makers affecting marine mammals in Canada).

I have presented my research from Chapters 3 & 4 at annual North Atlantic Right Whale Consortium (NARWC) Meetings and at a workshop hosted by the Canadian Wildlife Federation – both meetings which bring together science, policy and stakeholders in the interest of North Atlantic right whale conservation.

Colleagues at the National Ocean and Atmosphere Administration (NOAA) in the United States (the primary policy makers affecting marine mammals in the United States) became aware of my research both from my publications and from their presence at the NARWC meeting. I was invited to attend their Species in the Spotlight Health Assessment workshop in June 2024 which focused on the health and limitations to recovery in three species/populations (including North Atlantic right whales and Southern Resident killer whales). I was asked to speak at the meeting on the role of epigenetics for informing species recovery measures.

I have also worked to communicate our findings to the broader public by working with a scientific illustrator to develop cartoons to easily translate the results of our genomic work on North Atlantic right whales to the public.

Lastly, as part of a project on the conservation genomics of the North Atlantic right whale - which is funded by Genome Canada, Genome Atlantic, and Research Nova Scotia - we have been integrating social sciences and these genomic results to better engage with relevant entities associated with, or impacted by, right whale conservation; so that these results have the most appropriate impact on policy and industry (e.g., the fishing and shipping industries). This work is ongoing, and will extend beyond the time of the completion of my thesis, but I have played a key role in obtaining the results that will be a key part of this process, and have also been an active part of collaborating with the social scientists involved to aid the integration of these two approaches.

6.3 Future directions

Each of my chapters can act as stepping stones to future research to improve and inform conservation efforts for Species at Risk of extinction.

I demonstrated the application of using methylation patterns to measure cumulative stress in wildlife. Since this paper was published, the application of epigenetics to conservation is appearing more frequently in the literature, including the use of methylation patterns to detect accelerated aging as a measure of health

(Barratclough et al. 2024), and it is encouraging to see research progress in this area. In order to best incorporate epigenetics into cumulative effects modelling, further studies on the effects of stressors across the entire methylome will be extremely valuable.

My work on right whale demographic history was important in resolving that North Atlantic right whales have lived with a small population size for thousands of years. Future work to understand how long-term small effective population size has impacted the species was an important next step for this work and has been recently completed (Orton et al n.d.).

I was able to demonstrate that inbreeding depression is likely limiting the reproductive output of North Atlantic right whales, future work will be needed to quantify the reproductive deficit attributed to inbreeding, and/or pinpoint when the loss of inbred calves is occurring. This will be important for understanding the energetic burden that inbreeding depression is having on females, in addition to help set realistic recovery goals and expectations for population growth.

Finally, the SNP panel I developed should be applied to the entire North Atlantic right whale genetic database and could have utility for studies on southern right whales. In North Atlantic right whales, the panel will be used to reassess paternity and

help resolve unknown maternity in rare instances when not known from field data, in addition to ongoing genetic monitoring.

6.4 Concluding Remarks

Lag times still exists for implementing conservation actions from COSEWIC and SARA reports in Canada (Findlay et al. 2009; Mooers et al. 2010; Pynn 2019; Turcotte et al. 2021). Hopefully my work provides clear examples of the utility of genomic tools to better address outstanding questions of concern for Species at Risk in Canada in ways that may be more difficult or impossible with other methods.

6.5 References

Barratclough A et al. 2024. Estimating age and investigating epigenetic changes related to health across multiple bottlenose dolphin populations. Biological Conservation 293:110570.

COSEWIC. 2013. COSEWIC Assessment and Status Report on the North Atlantic Right Whale *Eubalaena glacialis*. Ottawa.

Findlay CS, Elgie S, Giles B, Burr L. 2009. Species listing under Canada's species at risk act. Conservation Biology 23:1609–1617.

Fisheries and Oceans Canada. 2017. Action Plan for the Northern and Southern Resident Killer Whale (*Orcinus orca*) in Canada. Page Species at Risk Act Action Plan Series.

Fisheries and Oceans Canada. 2021. Action Plan for the North Atlantic Right Whale (*Eubalaena glacialis*) in Canada. Page Species at Risk Act Action Plan Series.

Formenti G et al. 2022. The era of reference genomes in conservation genomics. Trends in Ecology and Evolution 37:197–202.

Hohenlohe PA, Funk WC, Rajora OP. 2021. Population genomics for wildlife conservation and management. Molecular Ecology 30:62–82.

Mooers AO, Doak DF, Scott Findlay C, Green DM, Grouios C, Manne LL, Rashvand A, Rudd MA, Whitton J. 2010. Science, Policy, and Species at Risk in Canada. BioScience 60:843–849.

Orton RW, Hamilton PK, Frasier TR. (n.d.) Genomic Evidence for the Purging of Deleterious Genetic Variation Through Inbreeding in the Endangered North Atlantic Right Whale. Evolutionary Applications. *Submitted*.

Ouborg NJ, Pertoldi C, Loeschcke V, Bijlsma RK, Hedrick PW. 2010. Conservation genetics in transition to conservation genomics. Trends in Genetics 26:177–187.

Pynn L. 2019. 'It just takes too damn long': How Canada's law for protecting at-risk species is failing. The Narwhal: https://thenarwhal.ca/ it-just-takes-too-damn-long-how-canadas-law-for-protecting-at-risk-species-is-failing/

Theissinger K et al. 2023. How genomics can help biodiversity conservation. Trends in Genetics 39:545–559.

Turcotte A, Kermany N, Foster S, Proctor CA, Gilmour SM, Doria M, Sebes J, Whitton J, Cooke SJ, Bennett JR. 2021. Fixing the Canadian species at risk act: Identifying major issues and recommendations for increasing accountability and efficiency. Facets 6:474–494.

7 Appendices

A2. An example of DNA methylation as a means to quantify stress in wildlife using killer whales – Supplemental Methods Details

Primer Design

We performed a literature search and identified genes involved in stress response that had demonstrated methylation changes in response to stressors in controlled studies. These included brain derived neurotropic factor (BDNF), corticotropin releasing factor (CRF) and glucocorticoid receptor (NR3C1) as well as a common control gene, β-actin (ACTB) whose methylation patterns should not be influenced by age, sex, or exposure to external stressors[1]. We identified reference sequences for the promoter region of each of these genes from the published killer whale genome (GenBank accession GCA_000331955.1) using the provided annotations for the desired genes on NCBI Genome Browser.

We also confirmed the annotations by aligning the killer whale reference sequences to genes identified in a similar manner from species with better reference genomes (e.g. human, mouse, cow; Fig A2.1).

The sodium bisulfite treated DNA is single stranded and the strands are no-longer complimentary; therefore, a different primer pair is required to amplify each

strand. We designed primer pairs for each strand. However, because the methylation patterns are complimentary on original strands, we only amplified samples using a single primer pair for each locus that showed the most reliable amplification (Table A2.1). Due to the low GC content in the DNA post-bisulfite conversion, long primers were needed to achieve the necessary specificity and annealing temperatures. Once we felt we had optimized conditions for each primer set, we performed Sanger sequencing to test that: (i) we had amplified the correct region, and (ii) to confirm successful conversion in the bisulfite treatment where the resulting sequences would be absent of cytosine except in potentially methylated sites.

Library Preparation

Following our initial PCR, we visualized a small amount of our products on 2% agarose gels. Using the remaining product, we performed a magnetic bead cleanup with Ampure XP beads using bead:sample ratios optimized for each primer set (ratios between 1.0:1 and 1.2:1) to eliminate all primer-dimers. In attempts to standardize the concentration of products across samples, the final elution volume at the end of the cleanup was adjusted based on the concentration determined from the pre-cleanup gel. (*i.e.* samples with concentrations <2.5ng/µl were eluted in 17.5µl, 2.5-10ng/µl were

eluted in 27.5µl and >10ng/µl were eluted in 52.5µl.) The cleaned products were size-separated and visualized on another 2% agarose gel to confirm concentrations. We used 5µl of cleaned product in an indexing PCR to bind sample-specific index combinations and adapters for Illumina sequencing using the Nextera xt Indexes (Illumina).

We performed a final bead cleanup after the indexing PCR using bead:sample ratios between 0.8:1 and 1:1. Cleaned libraries were eluted in a final volume of 27.5µl of Tris-HCl except where product concentration following the first bead cleanup was still <2.5ng/µl in which case the elution volume was reduced to 17.5µl.

Processing Amplicon Sequences

In CUTADAPT 2.6[2], we used an allowable error rate of 0.2 to separate loci and trim primer sequences. Our allowable error rate was higher than the default (0.1) to allow for more mistakes/mismatches within the primer sequence which CUTADAPT uses to identify distinct loci. Our main intention was to separate loci using the primer sequences and our primer sequences were sufficiently different from each other to allow this error rate. Trimming the primer sequences was less important for this study because they did not contain CpG sites that would be analyzed. We wanted to be more

conservative in including reads at this stage as reads with poor alignment or poor quality would be filtered out later in the pipeline.

Bayesian Model

We built a Bayesian model using the percent methylation of each site as the predicted variable in a Bayesian regression analysis where individual, age, sex, population and CpG site were the predictor variables, as well as the interaction between specific CpG sites and population. We also allowed for different standard deviations in each population. We built similar models investigating other pairwise interactions and none had an effect on our results, so we left them out of our final model. The model was run with a combination of R and RStan with 2000 steps as the warm-up and 12000 steps with recorded data.

$$mu = \beta_0 + \beta_{1[i]} ind_{[i]} + (\beta_2 * age) + \beta_{3[j]} sex_{[j]} + \beta_{4[k]} site_{[k]} + \beta_{5[l]} pop_{[l]} + \beta_{6[k,l]} site_{[k]} pop_{[l]}$$

$$percent.methylation \sim normal(\underline{mu, sigma_{[l]}})$$

To aid model performance, the continuous variables (percent methylation and age) were standardized as a z-scores (by subtracting each value from the mean, then dividing by the standard deviation), which transformed each into a normal distribution with a mean of zero and a standard deviation of one. Therefore, the prior probabilities for β_0 and β_2 were normal distributions with a mean of zero and a standard deviation of one. The coefficients for the other (categorical) variables were estimated in a hierarchical manner, where the hyper-priors for each mean was a normal distribution with a mean of zero and a standard deviation of one, and the hyper-priors for each standard deviation was a Cauchy distribution with a location parameter of one and a scale parameter of one.

The model was run with a combination of R v.3.6.0 and RStan. Performance of the MCMC process was assessed by examination of Rhat scores, effective sample size (ESS) estimates, and examination of trace plots. Performance of the model itself was tested using posterior predictive checks. The model was run with 2000 steps as the warm-up and 12000 steps with recorded data.

We reran the model under the same conditions omitting the sample that was collected from a freshly stranded carcass and this did not have an effect on our results.

1. Chen, I.-H. et al. Selection of reference genes for RT-qPCR studies in blood of beluga whales (*Delphinapterus leucas*). PeerJ 4, e1810 (2016).

2. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 17, 10–12 (2011).

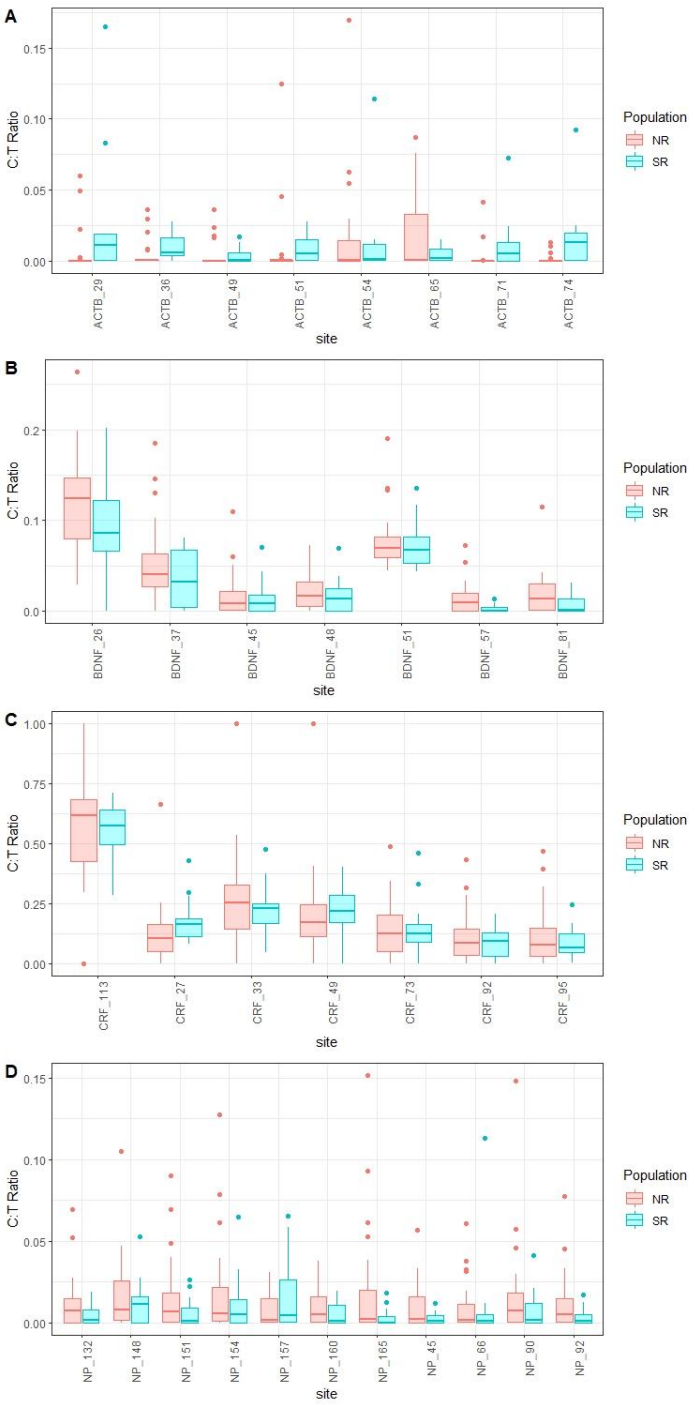

Figure A2.1. A comparison of the degree of methylation (C:T ratio in amplicon sequences – methylated : unmethylated bases) for CpG sites between populations for (A) ACTB, (B) BDNF, (C) CRF, (D) NR3C1. Northern Residents are represented in pink and Southern Residents in blue.

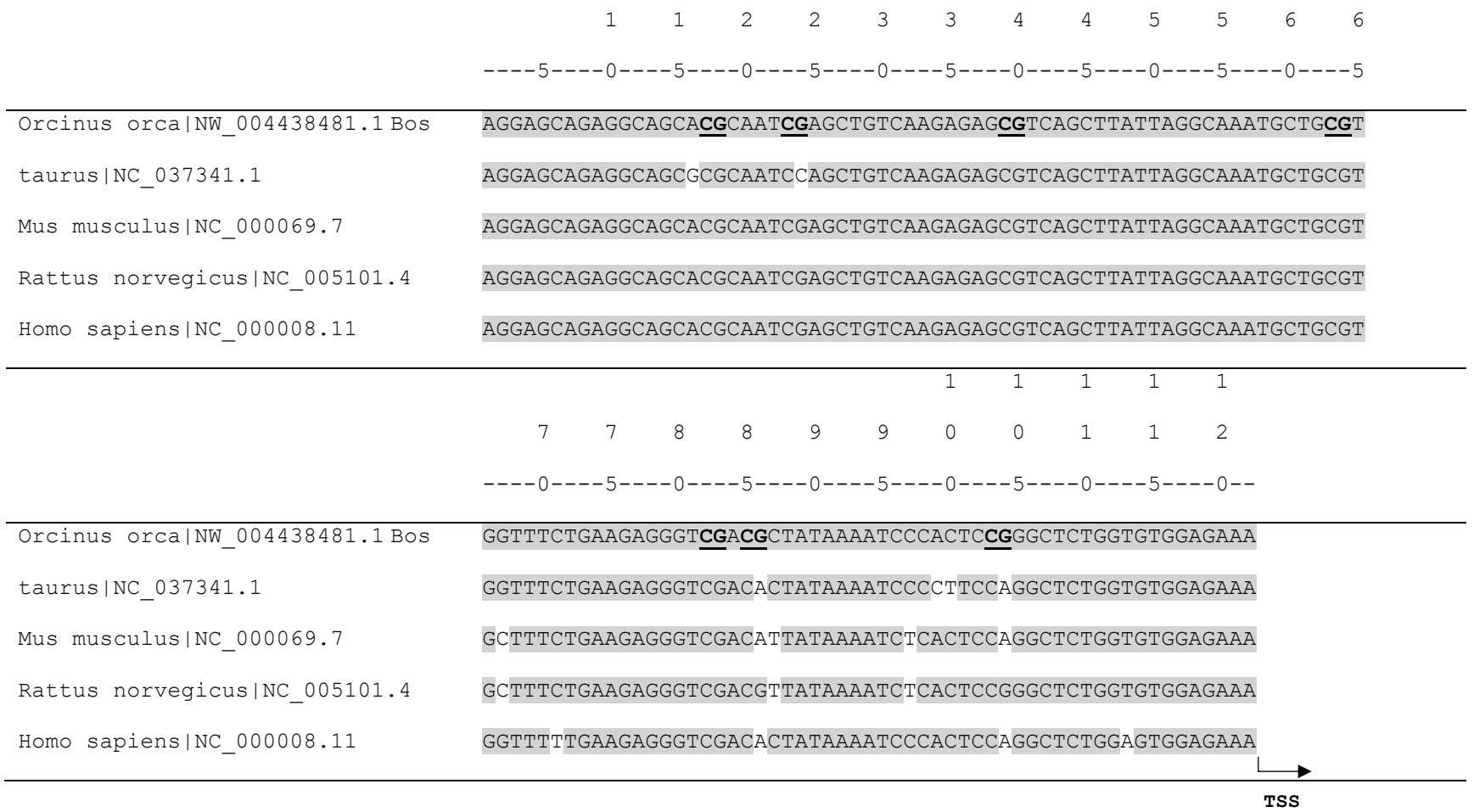

Figure A2.2. Alignment of the CRF promoter region between the killer whale (*Orcinus orca*), cow (*Bos taurus*), mouse (*Mus musculus*), rat (*Rattus norvegicus*), human (*Homo sapiens*). Transcriptional start site (TSS) and CpG sites in the killer whale sequence (bold & underlined) are identified.

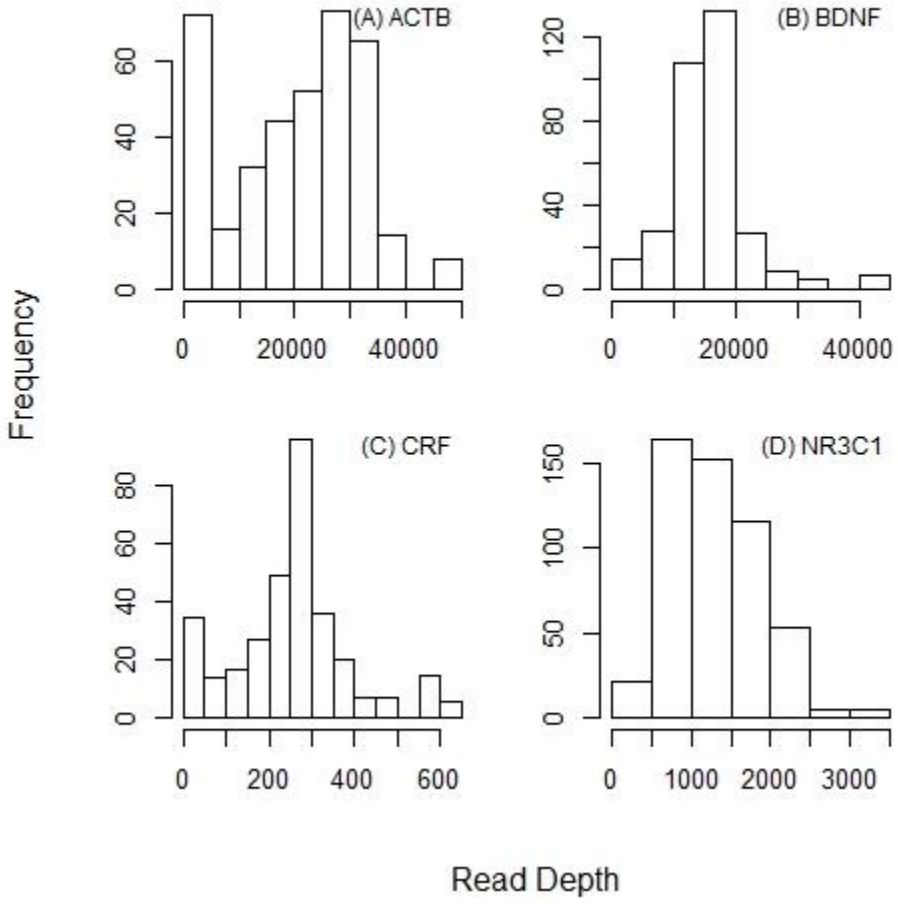

Figure A2.3. Histograms of read depth as calculated for each CpG site for (A) ACTB, (B) BDNF, (C) CRF and (D) NR3C1 used to determine threshold for omitting samples

Table A2.1. Loci and primers used in this study.

Gene and Region	Abbr.	Locus-Specific Primer Sequence	Ta	Product Length
β-actin	ACTB	F: 5'-CTCTCTACCAATCCATCTCTC-3' R: 5'-GAGTTATAAAAGGTAATTTTTGAA-3'	58°C	158bp
Brain Derived Neurotropic Factor	BDNF	F: 5'-TAAAGGAGTTATAATGAGTTGGT-3' R: 5'-AACCCAACCTACACACTTACC-3'	62°C	276bp
Corticotropin Releasing Factor	CRF	F: 5'-TGATTTATGTAGGAGTAGAGG-3' R: 5'-TAACAACTCAAACAATACAAAATTAAA-3'	60°C	351bp
Glucocorticoid Receptor	NR3C1	F: 5'-TTTTTTTAAAAATAATAATTTAAAAATGT-3' R: 5'-AAATTACAAAACAAAACCCACCCTC-3'	58°C	232bp

Table A2.2. Samples omitted from our model for insufficient read depth

Locus	Samples Omitted			
	N	Population	Sex	Age
ACTB	11	SR	F	38
		SR	F	23
		SR	F	26
		SR	M	22
		SR	M	19
		NR	F	21
		NR	F	26
		NR	F	24
		NR	F	19
		NR	F	26
		NR	M	6
BDNF	2	SR	M	22
		SR	F	26
CRF	7	SR	F	38
		SR	F	26
		SR	F	30
		NR	F	26
		NR	F	17
		NR	M	38
		NR	M	20
NR3C1	2	SR	M	22
		SR	F	26

A3. A comparison of genomic diversity and demographic history of the North Atlantic and Southwest Atlantic southern right whales

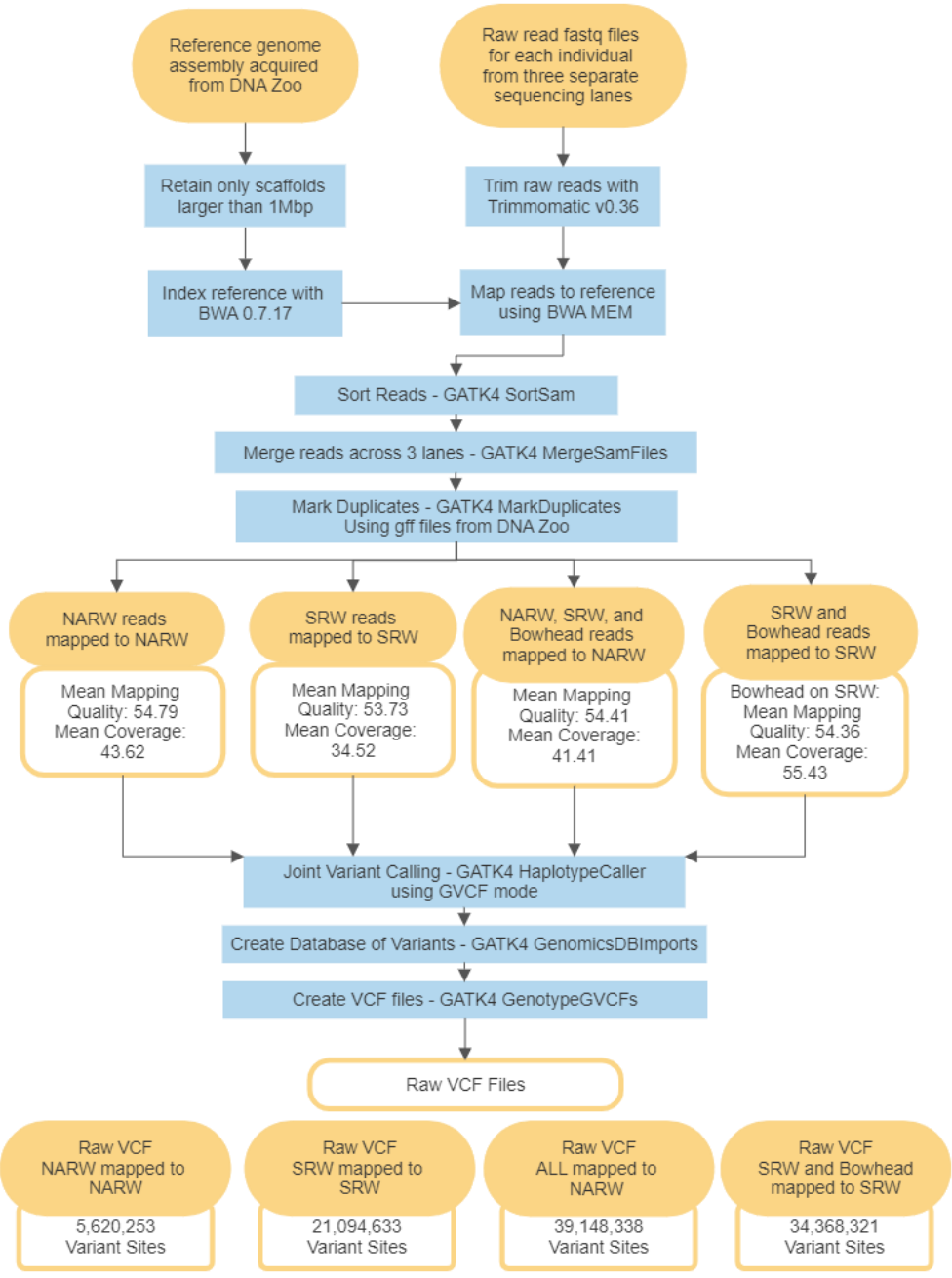

Figure A3.1. Variant (SNP and INDEL) calling pipeline used in this study.

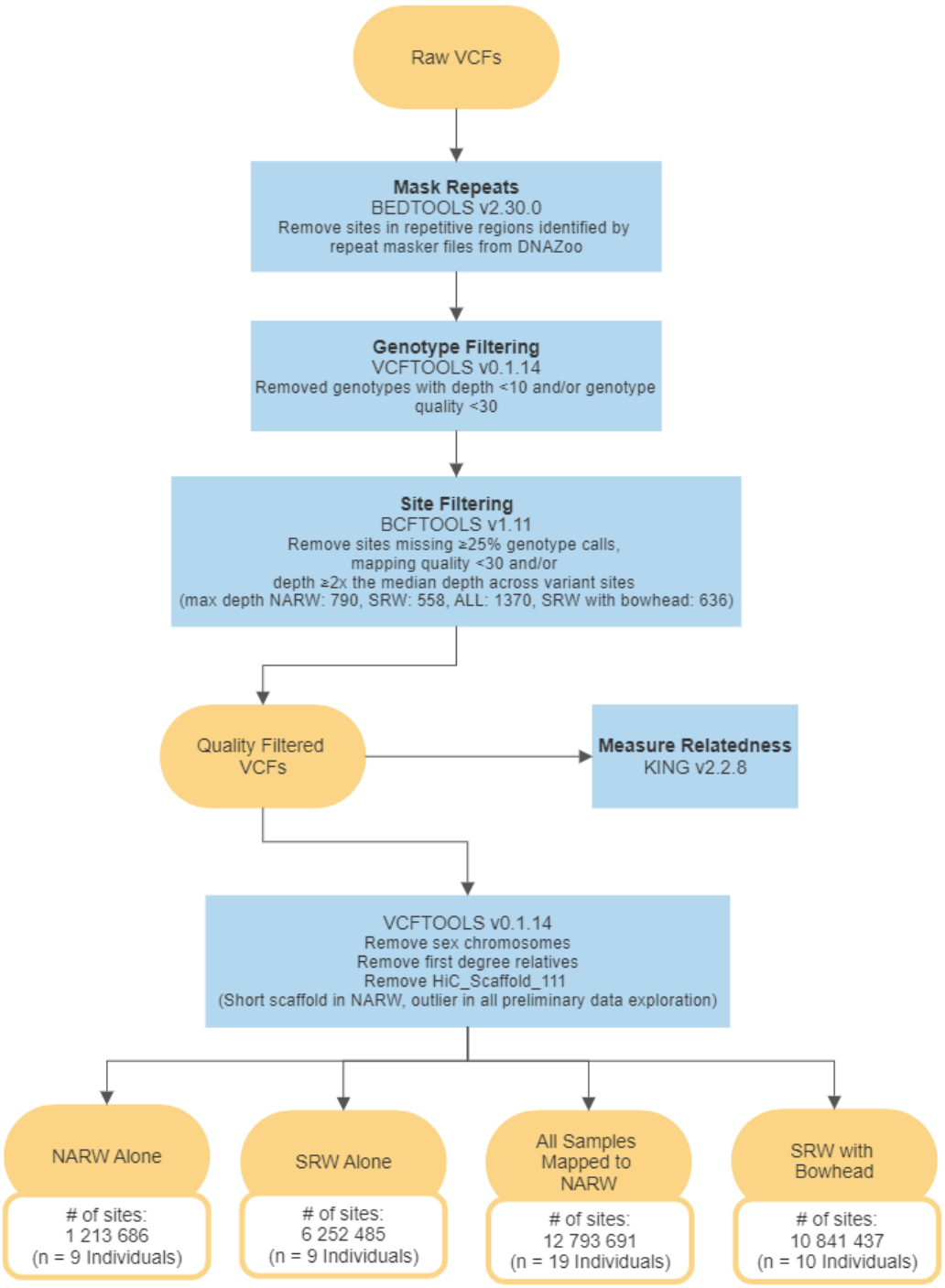

Figure A3.2. Variant (SNP and INDEL) filtration pipeline used in this study.

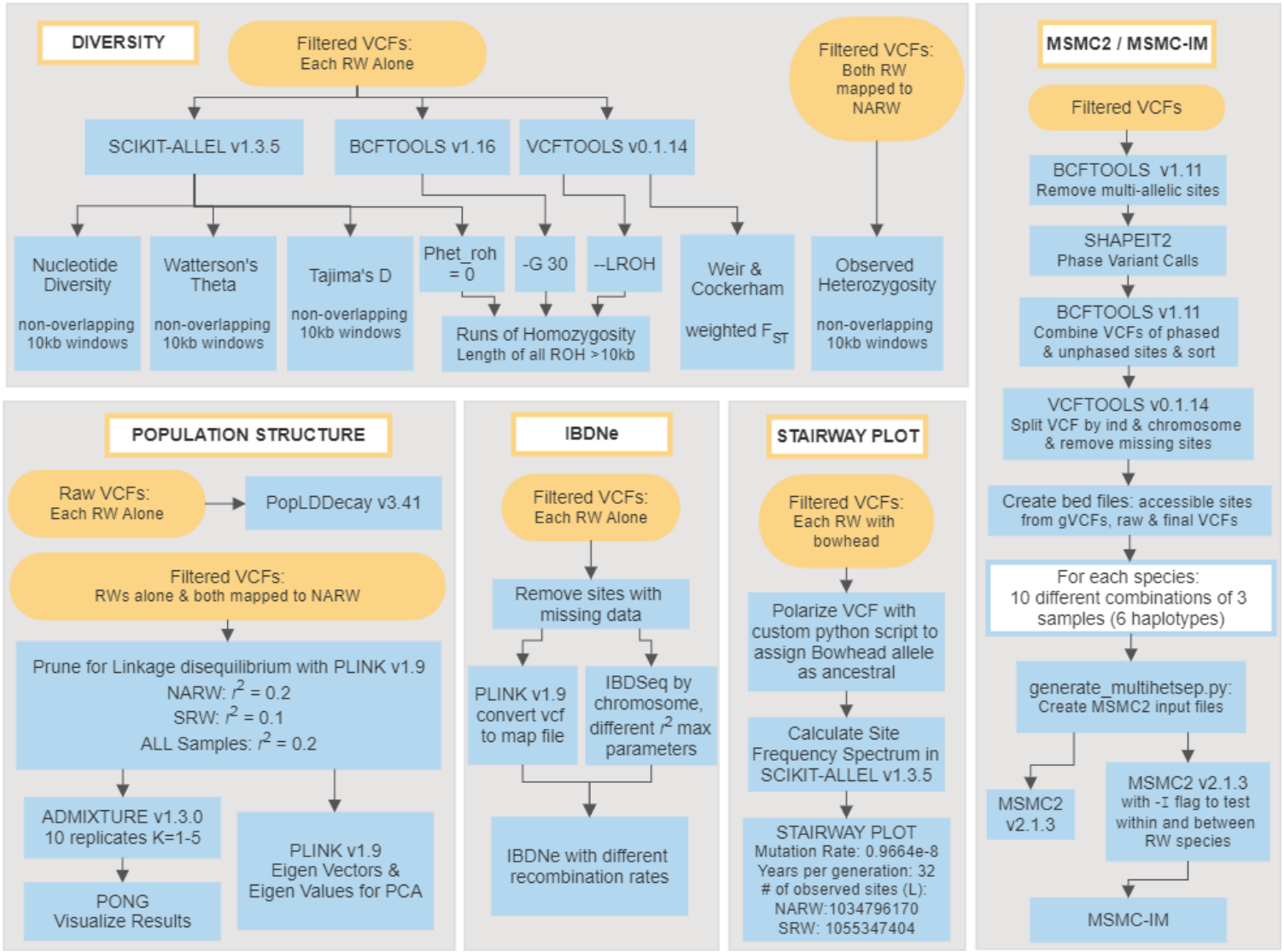

Figure A3.3. Schematic of tests performed in this study.

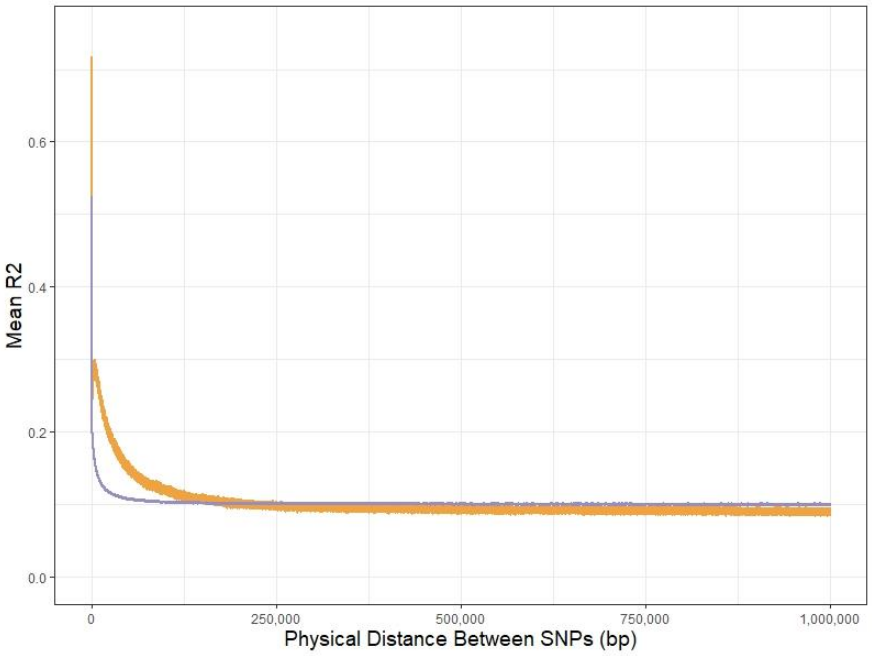

Figure A3.4. Decay in linkage between SNPs generated by PopLDdecay for North Atlantic (orange) and southern right whales (purple).

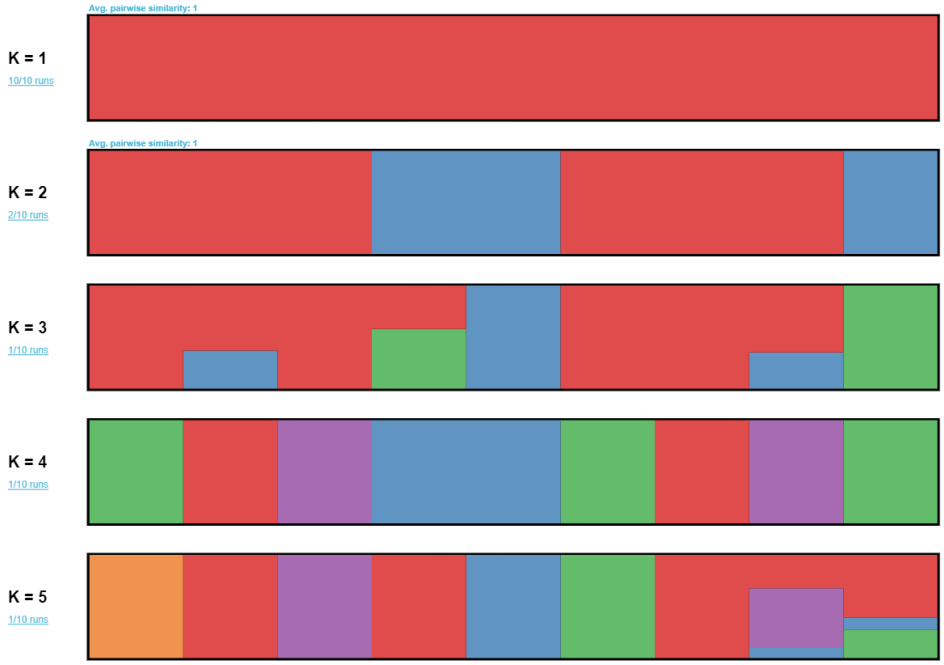

Figure A3.5. Visualization of 10 runs of ADMIXTURE from PONG for North Atlantic right whales. There is a lack of convergence across values of K>1.

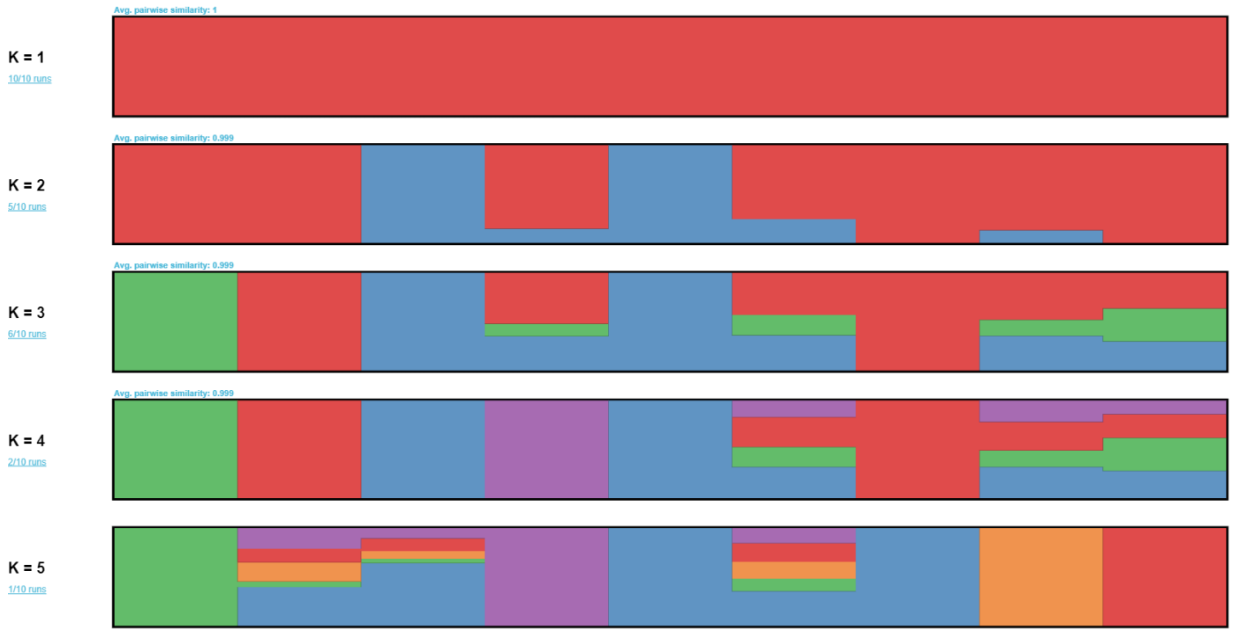

Figure A3.6. Visualization of 10 runs of ADMIXTURE from PONG for southern right whales. There is a lack of strong convergence across values of K>1.

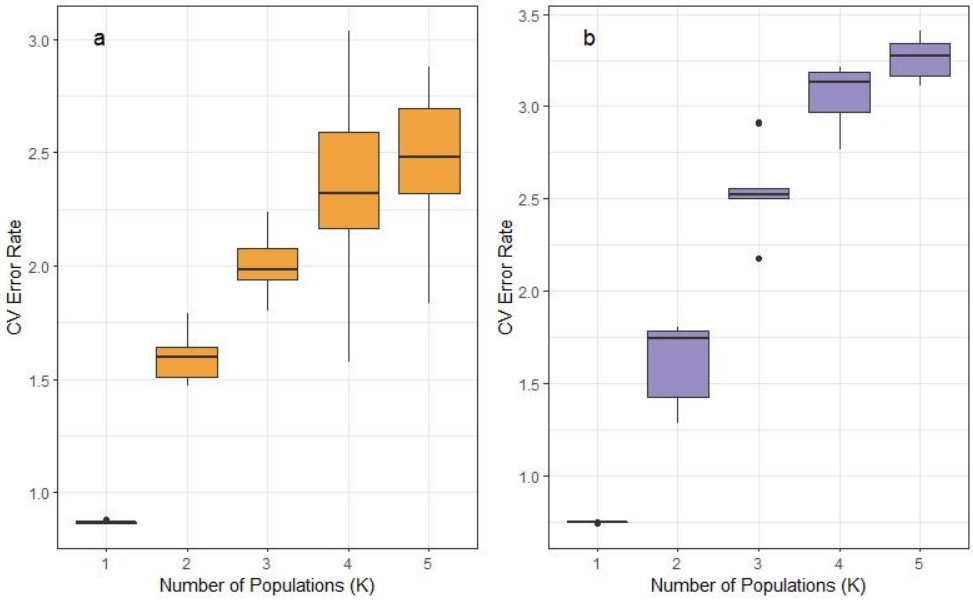

Figure A3.7. Cross validation error rate of 10 iterations of ADMIXTURE for different values of K for a) North Atlantic right whales and b) southern right whales.

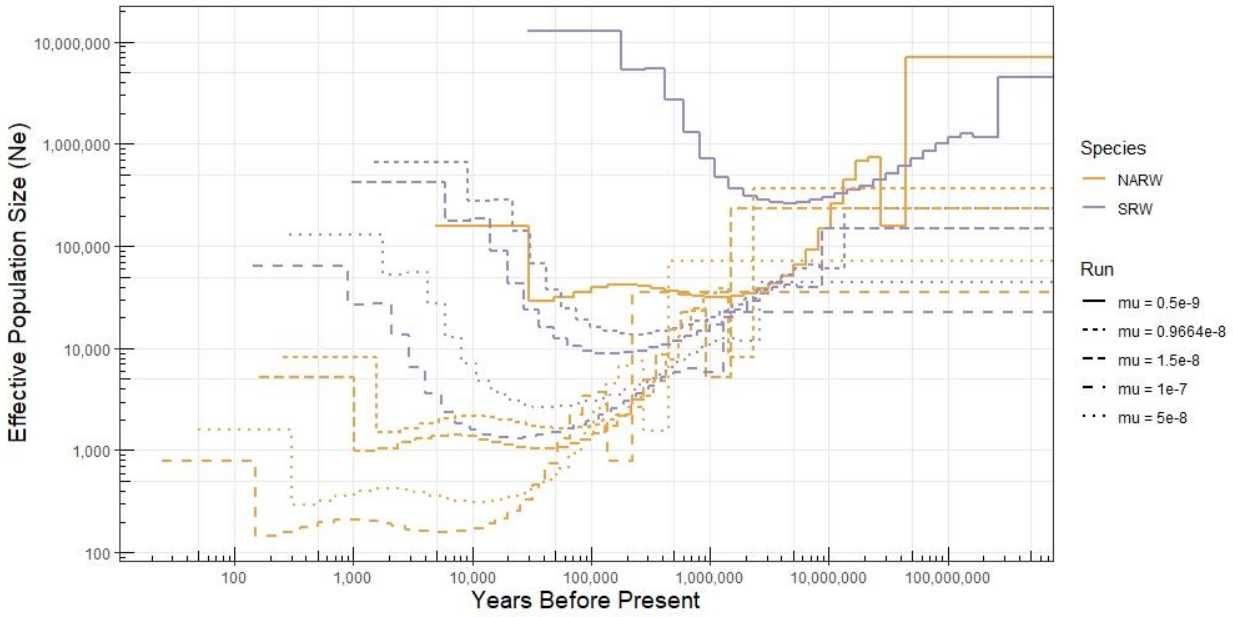

Figure A3.8. Effect of different mutation rates on the results of one run of MSMC on North Atlantic (orange) and southern right whales (purple). The timing of events and the effective population sizes are rescaled by the choice of mutation rate, but the broad pattern of changes in population size over time remains the same. Based on previous estimates for mysticetes (Dornburg et al. 2012), we used a mutation rate of 0.9664e-8 mutations/site/generation in this study.

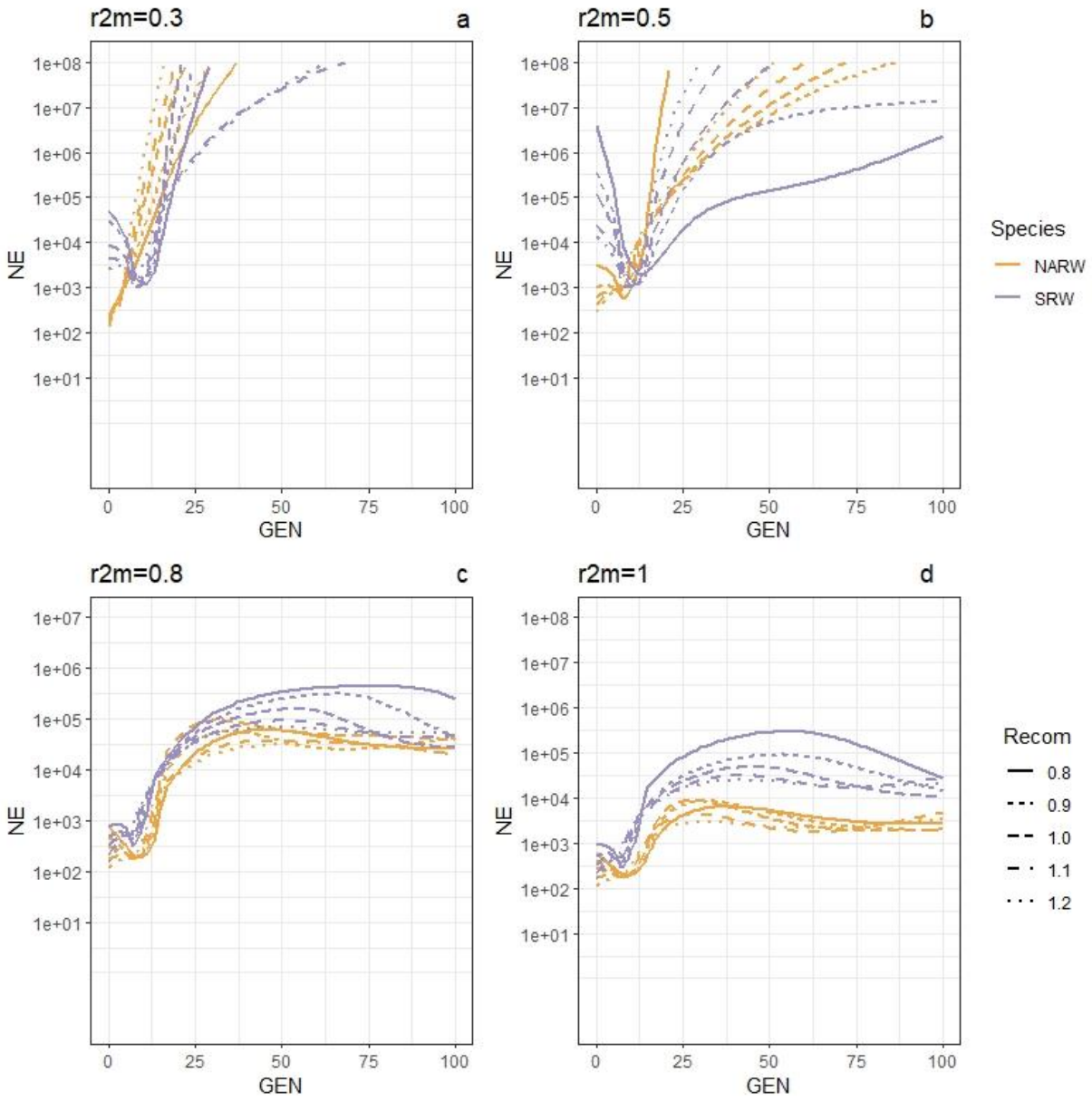

Figure A3.9. Effect of recombination rate and r^2max parameter choice on estimates of effective population size (N_e) of North Atlantic (orange) and southern (purple) right whales generated by IBDNe.

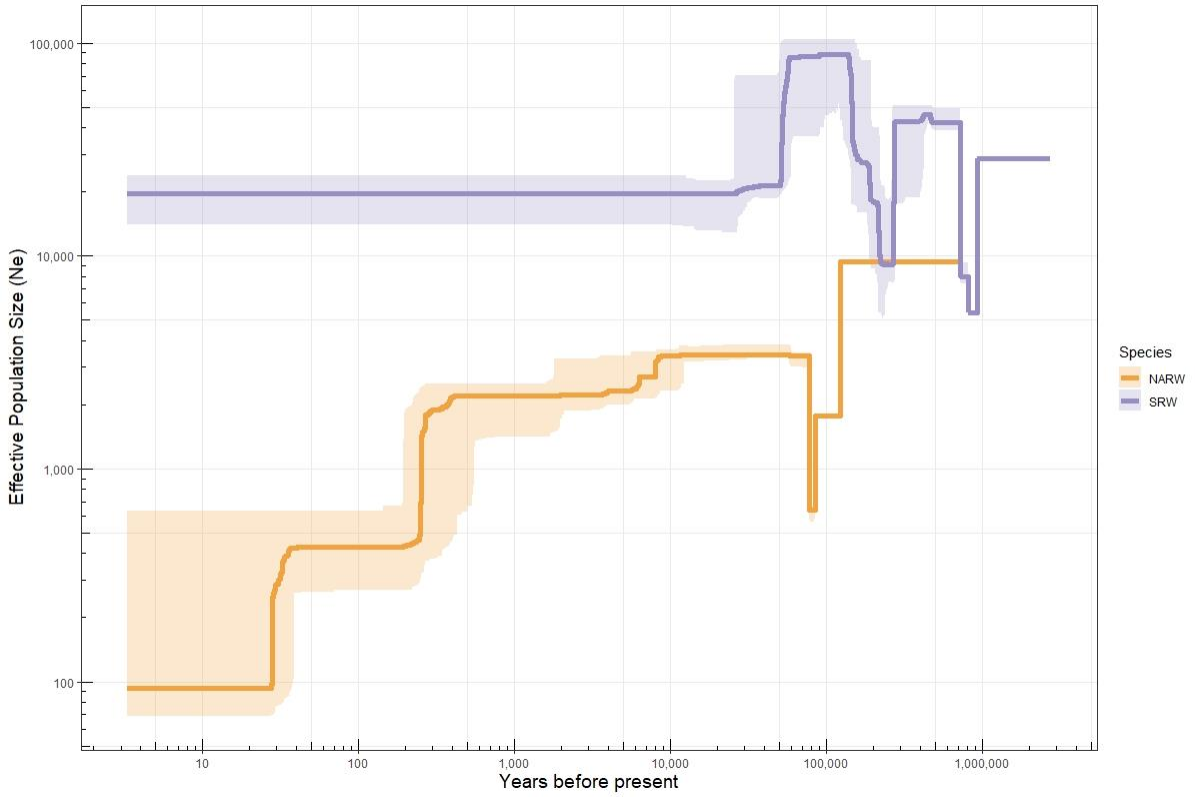

Figure A3.10. Estimates of effective population size of North Atlantic (orange) and southern right whales (purple) through time based on the site frequency spectrum in STAIRWAY PLOT2 excluding one southern right whale sample (Eau283) collected from South Georgia to confirm our main findings were not influenced by subtle population structure. The shaded regions represent the 2.5%-97.5% confidence limits for 200 estimates.

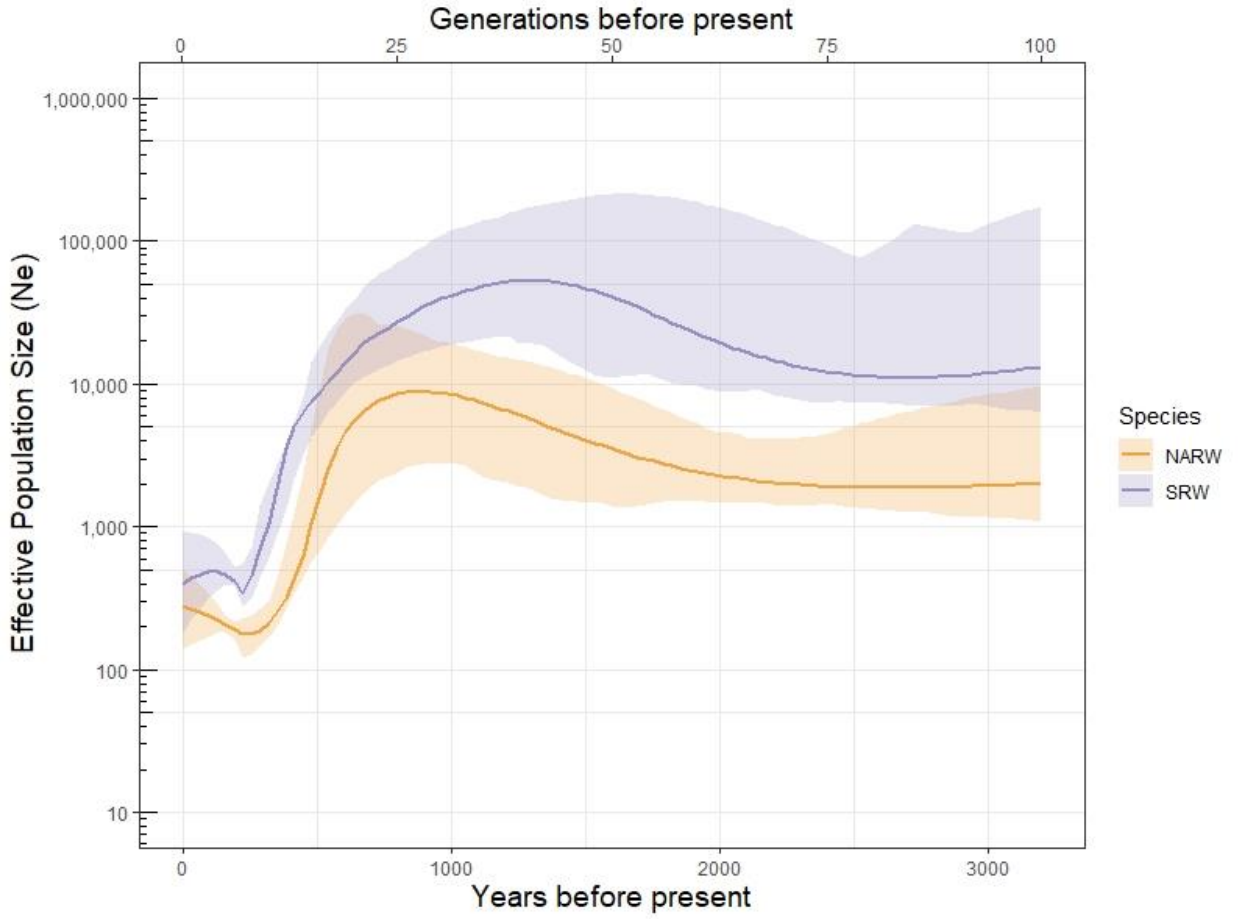

Figure A3.11. Effective population size in recent history as estimated by IBDNe for North Atlantic (orange) and southern right whale (purple) excluding one southern right whale sample (Eau283) collected from South Georgia to confirm our results were not influenced by subtle population structure. Key parameters in this additional test were the same as presented in the main text: a constant recombination rate of 1.0 cm/Mb, an r2max = 1.0 and a generation time of 32 years. 95% confidence intervals are depicted by the shaded areas.

Table A3.1. Descriptive statistics for samples used in this study. The latter statistics pertain to results from mapping to their respective species' reference assembly.

Sample ID	Ind ID	Species	Sex	Sample Collection Location	Sample Collection Date	NCBI Accession ID	Total number of raw reads	Sequencing Depth (mean±SD)	% of genome covered by ≥10X	Variable sites with genotype calls passing filters	Ind hetero-zygosity (F)
Egl 00252-1	NEA 1706	NARW	F	Bay of Fundy 44.68833, -66.46167	Aug 4, 1997	SRR22863755	357,409,922	41.45 ± 2102.56	99.02	1,198,058	-0.08305
Egl 254-1	NEA 1209	NARW	F	Bay of Fundy 44.585, -66.63167	Aug 18, 1997	SRR22863754	461,289,854	53.92 ± 2479.9	99.06	1,200,876	-0.09288
Egl 308-1a	NEA 1968	NARW	F	Bay of Fundy 44.64667, -66.37333	Sept 21, 1998	SRR22863743	412,868,996	47.56 ± 2223.27	97.89	1,195,032	-0.10973
Egl 013-3qa	NEA 1027	NARW	F	Cape Cod Bay 41.88333, -70.11833	Mar 19, 1997	SRR22863740	157,879,013	18.16 ± 858.76	81.18	942,617	-0.13296
Egl 183-1	NEA 2040	NARW	F	Bay of Fundy 44.55, -66.41333	Aug 27, 1995	SRR22863739	738,310,854	85.59 ± 5541.17	99.10	1,202,668	-0.05865
Egl 276-1†	NEA 1934	NARW	F	Bay of Fundy, 44.605, -66.515	Sept 11, 1997	SRR22863738	233,564,107	26.92 ± 1136.15	98.31	–	–
Egl 312-1a†	NEA 1240	NARW	F	Bay of Fundy 44.68167, -66.50833	Aug 17, 1995	SRR22863737	443,996,567	50.61 ± 3061.98	97.98	–	–
Egl 336_1b	NEA 1315	NARW	F	Florida, 30.635, -81.22667	Jan 12, 1998	SRR22863736	378,965,079	43.01 ± 2696.75	83.22	955,357	-0.10277
SID 179132	NEA 1204	NARW	F	Florida 30.45833, -81.14167	Mar 7, 2009	SRR22863735	453,242,967	52.66 ± 2371.98	99.07	1,200,280	-0.08794
SID 181803†	NEA 1334	NARW	F	Florida 30.27573, -81.30452	Feb 21, 2013	SRR22863734	245,549,525	28.61 ± 1541.08	95.59	–	–
Egl 140-1	NEA 1131	NARW	M	Roseway Basin, 42.89333, -65.36167	Sept 10, 1991	SRR22863753	297,155,891	34.51 ± 1763.5	98.46	1,191,098	-0.10677

Sample ID	Ind ID	Species	Sex	Sample Collection Location	Sample Collection Date	NCBI Accession ID	Total number of raw reads	Sequencing Depth (mean±SD)	% of genome covered by ≥10X	Variable sites with genotype calls passing filters	Ind hetero-zygosity (F)
Egl 272-1	NEA 1037	NARW	M	Bay of Fundy, 44.61, -66.43	Aug 18, 1997	SRR22863752	348,623,351	40.49 ± 1953.07	98.90	1,197,656	-0.10701
Eau 7	N/A	SRW	M	Argentina	1988	SRR22863751	184,153,280	19.66 ± 1357.24	74.48	5,799,668	-0.00221
Eau 9c	N/A	SRW	F	Peninsula Valdes	1989	SRR22863750	226,812,439	24.73 ± 1532.19	91.54	6,013,151	-0.00217
Eau 10b	N/A	SRW	M	Peninsula Valdes	1989	SRR22863749	518,434,524	54.39 ± 2983.71	93.89	6,048,321	-0.61309
Eau 017	N/A	SRW	F	Peninsula Valdes	1989	SRR22863748	394,616,387	42.48 ± 2617.96	93.14	6,239,654	0.01471
Eau 018	N/A	SRW	M	Peninsula Valdes	1989	SRR22863747	303,091,742	32.15 ± 2518.82	80.42	5,997,087	-0.00114
Eau 019[†]	N/A	SRW	F	Peninsula Valdes	1989	SRR22863746	496,312,699	50.16 ± 4145.34	79.43	–	–
Eau 023	N/A	SRW	M	Peninsula Valdes	1989	SRR22863745	345,538,989	36.6 ± 2786.91	82.92	6,175,009	0.01386
Eau 029	N/A	SRW	F	Peninsula Valdes	1989	SRR22863744	232,417,002	24.81 ± 1896.21	76.12	5,955,205	0.00860
Eau 034A	N/A	SRW	F	Peninsula Valdes	1989	SRR22863742	419,385,099	44.58 ± 4889.47	88.64	6,206,385	0.02182
Eau 283	N/A	SRW	M	South Georgia	1989	SRR22863741	148,959,155	15.6 ± 883.38	75.28	4,845,130	0.02488

[†] These samples were removed from analyses due to high relatedness with another sample (equivalent to first-order relatives or closer).

Table A3.2. Descriptive statistics of the reference genomes used in this study

	North Atlantic right whale	Southern right whale
Total Scaffold Length	2,369,417,546 bp	2,316,908,615 bp
Number of Scaffolds	172,124	3,234
Scaffold N50	101,413,572 bp	112,042,483 bp
Number of Scaffolds >1Mb	23	21
Size of Reference Genome Scaffolds >1Mb	2,170,759,585 bp	2,296,311,778 bp

Table A3.3. F_{ROH} estimates generated with different methods and different thresholds for identifying ROHs.

		BCFTOOLS			SCIKIT-ALLEL			VCFTOOLS		
SAMPLE	Species	$F_{ROH\ 100Kb}$	$F_{ROH\ 300Kb}$	$F_{ROH\ 1Mb}$	$F_{ROH\ 100Kb}$	$F_{ROH\ 300Kb}$	$F_{ROH\ 1Mb}$	$F_{ROH\ 100Kb}$	$F_{ROH\ 300Kb}$	$F_{ROH\ 1Mb}$
EGL00252-1	NARW	0.095	0.058	0.021	0.136	0.038	0.007	0.069	0.034	0.009
EGL013-3QA	NARW	0.064	0.036	0.009	0.166	0.031	0.003	0.054	0.026	0.003
EGL140-1	NARW	0.080	0.042	0.010	0.121	0.024	0.001	0.054	0.020	0.002
EGL183-1	NARW	0.119	0.081	0.029	0.165	0.057	0.008	0.089	0.053	0.008
EGL254-1	NARW	0.093	0.054	0.016	0.132	0.034	0.004	0.061	0.031	0.007
EGL272-1	NARW	0.071	0.034	0.001	0.113	0.014	0.000	0.047	0.016	0.000
EGL308-1A	NARW	0.091	0.057	0.015	0.135	0.039	0.003	0.070	0.038	0.006
EGL336_1B	NARW	0.071	0.041	0.015	0.170	0.033	0.005	0.054	0.024	0.004
SID179132	NARW	0.088	0.051	0.015	0.131	0.034	0.006	0.063	0.029	0.004
MEAN NARW		0.086	0.050	0.014	0.141	0.034	0.004	0.062	0.030	0.005
SD NARW		0.017	0.015	0.008	0.021	0.012	0.003	0.012	0.011	0.003
EAU017	SRW	0.020	0.006	0.001	0.009	0.002	0.000	0.008	0.002	0.000
EAU018	SRW	0.019	0.003	0.001	0.010	0.001	0.000	0.009	0.000	0.000
EAU023	SRW	0.022	0.006	0.004	0.012	0.003	0.000	0.010	0.003	0.001
EAU029	SRW	0.019	0.005	0.000	0.010	0.001	0.000	0.008	0.001	0.000
EAU034A	SRW	0.025	0.011	0.003	0.015	0.006	0.002	0.014	0.007	0.002
EAU10B	SRW	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
EAU283	SRW	0.050	0.037	0.024	0.043	0.025	0.005	0.024	0.014	0.005
EAU7	SRW	0.020	0.005	0.000	0.011	0.001	0.000	0.008	0.001	0.000
EAU9C	SRW	0.027	0.013	0.007	0.014	0.003	0.000	0.009	0.003	0.001
MEAN SRW		0.023	0.010	0.004	0.014	0.005	0.001	0.010	0.004	0.001
SD SRW		0.013	0.011	0.008	0.012	0.008	0.002	0.006	0.004	0.002

Table A3.4. Genome-wide nucleotide diversity estimates for cetaceans used in Figure 3.5. Data were compiled in Robinson et al. (2016) and Morin et al. (2021), except those from right whales generated in this study.

Species	Nucleotide diversity (π)	Original Source(s)
Vaquita (*Phocoena sinus*)	0.00010	Morin et al. 2021
Baiji (*Lipotes vexillifer*)	0.00012	Zhou et al. 2013.
Narwhal (*Monodon monoceros*)	0.00014	Westbury et al. 2019.
North Atlantic right whale (*Eubalaena glacialis*)	0.00019	This study
Killer whale (*Orcinus orca*)	0.00021	Westbury et al. 2018.
Beluga whale (*Delphinapterus leucas*)	0.00029	Westbury et al. 2019.
Minke whale (*Balaenoptera acutorostrata*)	0.00061	Yim et al. 2014.
Southern right whale (*Eubalaena australis*)	0.00080	This study
Finless porpoise (*Neophocaena phocaenoides*)	0.00086	Yim et al. 2014.
Indo-Pacific finless porpoise (*Neophocaena phocaenoides*)	0.00093	Morin et al. 2021. Zhou et al. 2018.
Yangtze finless porpoise (*Neophocaena asiaeorientalis asiaeorientalis*)	0.00105	Morin et al. 2021. Zhou et al. 2018.
Narrow-ridged finless porpoise (*Neophocaena asiaeorientalis*)	0.00134	Morin et al. 2021. Zhou et al. 2018.
Bottlenose dolphin (*Tursiops truncatus*)	0.00142	Yim et al. 2014.
Fin whale (*Balaenoptera physalus*)	0.00151	Yim et al. 2014.
Blue whale (*Balaenoptera musculus*)	0.00210	Morin et al. 2021.
Sperm whale (*Physeter macrocephalus*)	0.00228	Morin et al. 2021.; Fan et al. 2019.

Table A3.4 References

Fan G, Zhang Y, Liu X, Wang J, Sun Z, Sun S, *et al.* (2019). The first chromosome-level genome for a marine mammal as a resource to study ecology and evolution. *Mol Ecol Resour* **19**: 944–956.

Morin PA, Archer FI, Avila CD, Balacco JR, Bukhman Y V., Chow W, *et al.* (2021). Reference genome and demographic history of the most endangered marine mammal, the vaquita. *Mol Ecol Resour* **21**: 1008–1020.

Robinson JA, Ortega-Del Vecchyo D, Fan Z, Kim BY, Vonholdt BM, Marsden CD, *et al.* (2016). Genomic Flatlining in the Endangered Island Fox. *Curr Biol* **26**: 1183–1189.

Westbury M V., Hartmann S, Barlow A, Wiesel I, Leo V, Welch R, *et al.* (2018). Extended and continuous decline in effective population size results in low genomic diversity in the world's rarest hyena species, the brown Hyena. *Mol Biol Evol* **35**: 1225–1237.

Westbury M V., Petersen B, Garde E, Heide-Jørgensen MP, Lorenzen ED (2019). Narwhal Genome Reveals Long-Term Low Genetic Diversity despite Current Large Abundance Size. *iScience* **15**: 592–599.

Yim HS, Cho YS, Guang X, Kang SG, Jeong JY, Cha SS, *et al.* (2014). Minke whale genome and aquatic adaptation in cetaceans. *Nat Genet* **46**: 88–92.

Zhou X, Sun F, Xu S, Fan G, Zhu K, Liu X, *et al.* (2013). Baiji genomes reveal low genetic variability and new insights into secondary aquatic adaptations. *Nat Commun* **4**: 2708.

Zhou X, Guang X, Sun D, Xu S, Li M, Seim I, *et al.* (2018). Population genomics of finless porpoises reveal an incipient cetacean species adapted to freshwater. *Nat Commun* **9**: 1276

A4. Effects of inbreeding on reproductive success in endangered North Atlantic right whales

A4.1 Supplementary Methods

Reference Genome

The draft reference assembly of the North Atlantic right whale genome was obtained from DNAZoo (www.DNAZoo.org) in March 2023. We are grateful for the work that went into this assembly and would like to acknowledge their work as requested:

> "The draft assembly was generated by the DNA Zoo team from short insert-size PCR-free DNA-Seq data using w2rap-contigger (Clavijo et al. 2017), see (Dudchenko et al., 2018) for details. This work was performed under Marine Mammal Health and Stranding Response Program (MMHSRP) Permit No. 18786-03 issued by the National Marine Fisheries Service (NMFS) under the authority of the Marine Mammal Protection Act (MMPA) and Endangered Species Act (ESA). The specimen used in this study was collected by NOAA/T. Rowles/B. Bonde and provided by the National Marine Mammal Tissue Bank, which is maintained by the National Institute of Standards and Technology (NIST) in the NIST Biorepository, and which is operated under the direction of NMFS with the collaboration of USGS, USFWS, MMS, and NIST through the Marine Mammal Health and Stranding Response Program.
>
> Dudchenko, O., Batra, S.S., Omer, A.D., Nyquist, S.K., Hoeger, M., Durand, N.C., Shamim, M.S., Machol, I., Lander, E.S., Aiden, A.P., Aiden, E.L., 2017. De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. Science 356, 92–95. https://doi.org/10.1126/science.aal3327.
>
> Dudchenko, O., Shamim, M.S., Batra, S., Durand, N.C., Musial, N.T., Mostofa, R., Pham, M., Hilaire, B.G.S., Yao, W., Stamenova, E., Hoeger, M., Nyquist, S.K., Korchina, V., Pletch, K., Flanagan, J.P., Tomaszewicz, A., McAloose, D., Estrada, C.P., Novak, B.J., Omer, A.D., Aiden, E.L., 2018. The

Juicebox Assembly Tools module facilitates de novo assembly of mammalian genomes with chromosome-length scaffolds for under $1000. bioRxiv 254797. https://doi.org/10.1101/254797."

[www.dnazoo.org/assemblies/Eubalaena_glacialis]

Restriction Enzyme Selection

To estimate the number of fragments we would obtain with different combinations of restriction enzymes and in different size classes, we simulated digestion of the North Atlantic right whale genome using the package simRAD (Lepais and Weir 2014) in R v3.6.0. We selected NlaIII and EcoRI as the best combination of enzymes to maximize the proportion of fragments cut by both enzymes that would be retained through library preparation (Maximize A+B : A+A fragment ratio). A+A fragments will bind to the flow cell, but will not amplify during bridge amplification, so we want to minimize the proportion of fragments with A+A cut sites in the final library. B+B fragments are not a concern as they will not amplify during indexing due to the Y-shaped adapter.

Annealing Single-stranded Adapters

In order to bind unique indexes to each sample, we needed to ligate double stranded adapters onto the cut sites of each fragment that are complementary to the end of the indexing sequences. We annealed single stranded adapters together to a final concentration of 10μM for the NlaIII adapter and 1μM for the EcoRI adapter and

included NaCl in each reaction at a final concentration of 50mM. Single-stranded oligos were annealed in a thermocycler at 95°C for 5 min, followed by a ramp-down of 2°C/min to 21°C. To test that annealing took place, we ran annealed (double-stranded) and single stranded oligos on an agarose gel. Ethidium bromide should bind to double stranded DNA better and the double stranded adapters will have a higher molecular weight. Therefore, if the primers have annealed properly, they should be brighter and appear larger than the single stranded oligos.

Restriction Enzyme Digest & Adapter Ligation

A schematic of the entire library preparation steps is depicted in Figure A4.1. For each sample, 400ng of DNA was digested with 10U NlaIII and 20U EcoRI-HF (New England Biolabs, NEB; Table A4.1) at 37°C for one hour followed by 20 minutes at 65°C to denature the enzymes.

Table A4.1 Restriction enzyme digest reaction

Component	***Initial Concentration***	***Desired Concentration***	***Volume to add (µL)***
DNA (400ng)	40 ng/µL		10 µL
CutSmart Buffer	10X	1X	3.0 µL
NlaIII	10 000 U/mL	10 units	1.0 µL
EcoRI-HF	20 000 U/mL	20 units	1.0 µL
Nuclease-free Water		to 30 µL	15 µL

The required amount of each adapter is determined based on the number of cut sites generated by each restriction enzyme as estimated by simRAD (Lepais and Weir 2014). Each sample required 4.5 x 10-5 μmol of adapter for sites cut by NlaIII (common cutter) and 3.5 x 10-6 μmol of adapter for sites cut by EcoRI-HF (rare cutter). We ligated these adapters with 800U of T4 DNA ligase in a 50μL reaction (Table A4.2) at room temperature for one hour followed by 65°C for 10 minutes. We cleaned each reaction with Ampure XP beads (1.6:1.0 beads : sample ratio) and eluted into 17.5uL of Tris-HCl.

Table A4.2 Ligation reaction of double stranded adapters onto digested DNA

Component	***Initial Concentration***	***Desired Concentration***	***Volume to add (μL)***
Digested DNA			30μL
T4 Ligase buffer	10X	1X	5.0μL
P1 EcoRI adapter	1μM	3.5×10^{-6} μmol	3.5μL
P2 NlaIII adapter	10μM	4.5×10^{-5} μmol	4.5μL
T4 DNA ligase	400 U/μL	800U	2.0μL
Nuclease-free water		to 50μL	5μL

Indexing & Size Selection

We dual-indexed each sample with a unique combination of Nextera xt indexes (Illumina; Table A4.3) and performed another bead clean up with Ampure XP beads with a 1.6:1.0 beads to sample ratio. We pooled pairs of samples and ran each pool in a separate lane of a Pippin Prep (2% agarose cassettes with ethidium bromide) selecting

fragments 440bp-540bp in size (fragment size without adapters: 293 bp – 393 bp). Eluted product from two lanes of the Pippin Prep were combined for a final bead clean-up using approximate bead ratios of 1.8-1.3 : 1.0 and eluted into 22.5uL of Tris-HCl. Four of the original pools (each representing eight individually barcoded samples) were combined prior to normalizing.

Table A4.3 Index reaction

Component	***Volume per Reaction (µL)***
Digested DNA with ligated adapters	15µL
P5 adapter	5µL
P7 adapter	5µL
KAPA HiFi HotStart ReadyMix	25µL

Read Mapping

Figure A4.2 provides an overview of the bioinformatic pipelines carried out from raw reads to variant calling. We prepared the reference genome using bwa index and SAMtools (Li et al. 2009). We used bwa-mem followed by SAMtools view to generate bam files for each sample.

Variant Calling

Figure A4.3 provides details of each step and key parameters used in our variant calling pipeline. We carried out variant calling on four data sets: using Freebayes with two

minimum genotype depth thresholds (Garrison and Marth 2012) and two different alpha thresholds in the stacks_ref pipeline (Rochette et al. 2019).

We ran gstacks with the bam files generated by bwa-mem using two different alpha thresholds for detecting variants and calling genotypes (--var-alpha & --gt-alpha : 0.01 & 0.001). We ran the populations module for each gstacks data set requiring a locus to be present in 80% of individuals (populations -r 0.8) and generated a vcf file of all sites called (variant and invariant). We sorted for missingness by removing sites missing 60% of genotype calls, removed individuals missing genotype calls at 50% of sites, and then filtered sites again to retain only sites with <20% missing data. We sorted the vcf by position using bcftools v1.16 and removed repetitive regions identified by the RepeatMasker file that accompanied our reference genome in the DNAZoo repository. Finally, we used bcftools to retain only biallelic snps, and we removed low frequency variants (--maf 0.01 and --mac 3) with vcftools. Finally, we filtered our variants to retain only the first record of sites called when strands were called and reported separately.

In Freebayes, we called variants in consecutive 6Mbp regions of the genome and reported monomorphic sites as well. In complex regions, we included additional constraints with the --use-best-n-alleles 4 flag and in some cases the --skip-coverage 10000 flag to allow Freebayes to complete variant calling. As we did with Stacks, we

used two stringency thresholds and used vcftools (Danecek et al. 2011) to drop genotypes based on a depth of less than 5 or 10 reads, but we also excluded 11 samples (representing seven unique individuals) from subsequent analyses because mean read depth was extremely low. For both Freebayes datasets, we filtered for missingness in a multi-step process. First, we removed sites not called in at least 60% of individuals, we then concatenated the vcf files across all regions and removed individuals missing genotype calls at more than 50% of sites. Finally, we dropped sites in which less than 80% of samples had genotype calls. We removed multi-allelic sites to produce variant files for all monomorphic and bi-allelic SNPs with a mapping quality above 30. We then removed repetitive regions as identified by the supplied RepeatMasker file. We further filtered on depth to exclude repetitive or complex regions that may have been missed by removing sites with a maximum depth of greater than the mean site depth + 2 standard deviations (INFO/DP > 8732). We used the bcftools plug-in +fill-tags to correct allele frequencies and allele counts. We filtered to retain biallelic SNPs and removed one of each duplicated pair of samples. Finally, we performed a final correction of allele frequencies and allele counts, and removed rare variants where alternate allele frequencies were less than or equal to 0.01, and minimum allele count was less than or equal to 3 to ensure an allele was called in more than one individual.

De-lifing Fitness

For all females in the population, we identified the years she was alive and an adult, and the years in which she had a calf in that interval. We calculated the population fecundity each year by taking the total number of calves born each year, divided by the mean number of adult females alive each year as calculated by the Pace model (Pace et al. 2017), as it is considered to be the best age structured estimate of North Atlantic right whale population size.

For each female, we corrected her final year alive to be the final year she was sighted to take a conservative approach to calculating mean individual fecundity contribution for each female. As the typical reproductive cycle of a right whale female would be three years, we did not want to penalize a female for not having a calf in a year where many other calves were born because she was in a recovery or pregnancy year, therefore, we considered whether or not a female gave birth in a year with a sliding window approach meaning that her calving contribution for $year_i$ would consider whether or not she had a calf in $year_{i-1}$, $year_i$ or $year_{i+1}$. We calculated the mean individual fecundity contributions for each female over the years she was alive and adult (Equation A4.1). We omitted 2 females for which we had fewer than six years of fecundity values representing four years of calving data, to minimize the bias introduced by females with only one possible calving interval.

$$mean\left(\frac{reproduced_{year(i-1|i|i+1)}(1 \mid 0) - popfecundity_{year}}{NumAdultFemales_{yeari} - 1}\right)$$

Eq. A4.1 Mean annual fecundity contributions calculated for each female.

Bayesian Models

We built Bayesian models to identify the effects of inbreeding coefficients on individual fecundity for North Atlantic right whale females. As both fitness (as measured by individual fecundity) and inbreeding coefficients (F, sMLH, IR, HL or $F_{HBD>100KB}$) are continuous variables, we standardized these values as z-scores (by subtracting each value from the mean, then dividing by the standard deviation), which transformed each into a normal distribution with a mean of zero and a standard deviation of one. Therefore, the prior probabilities for β0 and β1 were normal distributions with a mean of zero and a standard deviation of one.

The model was run with a combination of R v.4.3.0 and RStan. We assessed performance of the MCMC runs by inspecting Rhat scores, effective sample size (ESS) estimates, and trace plots.

References

Danecek P, Auton A, Abecasis G, et al (2011) The variant call format and VCFtools. Bioinformatics 27:2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Garrison E, Marth G (2012) Haplotype-based variant detection from short-read sequencing. arXiv preprint arXiv:1207.3907 [q-bio.GN]

Lepais O, Weir JT (2014) SimRAD: An R package for simulation-based prediction of the number of loci expected in RADseq and similar genotyping by sequencing approaches. Mol Ecol Resour 14:1314–1321. https://doi.org/10.1111/1755-0998.12273

Li H, Handsaker B, Wysoker A, et al (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Pace RM, Corkeron PJ, Kraus SD (2017) State-space mark-recapture estimates reveal a recent decline in abundance of North Atlantic right whales. Ecol Evol 7:8730–8741. https://doi.org/10.1002/ece3.3406

Rochette NC, Rivera-Colón AG, Catchen JM (2019) Stacks 2: Analytical methods for paired-end sequencing improve RADseq-based population genomics. Mol Ecol 28:4737–4754. https://doi.org/10.1111/mec.15253

A4.2 Supplementary Results

Table A4.4 Summary results of final datasets

	Freebayes DP5	***Freebayes DP10***	***Stacks 0.01***	***Stacks 0.001***
Number of variant & invariant sites called/mapped ǂ (% of 21-scaffold assembly)	43 885 179 (2.03%)	34 358 890 (1.59%)	56 461 899 (2.61%)	56 462 262 (2.61%)
Number of biallelic SNPs passing filters	19 168	13 358	39 476	26 011
Number of individuals in final dataset	89	80	93	93
Repeatability (Duplicate pairs)	1.2 – 2.4% (n = 3)	1.4 – 1.9% (n = 2)	1.1 – 1.8% (n = 3)	0.57 – 1.36% (n = 3)

ǂ Total number of sites called after filtering for missingness, depth (Freebayes only), and mapping quality (Freebayes only) and after removing repetitive regions.

A4.2 Supplementary Figures

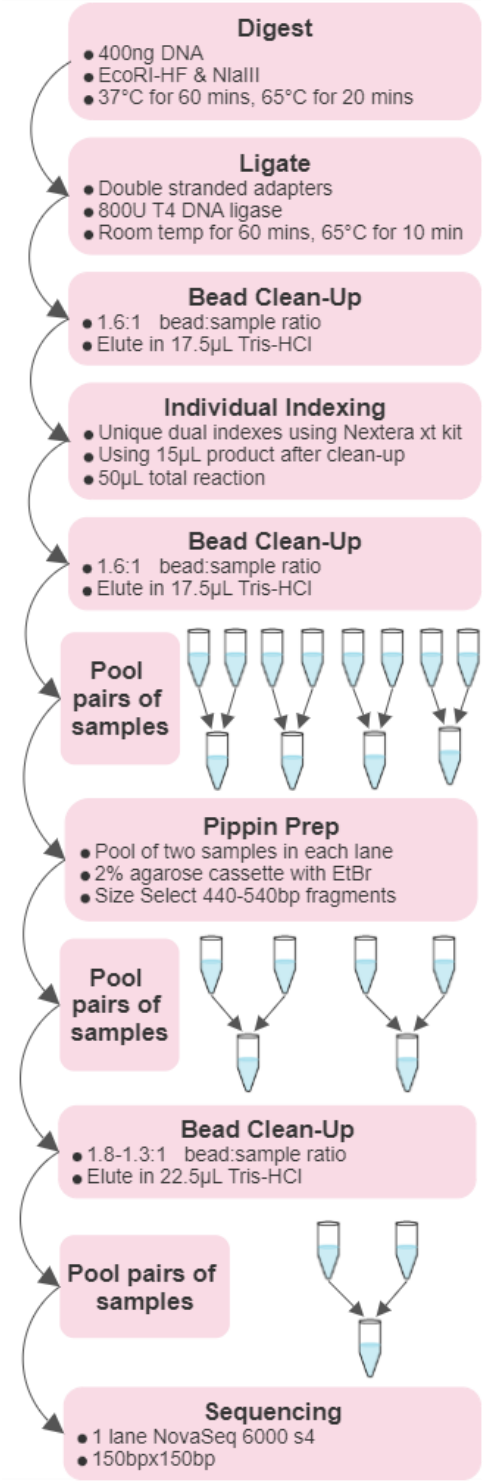

Figure A4.1 Summary of library preparation

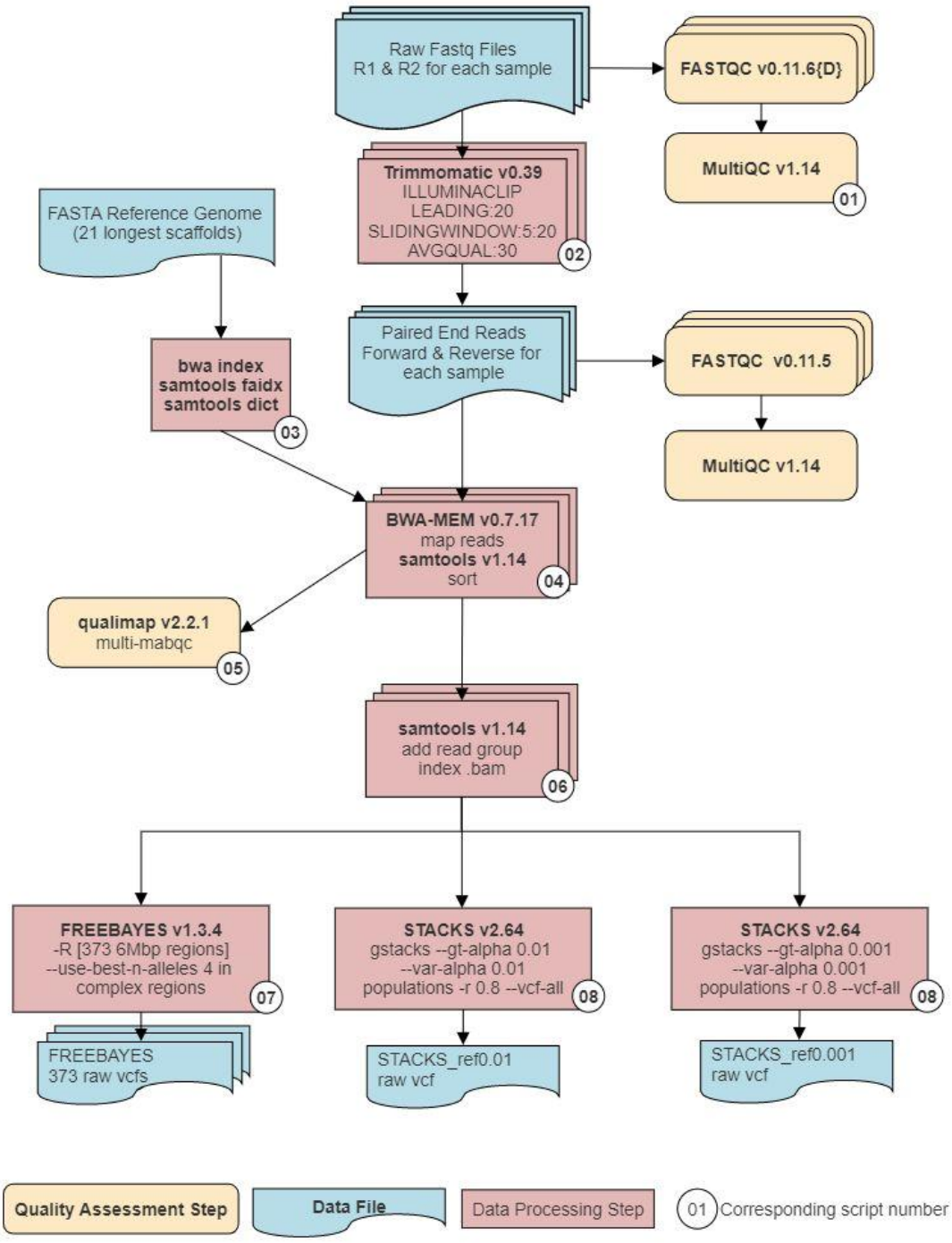

Figure A4.2 Overview of the bioinformatic pipeline and the key parameters used. Code for various steps is available at www.github.com/carlacrossman/NARW_ddRAD_InbreedingDepression

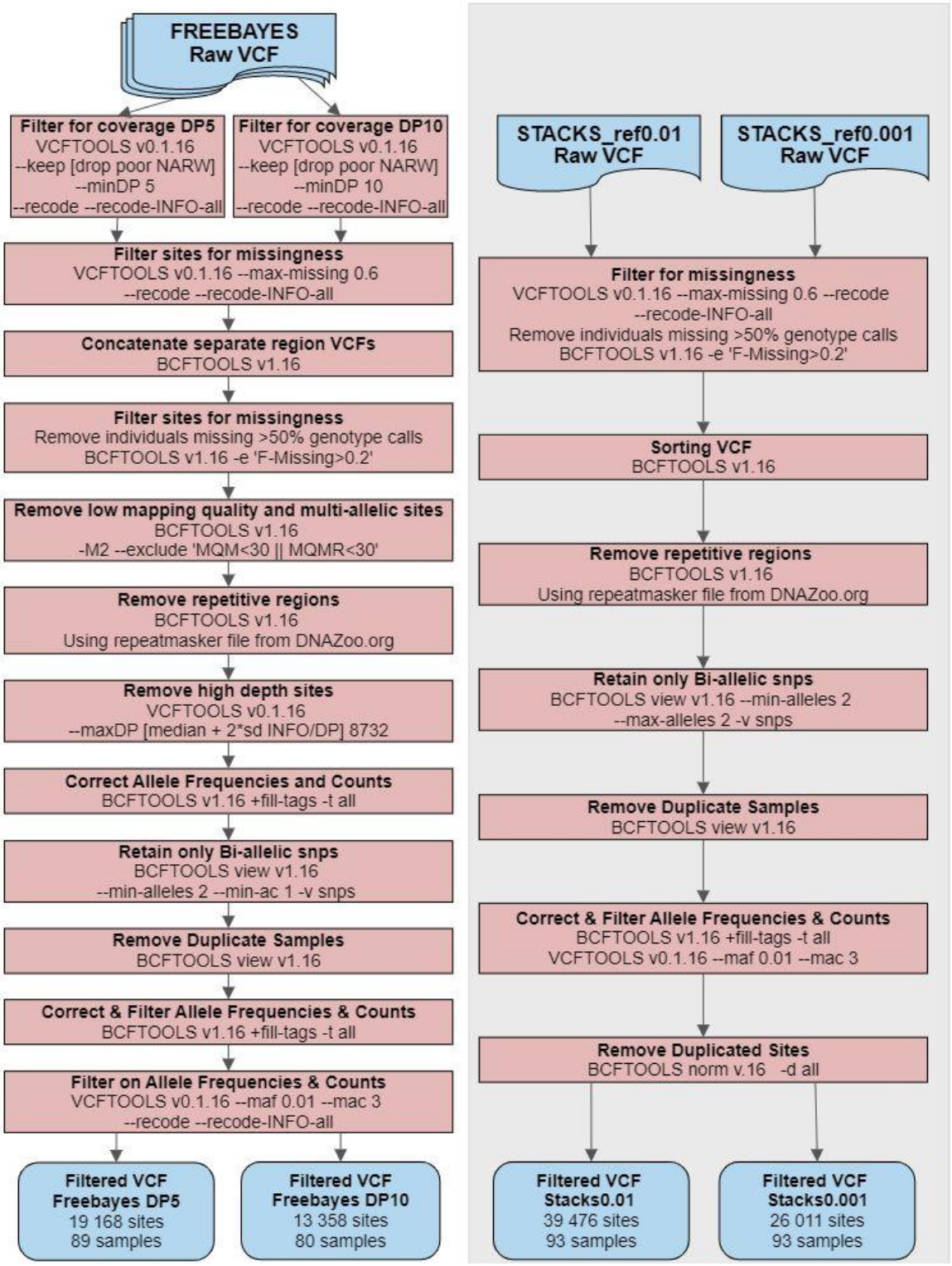

Figure A4.3 Overview of the variant filtration steps with key parameters listed. Code for various steps is available at www.github.com/carlacrossman/NARW_ddRAD_InbreedingDepression

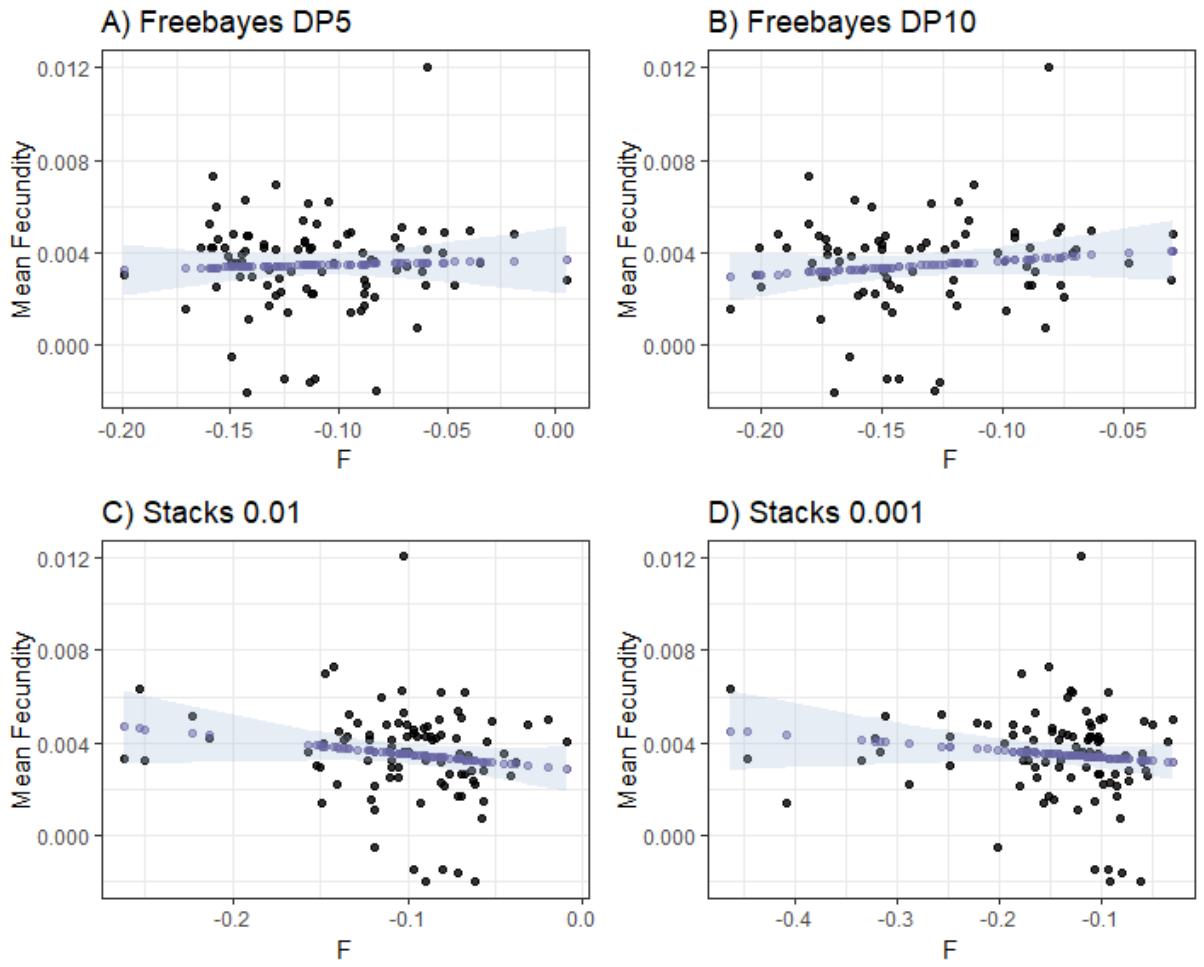

Figure A4.4. Relationship between the inbreeding coefficient F and reproductive fecundity for the A) Freebayes DP5, B) Freebayes DP10, C) Stacks 0.01 and D) Stacks 0.001 datasets. Posterior predicted means are plotted in blue and their 95% HDI are shaded.

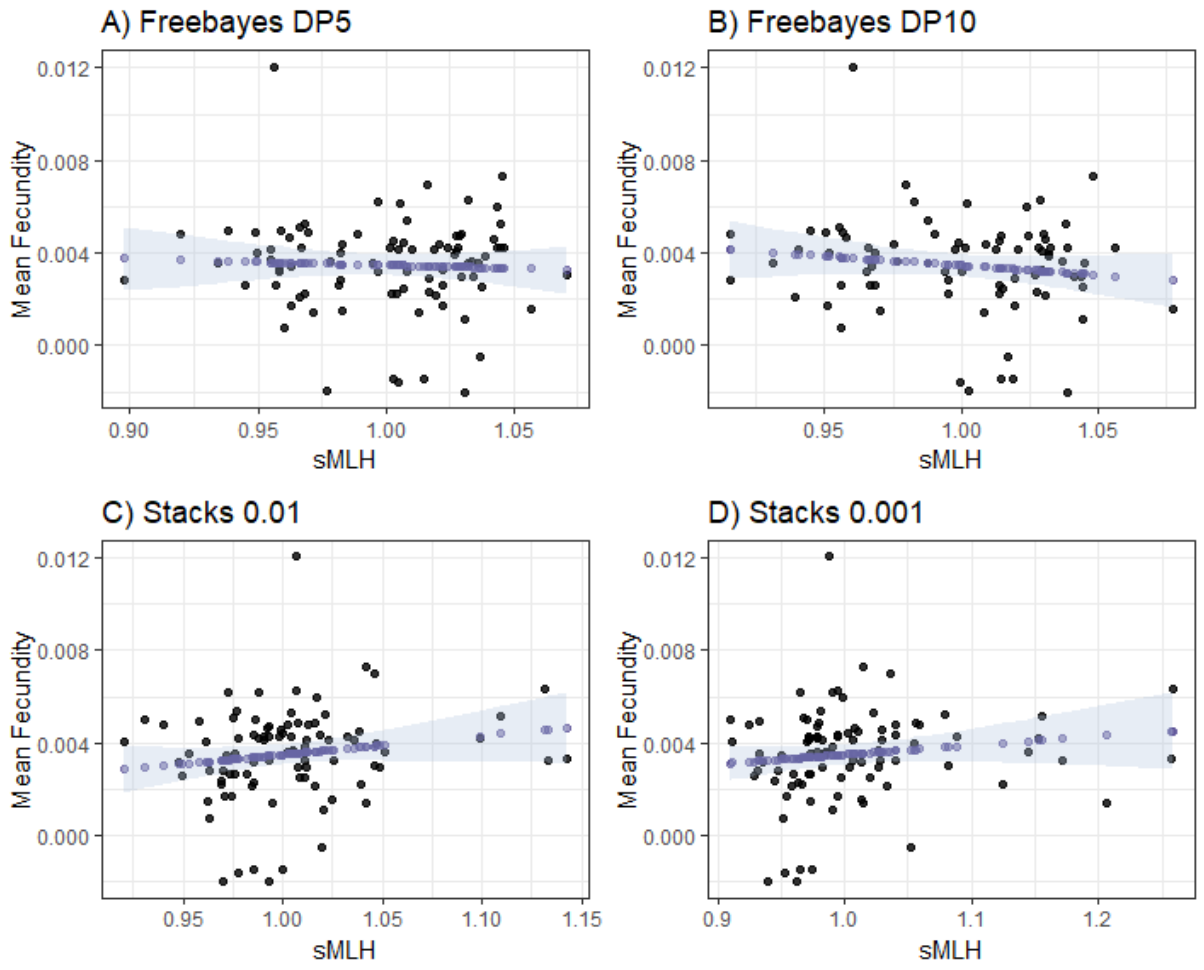

Figure A4.5. Relationship between the inbreeding coefficient sMLH and reproductive fecundity for the A) Freebayes DP5, B) Freebayes DP10, C) Stacks 0.01 and D) Stacks 0.001 datasets. Posterior predicted means are plotted in blue and their 95% HDI are shaded.

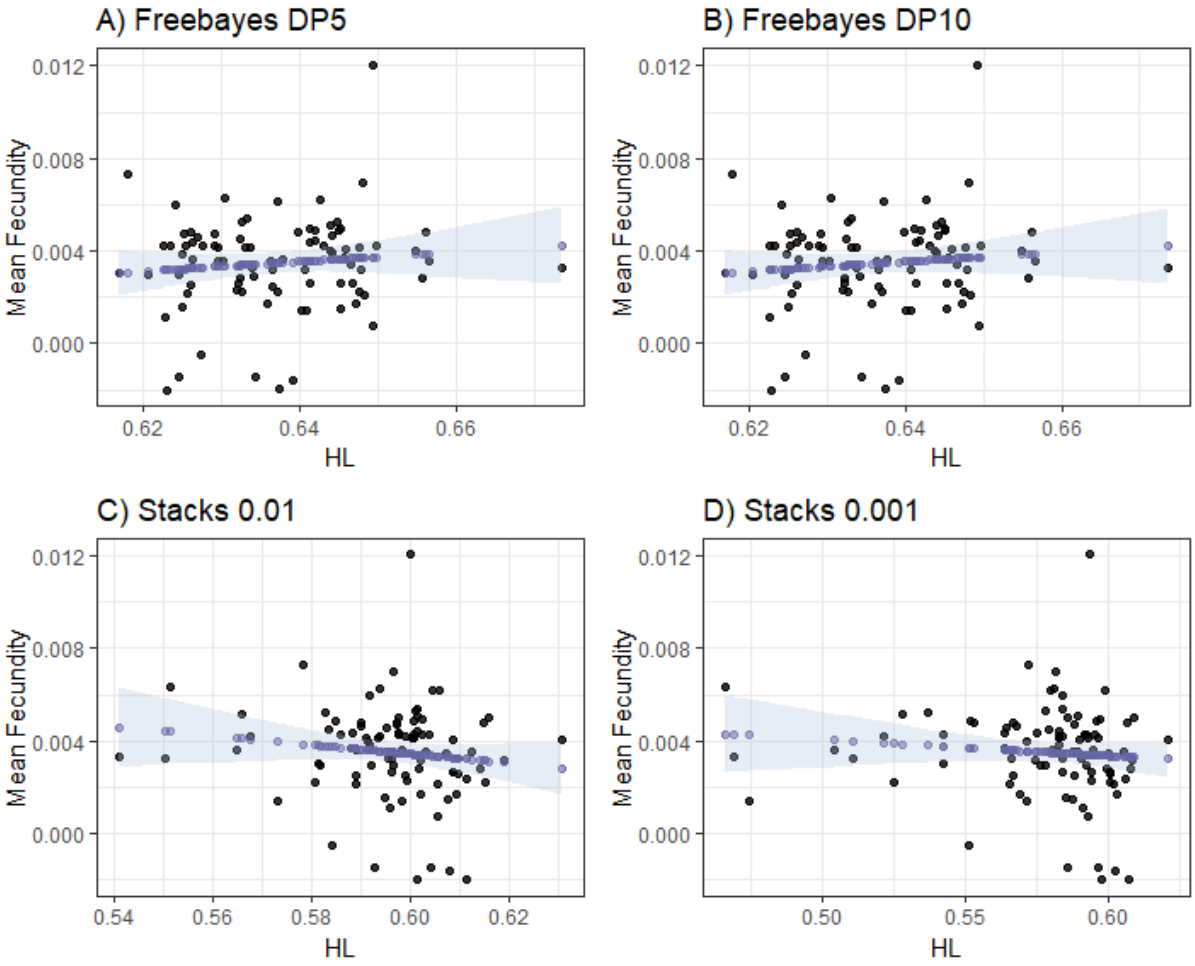

Figure A4.6. Relationship between the inbreeding coefficient HL and reproductive fecundity for the A) Freebayes DP5, B) Freebayes DP10, C) Stacks 0.01 and D) Stacks 0.001 datasets. Posterior predicted means are plotted in blue and their 95% HDI are shaded.

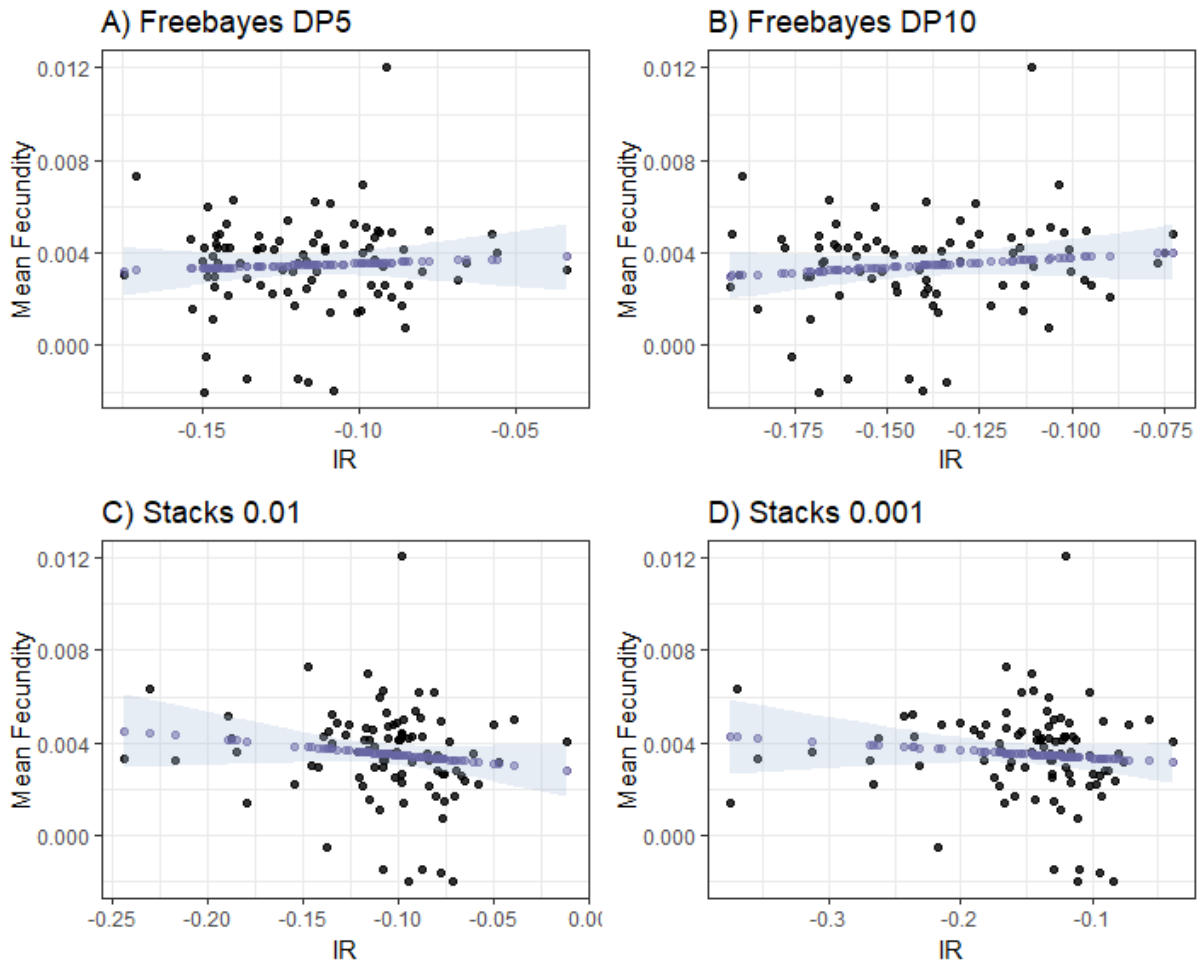

Figure A4.7. Relationship between the inbreeding coefficient IR and reproductive fecundity for the A) Freebayes DP5, B) Freebayes DP10, C) Stacks 0.01 and D) Stacks 0.001 datasets. Posterior predicted means are plotted in blue and their 95% HDI are shaded.

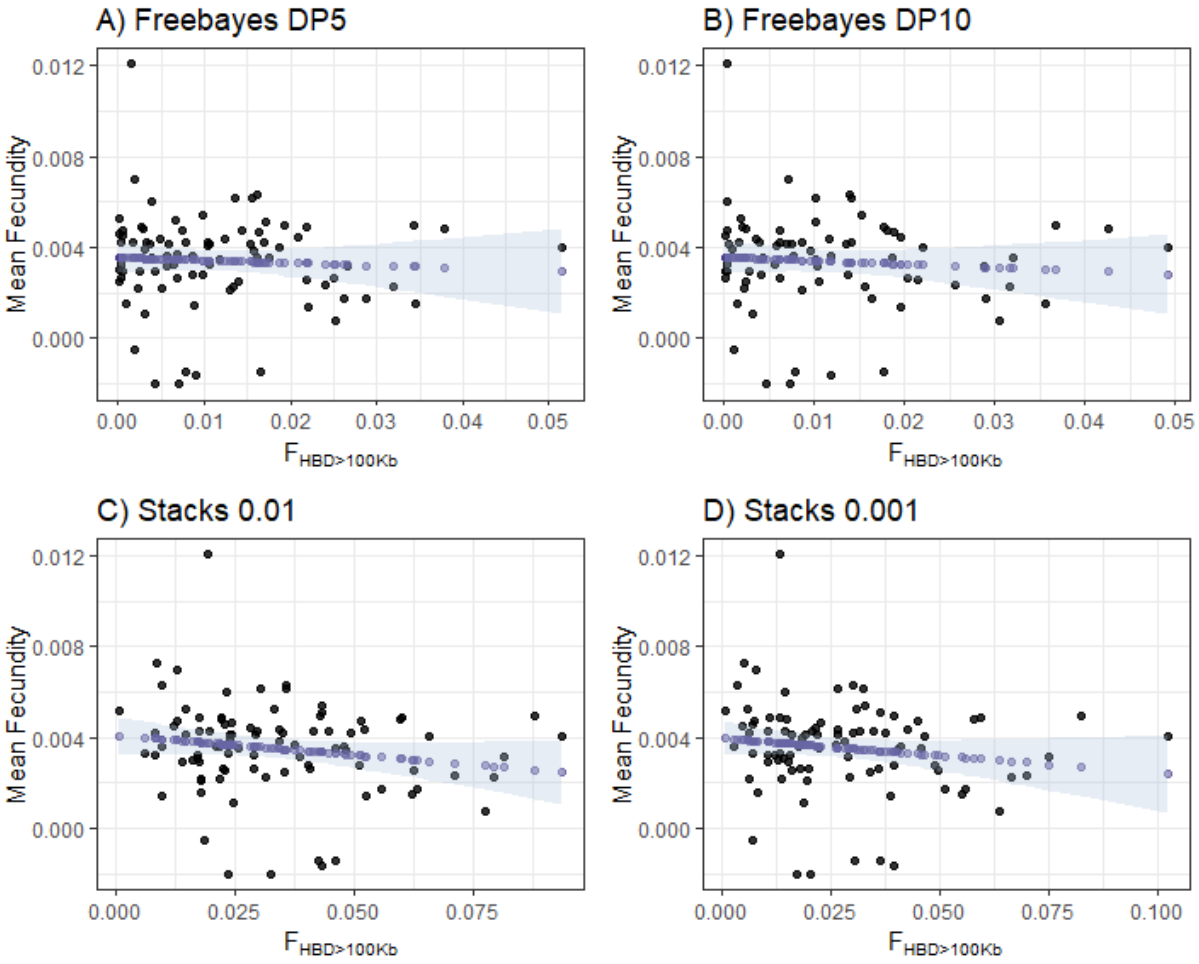

Figure A4.8. Relationship between the inbreeding coefficient $F_{HBD>100KB}$ and reproductive fecundity for the A) Freebayes DP5, B) Freebayes DP10, C) Stacks 0.01 and D) Stacks 0.001 datasets. Posterior predicted means are plotted in blue and their 95% HDI are shaded.

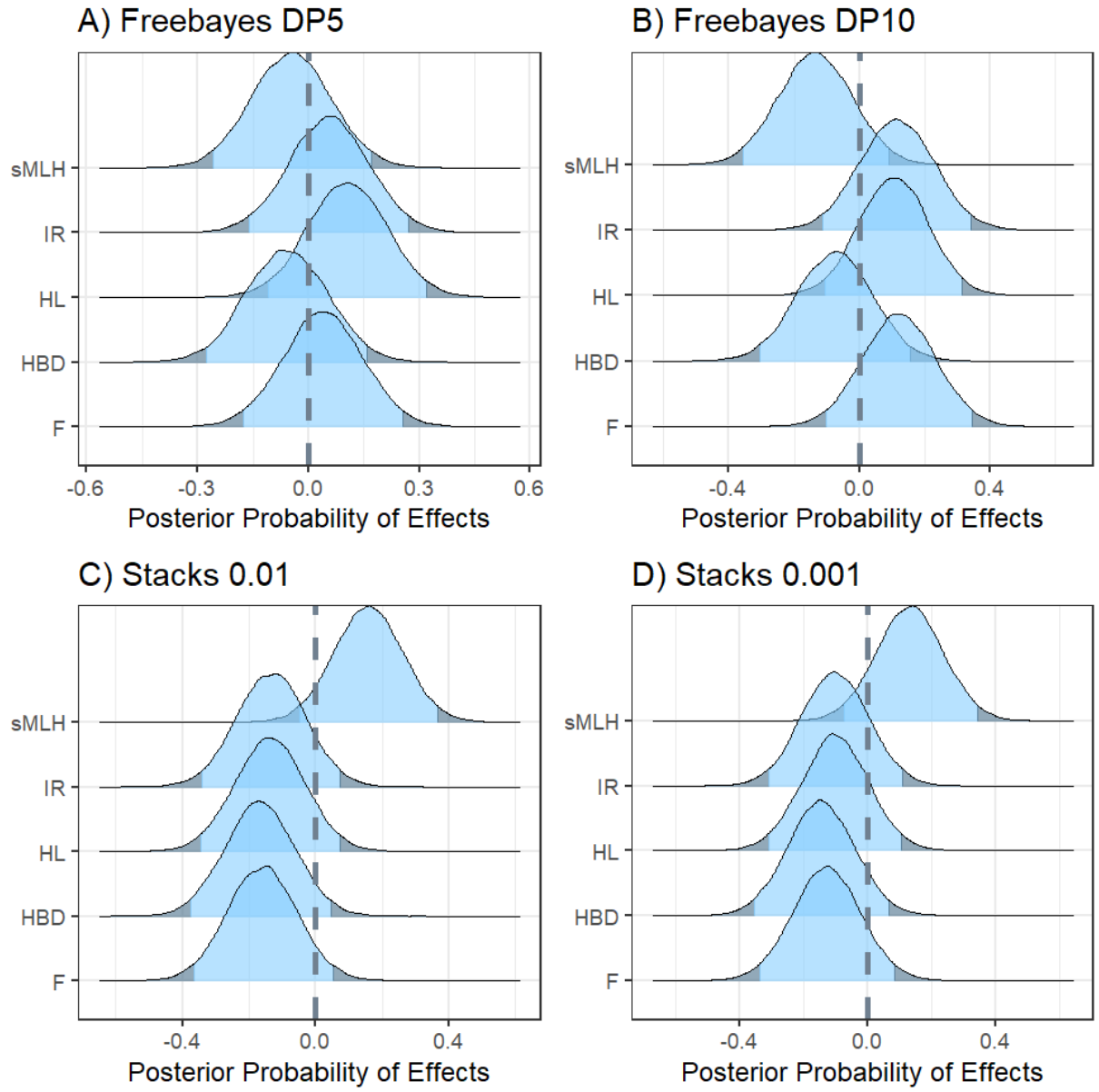

Figure A4.9. Posterior probabilities of the effect of each inbreeding coefficient on individual fecundity for the A) Freebayes DP5, B) Freebayes DP10, C) Stacks 0.01 and D) Stacks 0.001 datasets. Tails of the distribution that fall outside the top or lowest 2.5% are shaded.

A5. A SNP panel designed for monitoring North Atlantic right whales

A5.1 Supplemental Methods

Variant Site Selection

A previous study on North Atlantic right whales used ddRAD sequencing of over 100 females to identify variant sites throughout the genome with two different variant calling pipelines (Crossman et al. 2024). As a starting point for selecting SNPs for this study, we started with 7,544 SNP loci that were called by both variant calling pipelines (Crossman et al. 2024). We used bcftools v.1.16 (Danecek et al. 2021) to exclude loci with a minor allele frequency less than 10%, resulting in 2,872 potential SNP loci to include. In order to reduce the presence of null alleles resulting from variants within priming sites, we used bedtools intersect v.2.30 (Quinlan and Hall 2010) to identify variant sites located within 200bp of our SNP of interest from datasets of unfiltered variant calls from both the same ddRADseq study noted above (Crossman et al. 2024) and from whole genome sequence data (Crossman et al. 2023). We excluded loci with potential variant sites in possible primer sequences resulting in 885 loci of interest.

SNP Panel Primer Development

Using bedtools getfasta v.2.30 (Quinlan and Hall 2010), we extracted fasta sequences from the same reference genome used to call variants 200bp up- and down-stream of each SNP of interest. We were able to design primers for 873 sites using Primer3 v.2.5.0 (Koressaar and Remm 2007; Untergasser et al. 2012) that would have an optimal melting temperature of 60°C, an optimum length of 23bp and generate a product 100-150bp in length. Additional flags in Primer3 provided a 30bp buffer around the SNP of interest, specified a primer size range of 18bp - 27bp, and a melting temperature range of 59.5°C - 60.5°C. We selected the primer pair with the lowest penalty score designed for each site to proceed. Using mfeprimer v.3.3.0 (Wang et al. 2019) we estimated the possible formation of hairpins and dimers within our set of primers. We removed one site where there was a high possibility of hairpin formation by one of the primers and excluded a further 503 sites where at least one of the primers could potentially form five or more dimers. Finally, we used mfeprimer to test for specificity of binding across the genome using the reference genome (Genbank Assembly: GCA_028564815.2). We excluded 62 additional sites where at least one of the primers could create more than five potential amplicons.

We developed this SNP panel to be used for ongoing genetic monitoring and profiling of individual North Atlantic right whales and as a result, we wanted to include loci for sex determination and haplotype identification. Two sets of primers designed

for sex determination in cetaceans in previous studies (SRY: Einfeldt, Orbach, and Feyrer, 2019, Richard, McCarrey, and Wright, 1994; ZFX/ZFY: Konrad, Dupuis, Gero, and Frasier, 2017) were included in our panel. Finally, we designed two sets of primers that capture the known variant sites in the mitochondrial control region that are used to assign mitochondrial haplotypes (Malik et al. 2000). We ordered primer pairs for 311 loci with added tails on the 5′ end of each priming sequence that allow binding of Nextera indexes (Illumina). Primer sequences (including Nextera compatible tails) are available in Table A5.1.

SNP Panel Optimization

For the initial primer pool, primers were pooled in equimolar concentration to a final concentration of 250nM each. We prepared the same 7uL PCR reaction described in Campbell et al. (2015) using 30ng of template DNA. The reaction conditions for the first PCR were: 95 °C for 15 min; 5 cycles of: 95 °C for 30 sec, 5% ramp down to 57 °C for 30 sec, 72 °C for 2 min; 15 cycles of: 95 °C for 30 sec, 65 °C for 30 sec, 72 °C for 30 sec; hold at 4 °C. We visualized the performance of the reaction on a 1.5% agarose gel using 2uL of product diluted in 2uL of Tris-HCl. We made a 10x dilution of the product and performed a bead clean-up with Ampure XP Beads with a 1.8:1 bead to diluted sample ratio. We indexed samples with a unique combination of Nextera XT indexes (Illumina,

San Diego, USA) in a 50uL reaction with KAPA HiFi HotStart ReadyMix (KAPA Biosystems), adding 5uL of product from PCR1 as template for samples with good amplification and 15uL for samples with poor amplification as assessed on the agarose gel. We cleaned the indexed product with a 1:1 ratio of Ampure XP beads to product, eluting in 37.5uL of 10mM Tris-HCl ph 8.0. We quantified each sample on a Qubit fluorometer (ThermoFisher, Waltham, USA) and sent the completed libraries to the McGill Genome Centre for sequencing on an Illumina MiSeq with a v2 Nano reagent kit and a 2x150bp run.

Demultiplexed sequences were processed with fastp v.0.23.4 (Chen et al. 2018) to remove paired end adapters, filter reads with an average quality below 30, trim poly-X tails and merge paired-end reads that both successfully passed filters. Merged reads for each sample were analyzed through the GT-Seq pipeline (https://github.com/GT-Seq/GT-Seq-Pipeline) to compile genotypes and read counts for each sample at each locus. Loci that could not be genotyped in any samples were excluded. Loci that had extremely high amplification were added at lower concentrations to the primer pool. After the first round of sequencing, loci that had an average depth of >400 were added to the primer pool for the second round of sequencing at 80% of the concentration of the other primers. The second round of optimization followed the same protocol and the

final round of testing was prepared for more samples and run on an Illumina MiSeq with a v2 Micro 2 x 150bp run.

References

Campbell NR, Harmon SA, Narum SR. 2015. Genotyping-in-Thousands by sequencing (GT-seq): A cost effective SNP genotyping method based on custom amplicon sequencing. Molecular Ecology Resources 15:855–867.

Chen S, Zhou Y, Chen Y, Gu J. 2018. Fastp: An ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34:i884–i890.

Crossman CA, Fontaine MC, Frasier TR. 2023. A comparison of genomic diversity and demographic history of the North Atlantic and Southwest Atlantic southern right whales. Molecular Ecology 00:1–17.

Crossman CA, Hamilton PK, Brown MW, Conger LA, George RC, Jackson KA, Radvan SN, Frasier TR. 2024. Effects of inbreeding on reproductive success in endangered North Atlantic right whales. Royal Society Open Science Accepted.

Danecek P et al. 2021. Twelve years of SAMtools and BCFtools. GigaScience 10:1–4. Oxford University Press.

Einfeldt AL, Orbach DN, Feyrer LJ. 2019. A method for determining sex and chromosome copy number: Sex-by-sequencing reveals the first two species of marine mammals with XXY chromosome condition. Journal of Mammalogy 100:1671–1677.

Konrad CM, Dupuis A, Gero S, Frasier T. 2017. A sexing technique for highly degraded cetacean DNA. Aquatic Mammals 43:655–660.

Koressaar T, Remm M. 2007. Enhancements and modifications of primer design program Primer3. Bioinformatics 23:1289–1291.

Malik S, Brown MW, Kraus SD, White BN. 2000. Analysis of mitochondrial DNA diversity within and between North and South Atlantic right whales. Marine Mammal Science 16:545–558.

Quinlan AR, Hall IM. 2010. BEDTools: A flexible suite of utilities for comparing genomic features. Bioinformatics 26:841–842.

Richard KR, McCarrey SW, Wright JM. 1994. DNA sequence from the SRY gene of the sperm whale (*Physeter macrocephalus*) for use in molecular sexing. Canadian Journal of Zoology 72:873–877.

Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG. 2012. Primer3-new capabilities and interfaces. Nucleic Acids Research 40:1–12.

Wang K et al. 2019. MFEprimer-3.0: Quality control for PCR primers. Nucleic Acids Research 47:W610–W613. Oxford University Press.

Table A5.1 Primers for the 278 amplicons used in the final panel. Primer names correspond to the start position of the original 400bp region we used to identify suitable primers and the SNP of interest in each position is 200bp downstream per the assembly available in 2022 at www.DNAZoo.org.

Primer Name	*Sequence including Illumina tails (5′-3′)*
HiC_scaffold_1_5487690_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCTGAAGAGGCTAAGACAGAATGT
HiC_scaffold_1_5487690_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCCTGTCGGACAAATACTGATGGA
HiC_scaffold_1_17156094_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTGGGAAACGAAAACCACTGAAAG
HiC_scaffold_1_17156094_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGATGGTAATGGTCCTGTGGTAACC
HiC_scaffold_1_42480390_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTGCTTTTAGTGTGGCTTTCTTGG
HiC_scaffold_1_42480390_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGTTTAAAACCTGTTATCAAACATTCCA
HiC_scaffold_1_50962272_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCTACTTTCTGGGAGGTTGGACTG
HiC_scaffold_1_50962272_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCATGCAGTCTAAGAATCATGCCA
HiC_scaffold_1_61058877_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTTAGACCTGACTGAAGTCTGTGC
HiC_scaffold_1_61058877_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGTGTGCAAGTAAATCAGGTTGCT
HiC_scaffold_1_82859260_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCAATGTATCCAACCTCCCTGCTA
HiC_scaffold_1_82859260_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAGGCAGAAACCCTGTGTTCTAAT
HiC_scaffold_1_84341219_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAAGCCTCAGCAGATGTATGAGAG
HiC_scaffold_1_84341219_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGCACCATGCTAAGCCATATCTTG
HiC_scaffold_1_86784517_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAGAAAGCTGTTTAGGATGGAGGC
HiC_scaffold_1_86784517_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGCACCCACCTCATTGTAGTGTAA
HiC_scaffold_2_10421859_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTGGCTCTCAAGGAATTTCACTCA
HiC_scaffold_2_10421859_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGACCCAGGAGTTCTACAAACACTC
HiC_scaffold_2_14874076_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGGTCTTGACCAAATTCTACTAGAACC
HiC_scaffold_2_14874076_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGATACTGCTGGTGTGCATTAGGAA
HiC_scaffold_2_19828980_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGCTTGAAAGACAACCGCTGAAC
HiC_scaffold_2_19828980_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTAGGTCCCAGGGAGAAGTTAACT
HiC_scaffold_2_21569305_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCAAAGCCCTTCTGTTCTAGGAA
HiC_scaffold_2_21569305_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCGAGGGGAGGAAACAGTTAGTAA
HiC_scaffold_2_31889491_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTCAAAACTGGCCATATAGACGGA
HiC_scaffold_2_31889491_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGCCTGACCTTCCCACTAAATTGA
HiC_scaffold_2_35073845_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTTCTTTGGAAGTTGGTGAAAGCG

Primer Name	*Sequence including Illumina tails (5′-3′)*
HiC_scaffold_2_35073845_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCACCCGGGTAGAACATAACTGTA
HiC_scaffold_2_37386810_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTCCCCATAAGTACGATAGAGTTTATCA
HiC_scaffold_2_37386810_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTCCTCTTAGCTGAATTCCAGGAC
HiC_scaffold_2_38925890_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTCTGAGGAAAATGAGATGGTGGG
HiC_scaffold_2_38925890_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGATGGTGCATGTTCATCTAACCG
HiC_scaffold_2_55964150_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAGTCCAGGTGAAGGAATTCAGAC
HiC_scaffold_2_55964150_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTTCCCCAAATTGTCTTAAACGGC
HiC_scaffold_2_56234510_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGATTGTGGCTCTGGCTCCTTATTT
HiC_scaffold_2_56234510_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTGTGAGGACTGGTGCTAATGTAC
HiC_scaffold_2_56632005_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTCAGATGATGAACTGCTTCTGCA
HiC_scaffold_2_56632005_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTGTGCTTTATCTAGAGCCTTGCA
HiC_scaffold_2_69018542_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTGAACCTAATGTAGCATCTGCCA
HiC_scaffold_2_69018542_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGACAAAAGATTCTTCAGGGGCAAC
HiC_scaffold_2_80157741_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCTGCAACCTCTCCTTCTCTGAAT
HiC_scaffold_2_80157741_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGCAGTCAAATGTATACAGGGCAC
HiC_scaffold_2_80986766_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTGTGGGAATTCTTGAGATGAGCA
HiC_scaffold_2_80986766_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTTGTGTCTCAAGAGCTTCCGAG
HiC_scaffold_3_2656529_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTATTGTGGCTTGAAATACGCAGC
HiC_scaffold_3_2656529_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCATTGGCACTTGGCGCTATTAAT
HiC_scaffold_3_3182320_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTGTTTTCAGCAGAGGGTAGAACT
HiC_scaffold_3_3182320_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGGGTGAGAGCTAAGTTCAGGTC
HiC_scaffold_3_3375977_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCTAGACTTCCTGCGTCTTCTTT
HiC_scaffold_3_3375977_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTTGTCCTTTACATTTTGCCTCGC
HiC_scaffold_3_9263986_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCCTCTGAACACTCCCTGAAAAA
HiC_scaffold_3_9263986_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCCCATTTGTGTGGCAGGAAATAC
HiC_scaffold_3_9879080_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGTTCTCTTGACTGTTTTTCCGCA
HiC_scaffold_3_9879080_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTCTTGCTATAAGACACCCCTAGG
HiC_scaffold_3_11540105_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGTGGAAGTGCCGAGATTTAGGAG
HiC_scaffold_3_11540105_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCTAAACCCACCACAGACTCTTGA
HiC_scaffold_3_18758973_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGCAGATCGCACGTTTATCTTGAG
HiC_scaffold_3_18758973_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTCTCCCCTCTCTGGTAAGCAATA
HiC_scaffold_3_19550612_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAAGCAAGATTCCTTCCTGACCTT
HiC_scaffold_3_19550612_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGATCTGTGTGTTGAGAGCCAAAAC
HiC_scaffold_3_19773172_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTAGCCAACCTTTCTAAAGAGGCC
HiC_scaffold_3_19773172_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTCGACTGTCTCTTCAAAGCTTCA
HiC_scaffold_3_28582000_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCAAGTGAATGAAGGAGCACACAT
HiC_scaffold_3_28582000_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAGCCAATGTTTACTACTGCATGA
HiC_scaffold_3_32964932_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCAGACTTCTTGTTACATGGGGA
HiC_scaffold_3_32964932_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTCATTAGAAGTTGGTCCCATTCA
HiC_scaffold_3_59892827_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGTATCAAGGAAATGATGCCACGT
HiC_scaffold_3_59892827_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTCCAAACTAGTGTTATCCACCATCA
HiC_scaffold_3_59987484_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGACTCCACGCTTATCAGTTGCATA
HiC_scaffold_3_59987484_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAGAAATCAAGCTAAGGGTGAGCA
HiC_scaffold_4_747586_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTCTGATTTTAAGAAAAGGCTCGG
HiC_scaffold_4_747586_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGACCAACATAAGCTGATACAGAGCA
HiC_scaffold_4_12891193_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTCTGGAAGGTGGGACAAATTCAA
HiC_scaffold_4_12891193_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTGCACCGGTCCTATGTAATAAGG
HiC_scaffold_4_16003593_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTGGAAACTCGGTGTCCTGTAAAA
HiC_scaffold_4_16003593_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGACAGGATTCTTCGACTTGTACCC

Primer Name	*Sequence including Illumina tails (5′-3′)*
HiC_scaffold_4_22349252_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCAGACAACAGGAAATGTGTGTG
HiC_scaffold_4_22349252_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTTGACAGACGCAATTCTTAAGCC
HiC_scaffold_4_26321802_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTTCTACCAAGATGTGCAGGTTGA
HiC_scaffold_4_26321802_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGATCACTTCCGTGTTCAGTAAGCT
HiC_scaffold_5_4504515_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAAGGGTGCAGTTACATTCGTTTG
HiC_scaffold_5_4504515_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTTAGGCTGATGGGCAGTGTTATT
HiC_scaffold_5_9571459_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCATGGGGCTACTACATCAGGAAA
HiC_scaffold_5_9571459_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTTAATATTTCCCCCACGCTCACA
HiC_scaffold_5_10174720_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCAATGGAGTCTGCGATTTCACTC
HiC_scaffold_5_10174720_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTCCCACCAATGAATTCTCTTTTAGA
HiC_scaffold_5_24863504_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAATGGGAAGATGCAACAATCTGC
HiC_scaffold_5_24863504_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTATTGACCAACAGTGCACCCTAA
HiC_scaffold_5_26286250_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGTGGTTGCTATGGTTGTGTTTGA
HiC_scaffold_5_26286250_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGACTGTGAAAGAAATGCTCGATCA
HiC_scaffold_5_34722835_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCAGTCAGAATACACCAGTGGAA
HiC_scaffold_5_34722835_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGTCACAGGAAACAGCACATGAAA
HiC_scaffold_5_53049486_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGCATTTACCACATTTCATCCCCC
HiC_scaffold_5_53049486_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTGGGGATATGTATTGCATAATGTGA
HiC_scaffold_5_56592860_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTCACGCACGATGTTGCTTTAAAT
HiC_scaffold_5_56592860_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTTACTGGAGAGCCTACCATATGC
HiC_scaffold_5_57903154_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTGACTCCTGGCTTCAAATGAGTT
HiC_scaffold_5_57903154_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTGGAGTCCTCACCTTAAATCACC
HiC_scaffold_5_68194396_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCTATTTCACAGCTGAGGCAATGG
HiC_scaffold_5_68194396_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCAATAAACGTGCCAGACTATGCC
HiC_scaffold_5_73768107_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGTACGCTGATTCCCTCCTTGTTA
HiC_scaffold_5_73768107_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGATTTCTGATGAGACCCAGCTGTT
HiC_scaffold_5_75420640_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCCCATTGAGTCCACTTCCTAAA
HiC_scaffold_5_75420640_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAAATCAACCAGCCCTACTCCAG
HiC_scaffold_5_76660773_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGGTTTCACGGGTCTACACATTTC
HiC_scaffold_5_76660773_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGACTGGAAGGAAACAGGGAAATGA
HiC_scaffold_6_22911369_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGAGATGTCAGCCGGATACAATCT
HiC_scaffold_6_22911369_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCTACACCCAGTTCCAGTCAGATC
HiC_scaffold_6_38872257_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTCATTCTGGAACTTTCAGTGACCT
HiC_scaffold_6_38872257_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCTCCATGGAGTGCCTAAAAGTCT
HiC_scaffold_6_41848789_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAGATGTCCAATGCTAGTGACCTC
HiC_scaffold_6_41848789_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGACAGCAATTGTTTCCCAACGATT
HiC_scaffold_6_42067375_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTCTCCAGGGAAAAAGGAATCTGA
HiC_scaffold_6_42067375_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCAGACCTATGCCATTTATGCTGC
HiC_scaffold_6_43669385_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAATGAGTCGGAGGAATTCTCTGC
HiC_scaffold_6_43669385_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTCCCTTGGAGACCTATAGTTCCC
HiC_scaffold_6_46418541_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTCTGTCAAGCCCAACAAATTACA
HiC_scaffold_6_46418541_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGACACTCATCTTGCAAATCCCCTA
HiC_scaffold_6_57266843_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGACCATGTGGCCATTATAAGAAGGT
HiC_scaffold_6_57266843_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGACAAAAACAAAGATCTCAACTAGCT
HiC_scaffold_6_66145158_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAACCAACGAACACCCACAAAATC
HiC_scaffold_6_66145158_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGATGTAGCTAGTGAACGATGGAGC
HiC_scaffold_6_69782590_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTTGCACTGGGGTGTGAAATTATC
HiC_scaffold_6_69782590_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCCAGATAACACCAGGTCTGTTCA
HiC_scaffold_7_2267624_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGCAGTCTGCAGAATTCATTTCGT

Primer Name	*Sequence including Illumina tails (5′-3′)*
HiC_scaffold_7_2267624_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGACTCCCTGCAGTAAGTTCCTTTT
HiC_scaffold_7_28676453_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTGCACACCCTATTAGGAGAATTCA
HiC_scaffold_7_28676453_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTCTGGATAGAAAGTAACTCTTTGCA
HiC_scaffold_7_44583407_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAATGATGACGCTCTCGACCAC
HiC_scaffold_7_44583407_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGTAGGAGAAGATGGAGATGGCG
HiC_scaffold_7_46149551_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCTTTTAGACAGAGCTGGCAATGG
HiC_scaffold_7_46149551_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGCCCATCACTTCATCAAAGATCC
HiC_scaffold_7_54033360_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGCCACAGAGGATTAAGGAAGATCA
HiC_scaffold_7_54033360_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCGATGAGCTAGGTATTGGGATACA
HiC_scaffold_7_55677965_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGGCAACCAAAAGTGAGAGATCTT
HiC_scaffold_7_55677965_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAAAGCCTGCATTTCTTGGAAAAA
HiC_scaffold_8_5145544_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAGATTTTTCTTTTAACCAGGTGGC
HiC_scaffold_8_5145544_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAGCTGGCTTACTGATATAACGTCA
HiC_scaffold_8_19797774_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGATCACTTTTCTCACGAGAGCACT
HiC_scaffold_8_19797774_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAGAATATGACATTGAACGCGCAA
HiC_scaffold_8_22052472_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCTTTAGGTCCTGAGTGAGGTACG
HiC_scaffold_8_22052472_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCACCATGTGCAACCCAGTATTTT
HiC_scaffold_8_37759473_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTGAGGAAGGTAATGCCATCGAAA
HiC_scaffold_8_37759473_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGGAACATAGCTCCATTGAATCCC
HiC_scaffold_8_50871305_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGGCTGAGACTCTTTAGGAGCTAA
HiC_scaffold_8_50871305_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTGCTAACAGCTTGGGAACTAGAT
HiC_scaffold_8_52648543_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGGCTGCAATGTGTTAGGTGTTAT
HiC_scaffold_8_52648543_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCCCTCCCCACCCAAAAATTACTT
HiC_scaffold_8_62456637_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTATTGTGTCACTGCAGGCTGTAT
HiC_scaffold_8_62456637_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGGAGAATTCACTGATGGGGCTTA
HiC_scaffold_8_67790355_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCACTCGTAGCAGTAATCTAGGC
HiC_scaffold_8_67790355_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTTCTCTCCAAAACCGGGACTATC
HiC_scaffold_8_75916647_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAGTTGTGTTTAATGCAACCAACA
HiC_scaffold_8_75916647_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGTATGTGTAAGGGAGGGATCTGC
HiC_scaffold_8_77736396_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTGAAGAAATACAAGTTGCAGGCG
HiC_scaffold_8_77736396_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAACCAAAGAAAGCATCATGACCG
HiC_scaffold_8_78734402_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTTGGAAGCCAGATGAAAATGTGG
HiC_scaffold_8_78734402_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTCCAACTTTGCTACTAATGGGCT
HiC_scaffold_8_79783489_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGACTCAAACATCAGTAAGTGCCAA
HiC_scaffold_8_79783489_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGGCACAGTTGTCCTGATTTATGG
HiC_scaffold_8_89195617_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGAGAAGTCTGAGTCAGGCAAAGA
HiC_scaffold_8_89195617_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGTAGCAGCTGGAACAAGTAGAGT
HiC_scaffold_8_91820234_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTTTGTGAAAGCCTAAGATGGACA
HiC_scaffold_8_91820234_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAAAAGTGCCACGTGTTTTAACCA
HiC_scaffold_8_92996820_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTTCAAAGTGTAGACCCAGTAAGAT
HiC_scaffold_8_92996820_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAGGTCACGAAAATCGAGAGGAAA
HiC_scaffold_8_104226048_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGGGCTATAATCTGACCTCTTGCA
HiC_scaffold_8_104226048_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCCATCTGTAAGTTTTACCTGAGGGA
HiC_scaffold_8_115632576_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGGAAAAGGAGGACCAAAAACAGG
HiC_scaffold_8_115632576_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTCTTCCTCGGTACCTCTCTATCC

Primer Name	*Sequence including Illumina tails (5′-3′)*
HiC_scaffold_8_119251933_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCTTCTCAGTTGCAGAAGAGTTGA
HiC_scaffold_8_119251933_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAACCTGCTAAGTACACATGGTGG
HiC_scaffold_8_120650931_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGCACTGAAATAAGGGGGAATGTG
HiC_scaffold_8_120650931_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGATCCTGGAACCTCCTTCTGATTG
HiC_scaffold_8_123029165_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTTCTGCTTTCTACTTGGACCTCC
HiC_scaffold_8_123029165_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCTCTTGAAACCCACGTGTCAATG
HiC_scaffold_8_132800814_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGACACTTTTACAAAATTAGGATCGATC A
HiC_scaffold_8_132800814_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTACGAGCTGCATTCAGTCGAG
HiC_scaffold_9_14110534_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTTGGAGTCAGTGTGTTAGATGCT
HiC_scaffold_9_14110534_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCTCCAATCTCCATACTCGGTCAG
HiC_scaffold_9_32368883_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAGGTTGCCTCATTCATTTCATTCT
HiC_scaffold_9_32368883_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGCATCGGTCCTCCTGTGATATTA
HiC_scaffold_9_36421044_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTGTAAATTCCAAGCTGTAAAATCGA
HiC_scaffold_9_36421044_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGCTTTCAATTTGTCAACCGCATG
HiC_scaffold_9_53960741_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTTCATTTTAGGGTTTGTTGTGGC
HiC_scaffold_9_53960741_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTCCAGCTCTTTTATAAAACACGTCT
HiC_scaffold_9_55252946_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGATGCCACCGTAAAGGAAGATGAT
HiC_scaffold_9_55252946_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAGCTGGGTAAATGTGGGTAAGTT
HiC_scaffold_9_64980245_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAGAGGGTGGAAATTTGGCTTAGT
HiC_scaffold_9_64980245_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTGACAGACCCTTCTTGCTTAACA
HiC_scaffold_9_71222842_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTTGTGTTAGACCTGATCCTGCTG
HiC_scaffold_9_71222842_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCACCTTCAAAGCCCAGAAAGTTT
HiC_scaffold_9_76004948_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTGCATCTCACTTTAGCTGTAGGA
HiC_scaffold_9_76004948_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTTGTTAATCAGTTAGGGTGGAGAA
HiC_scaffold_9_78225358_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGCTTGGGTAGAAAGATACTGGGT
HiC_scaffold_9_78225358_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGCCTTTCCTGGAGTCCTAAAGAT
HiC_scaffold_9_82595155_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAAAGTTGGGGAAGGACTCAAAGA
HiC_scaffold_9_82595155_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGACAGTAGAGGAAAGACTTGACGG
HiC_scaffold_9_83374249_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCAGGCCTCAGTAAAGCTGTTCTA
HiC_scaffold_9_83374249_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGTGAATTCATCTTACGCAAACTGC
HiC_scaffold_9_84356946_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAAAATACGAGATGTTTGTCGAATAGT
HiC_scaffold_9_84356946_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCGGTGGCCTGAAACAGTTAATTT
HiC_scaffold_9_85427331_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGGAAGTCTTGAGGCTAACCAGAA
HiC_scaffold_9_85427331_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTGAGCCTAAATTAAGCCCCAAGT
HiC_scaffold_9_90981274_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGCAGCATGGCTCTGGAAATAAAT
HiC_scaffold_9_90981274_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTACAAAAACAGGTGTGGATGTGC
HiC_scaffold_9_98609779_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGGGAGGATAATGCTCTTGTCACA
HiC_scaffold_9_98609779_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGACAGACCGTGCTTTGAGATATGT
HiC_scaffold_10_5096600_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAGGGTTTGCCATCTCTATTCAGG
HiC_scaffold_10_5096600_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCGTAAGTATAGCAAAGTGGCCCA

Primer Name	*Sequence including Illumina tails (5′-3′)*
HiC_scaffold_10_5551738_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCTCTACCTGACAGACTGGTAAGC
HiC_scaffold_10_5551738_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCAACACGAGCAGAGAGGATCAG
HiC_scaffold_10_10808419_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCGTTCAGTTCACTCGTCAATTC
HiC_scaffold_10_10808419_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAGACAGGGTGATCAGATGGGATA
HiC_scaffold_10_21969595_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTTCATGCAAACACATGTCAAGCT
HiC_scaffold_10_21969595_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTGAATTCTCCCCACCTCACAAAT
HiC_scaffold_10_28620105_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGATGGCAGGGGCAAACATTTC
HiC_scaffold_10_28620105_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCATTGCTTGTTTCACGGACTACA
HiC_scaffold_10_40048610_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGACTTGGTACCATCAGCTTTCTTT
HiC_scaffold_10_40048610_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGGTGCCCAAACTGTGAGTTAAAT
HiC_scaffold_10_51090210_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTGAGATACGCTTTACTGGTGACA
HiC_scaffold_10_51090210_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTCGATCCAGCCCTTGTAAATGAA
HiC_scaffold_10_67129644_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAATAAAGCCCTTAGGATCCCTCC
HiC_scaffold_10_67129644_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAGGAAGTGCTGAGGAATTCATGT
HiC_scaffold_10_69031658_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTGTTTGTTTTCCCTTTGCATAGGT
HiC_scaffold_10_69031658_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTGAGAATCAACGGATCTTGCCAG
HiC_scaffold_10_69201923_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAGGATCATCTCAGAGCAAGACTA
HiC_scaffold_10_69201923_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTCTTAAGCTATGTAGGCAGTTGT
HiC_scaffold_10_90620979_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTGTAGCCAACTCTTTGTGGATCA
HiC_scaffold_10_90620979_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCTAGAAGCTCCCTCCCAGAAATC
HiC_scaffold_11_24611282_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGGCCTCTCTCTATTTGCTCTCAA
HiC_scaffold_11_24611282_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCTTCCACAATTCAGCGCATAGAG
HiC_scaffold_11_31475693_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTGCCAGTATCCACGTGAATAACA
HiC_scaffold_11_31475693_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTATTTTGCAAAACGGTCTCCAGG
HiC_scaffold_11_32940772_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGACAGTGCACAAGGGAGAGTATG
HiC_scaffold_11_32940772_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCTACCATTGCAACCAGGTTTAGC
HiC_scaffold_11_46883727_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGACCTTCATACATAACCCCTTACTT
HiC_scaffold_11_46883727_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCCTCACACCCACTCTCACAATTA
HiC_scaffold_11_48460194_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTCTGAAATGTCCCAAATTGCCAA
HiC_scaffold_11_48460194_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTTCATTAATTTAGCCAGCAAGCC
HiC_scaffold_11_50180119_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCACAAGGTAAGACATCAAACCGA
HiC_scaffold_11_50180119_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCAGAAACGCCCCGGTTAAATG
HiC_scaffold_11_70076581_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTGCTTACAAAGAGATTGGGACGA
HiC_scaffold_11_70076581_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTTTGAAATGCCTCTCAATCTTATTCT
HiC_scaffold_11_75969461_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAGACCAAACTACTTTTGCTGTTACA
HiC_scaffold_11_75969461_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTGTGTGGCTTTTAGTAAGTCGGA
HiC_scaffold_11_79497618_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGACAGGTTGATTGGTATGGACGAA
HiC_scaffold_11_79497618_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCACAGCAAGTGTCCCCAGTAAAT
HiC_scaffold_11_86491838_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAGTGCCTTGGGTTCCATTAGTTA
HiC_scaffold_11_86491838_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGACATAGCCTGGGGGAATCAATT
HiC_scaffold_11_97307493_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGATCACCTGGGCAATGATTGTCTT

Primer Name	*Sequence including Illumina tails (5′-3′)*
HiC_scaffold_11_97307493_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGTTGAGCAGACAGAAGGGAGTAA
HiC_scaffold_12_24851484_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGACAGATCATCCAGTACATATGCCC
HiC_scaffold_12_24851484_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAAAGCACATTCCCTAAGCAAACC
HiC_scaffold_12_42796578_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCGGTTCAGCATCACAAGAGAGAA
HiC_scaffold_12_42796578_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAGGCCCGACTTAGGATGATTCTA
HiC_scaffold_12_50557975_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTGAAGGGATACAATTCAGCACCA
HiC_scaffold_12_50557975_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCATGCTTACATCTTCACCACGAC
HiC_scaffold_12_51189838_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGTGCTAGAAATCCAATGTATGCCC
HiC_scaffold_12_51189838_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTGTGGTTACCTGAAAATGTTACTTC A
HiC_scaffold_12_57658822_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTGTGTGTGTGTGTGTGTCTAGAA
HiC_scaffold_12_57658822_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCCTTCAAGCACACAAGTCACAAA
HiC_scaffold_12_57688824_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAAGTAGGGGAAAGAGTTCTCTGC
HiC_scaffold_12_57688824_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAGAAAACTACGTGGCTTCATCCT
HiC_scaffold_12_59210174_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCGTGATCGTTTTTCTCTTCCAGA
HiC_scaffold_12_59210174_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTACTTTTACGCTGTAGAGGGTGG
HiC_scaffold_12_61235240_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGATGAACGAGGGAGTCAAGATAGC
HiC_scaffold_12_61235240_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCCTCCCATCTCCTTAGGTGTCTA
HiC_scaffold_12_66438797_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAATAGGCAGTAAGAGGAACAGCC
HiC_scaffold_12_66438797_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGCAGTTAGGGTAACAACGGTAGA
HiC_scaffold_12_70558205_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTTGGGAAGAATTCTGGCCAAATG
HiC_scaffold_12_70558205_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAGGAGCATACTAAAAATCTCCCAGT
HiC_scaffold_12_72911552_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAAAACTCTCAAGAACACACAGCG
HiC_scaffold_12_72911552_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTACATTAGACTTCCTCCAGGTGC
HiC_scaffold_12_91463386_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGTTCTTCTCCCACTGCTTCAAAG
HiC_scaffold_12_91463386_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGGGAGATCGATACAGGTGGAAAG
HiC_scaffold_13_4382975_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTGGTCCCACCATTTCACTCTAC
HiC_scaffold_13_4382975_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCTGAGTCTGGCTATGAGAGATGG
HiC_scaffold_13_4770066_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTTAACAGCCGTTTTCACCTTCAC
HiC_scaffold_13_4770066_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAAAGGTGGGAACGAGATTATGGA
HiC_scaffold_13_26505549_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGACAAGGCTTTGTTTCAGCACTTT
HiC_scaffold_13_26505549_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGGCTGCACTGAATGACTTGATAG
HiC_scaffold_13_32916117_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAAGCTTTCTGATCTGGGAGACTT
HiC_scaffold_13_32916117_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGCAATGTACACGTGTGACATCAT
HiC_scaffold_13_38167380_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTCCAGTCTACTCCAGTTTCAAGC
HiC_scaffold_13_38167380_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAGATAGGTGGCTATATGCATGCC
HiC_scaffold_13_53869548_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTTTGCATGAATAGTTTGGAGCCC
HiC_scaffold_13_53869548_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTGTACTTTCCACAGGAGGCAAC
HiC_scaffold_13_56240445_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGACCTATCAGGAGGCTCGTATAGG
HiC_scaffold_13_56240445_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGACAGTCTCTTCACCACATCGTC
HiC_scaffold_13_57659475_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGATAAAGAAGGAGCCACAACGAGT

Primer Name	*Sequence including Illumina tails (5′-3′)*
HiC_scaffold_13_57659475_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGGCTCAGAATTCTCACTCCTGTA
HiC_scaffold_13_90066320_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTGGAAAGCTTTGGCAAAGTATCG
HiC_scaffold_13_90066320_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGACCTAGATGGGGGATTTACATTTT
HiC_scaffold_13_101018574_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGGGTTCTTGATGCAAGCAAATGA
HiC_scaffold_13_101018574_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAGAGCTCTCCTTAGGTACTGCAT
HiC_scaffold_13_108977889_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTAAAGTGAGTGCCAAAAAGCAGG
HiC_scaffold_13_108977889_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCTCTGGCCCCAGTGTATTCTTC
HiC_scaffold_13_130460939_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAGGCCTGTTCAAGAAAGAAGTCA
HiC_scaffold_13_130460939_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTGCTTCCCAAATGGCCTACTTAT
HiC_scaffold_13_146830215_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCTCTGGCCACCATAACTGGATTA
HiC_scaffold_13_146830215_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGCAGAGAATTCAGCCTGTTACTG
HiC_scaffold_13_156448143_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCAGGGCTTGTCTACAAAAATGCA
HiC_scaffold_13_156448143_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAGAACAAAACTGCAGTAGCTCAA
HiC_scaffold_13_157315349_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGACTCTCCACAGGCCCGTAT
HiC_scaffold_13_157315349_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAGGGTACTTTGTGCTGGATGAAA
HiC_scaffold_14_2154506_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGACTTCTTCTCCACCTTCATAGCC
HiC_scaffold_14_2154506_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCATGCTTCCTACCATCATCATCG
HiC_scaffold_14_2653519_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAGCCCAATAAACCCTGGATGAAT
HiC_scaffold_14_2653519_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTCGTGGCTATGCTTTTCCAAATG
HiC_scaffold_14_5320721_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTTCGTTTCTCTGTCCCTGTACAG
HiC_scaffold_14_5320721_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGGGGGTCAGGTTTCTAATTTAGC
HiC_scaffold_14_8290929_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTGTGGACATGCACATTCCATAGA
HiC_scaffold_14_8290929_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCCTTTTGTCCGTTTGCTTTTGTC
HiC_scaffold_14_8538349_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTACCTGAGATTTTTCGTGGTTCG
HiC_scaffold_14_8538349_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAAATACATTTTGTGACCCGGTGC
HiC_scaffold_14_9036035_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGACGTTCTAGATTTTTCTGCTCATTCT
HiC_scaffold_14_9036035_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCAGCAGTAGCCACCAAATTGTTT
HiC_scaffold_14_16606417_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAATAAGTCATTTCCCTGCCGTTG
HiC_scaffold_14_16606417_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGGGACAAAAAGCAAGTTGATCCA
HiC_scaffold_14_21246735_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTAAACATTTGAAGCTGCACACCA
HiC_scaffold_14_21246735_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGACCTTTCATTTAAAACAAACTTAGCC T
HiC_scaffold_14_23411913_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGGGAGGAAACTTTAGCCACTTCT
HiC_scaffold_14_23411913_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTTCCCAGCAAGATTCTCCATTCT
HiC_scaffold_14_30405933_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGATCTCTGCCCTCAATGTGCATAA
HiC_scaffold_14_30405933_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTGCATTGTCTTTCTTCTCTCCCA
HiC_scaffold_14_42542421_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGGTGGGTGGAAACAATGGAAAAT
HiC_scaffold_14_42542421_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGCATCATTAGCATCGCAACAGAT
HiC_scaffold_14_49380366_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGCCCTCTGTTCTCCAAGATTCTA
HiC_scaffold_14_49380366_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAAACTGACCAAATTACATGCACT
HiC_scaffold_14_51317733_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAATCACATTGACCGAGTCCTTCA

Primer Name	*Sequence including Illumina tails (5′-3′)*
HiC_scaffold_14_51317733_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTGTGAATTCTGTGAATGTTGGGG
HiC_scaffold_14_56481354_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGGGTCAGACCAACTTGCTAATCT
HiC_scaffold_14_56481354_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGAATTCCAAAATGACGGTTGCCA
HiC_scaffold_14_63069263_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAAAGAGCAGTCTCGACAACAGAT
HiC_scaffold_14_63069263_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGGTTAGAGGCCTTTCCATGCTAT
HiC_scaffold_14_71637595_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCAAAATGGGGCCACATGAAAAGA
HiC_scaffold_14_71637595_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGATAGGAAATTAATGGGGAGCGCT
HiC_scaffold_14_80371087_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGACTCTTCTGCAACATTTCTTCCC
HiC_scaffold_14_80371087_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTCACAAAATCTGCAGGCAGTTTT
HiC_scaffold_14_84191379_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTTAGGAGCTGTTAGGGCCAAATT
HiC_scaffold_14_84191379_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTGGACAGTGTTAGTGGCATAGTG
HiC_scaffold_14_92223217_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGTCCAGGCTTGCTATCATCTCAT
HiC_scaffold_14_92223217_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGTTGCTTCTTGCCCAAATAGGTT
HiC_scaffold_14_94545156_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTCACCATTGAAGCATGGGATGTA
HiC_scaffold_14_94545156_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAGCCTGGCTCATTTTAGAAGTCA
HiC_scaffold_14_96116903_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTCAGGCATGGAATTCAGAAGAGT
HiC_scaffold_14_96116903_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGATTCTCCTCAGAGGGCATTTCAG
HiC_scaffold_14_111110477_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTCATAACTTCTAAGCAGGTCAGAA
HiC_scaffold_14_111110477_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTCATTATTTAAGTCCCCATTGAAATG C
HiC_scaffold_14_116748812_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCAGAATTCAGTGTCATTCCAGGC
HiC_scaffold_14_116748812_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCCAGAAAAGTGCCCACATTGTAA
HiC_scaffold_14_124928634_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCAAGAGAAGTGACCACCTGAACT
HiC_scaffold_14_124928634_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAATTAGACATCTGGCTACCTGGC
HiC_scaffold_14_125233069_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTAACAGCCCAGGGAAAGGTTAAA
HiC_scaffold_14_125233069_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGTTGGCACTGTGCTATTTGTCTT
HiC_scaffold_14_132859552_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGCGGTTAGTATCACTGATGACCA
HiC_scaffold_14_132859552_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGAAGACCAGAGAGACACTGAACC
HiC_scaffold_14_136734508_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGCTGAATCCAAACACTTTTCCCA
HiC_scaffold_14_136734508_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTGTAAGTTACCAGCTGGAGACAC
HiC_scaffold_14_139747457_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGCTGGGATTTGATATTTGGCGAA
HiC_scaffold_14_139747457_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCAATGCGTAAAAAGACAGGCTCA
HiC_scaffold_14_145169739_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGCTTTGAGAGGGGATGTGATTCT
HiC_scaffold_14_145169739_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGATCTAGTGGATGCCAAGAGACC
HiC_scaffold_14_150262720_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGCACACTGCTGCGAAGTATTTAA
HiC_scaffold_14_150262720_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGATTCGCAACGTCCCTATGTCTAG
HiC_scaffold_15_1681488_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGGTCTTATAGTGAGGCTAGACCG
HiC_scaffold_15_1681488_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCAGAAAAGACCTGGCGCAAC
HiC_scaffold_15_5492086_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGCAAAATGAGCCTATGTGGAGGA
HiC_scaffold_15_5492086_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGTCCTGCAAAATCAAGTCAGCAT
HiC_scaffold_15_17099628_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAGACAGTTTCAGTTACTGGACAA

Primer Name	*Sequence including Illumina tails (5′-3′)*
HiC_scaffold_15_17099628_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTTCAGACACCTTCAACTGCAGTA
HiC_scaffold_15_22308208_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAGAATAAGCAGTTTGAGGCTTCAC
HiC_scaffold_15_22308208_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGATTGGATGGCTTCATAACACACG
HiC_scaffold_15_24973250_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTATTCCTCCTTTGAGCCAAACCA
HiC_scaffold_15_24973250_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTCCTAGGAGGTTTGCAATTGACA
HiC_scaffold_15_27010468_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCTGAGCTGCTGGTGAAATGAATC
HiC_scaffold_15_27010468_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGTTATGGCAGTCCTGGGAAACTA
HiC_scaffold_15_38590960_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCTGTGAGAATGCAACTCCCATTC
HiC_scaffold_15_38590960_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGGTATGGCCAGTTTTGCATTGAT
HiC_scaffold_15_43440552_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTTTGGACTCAAACCTCTCTAGCC
HiC_scaffold_15_43440552_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCCTTGGGACTAGAATTCCAGTGT
HiC_scaffold_15_47475404_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTAATGATGGGGATTGCATTCTCC
HiC_scaffold_15_47475404_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTTGGGAGTCACAAAATTCCAAGG
HiC_scaffold_16_28265897_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGAATGATTTCAGCTCAGTGTGCC
HiC_scaffold_16_28265897_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAGTGGGCGTGCAAAGGTAG
HiC_scaffold_16_39743987_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAGTCATACAGGCTCCAGAGTTAC
HiC_scaffold_16_39743987_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAGGAAACTGAACAGTGGAGTTGT
HiC_scaffold_16_50344796_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTCAGTTTGTTCTTTTAGCAGAGAGA
HiC_scaffold_16_50344796_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCTTGAAAGTCAACTGAACGACCC
HiC_scaffold_16_55361369_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTGGTACAAACAAGCTCCCAAGAT
HiC_scaffold_16_55361369_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGACCCACCTTCCTGAATTCTCTTT
HiC_scaffold_16_57120133_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTCCACAAGCCTCAGAATTCTCTT
HiC_scaffold_16_57120133_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAGCCCTGAAAACAGAGCATATCA
HiC_scaffold_16_57456028_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGATAGCCCTGATTCTGTGGTTCTC
HiC_scaffold_16_57456028_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGACCTCCTCCCATTTCAGACCTAA
HiC_scaffold_16_75910031_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAGCAAAGCAAGGAAAATATGACA
HiC_scaffold_16_75910031_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGCCTAGGTGTGTTCTTTAGCACT
HiC_scaffold_16_78261548_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCCCACATCTTTGTTTTTGGACA
HiC_scaffold_16_78261548_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTCCATCATGTCTGGAACGCTTAT
HiC_scaffold_16_80069038_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGAGTTGGAAAATAAAGCTGCCGT
HiC_scaffold_16_80069038_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGAATGATCGCTCCAGTCTACCAT
HiC_scaffold_16_86594570_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTCATCCCACTGCCTGAATATTCA
HiC_scaffold_16_86594570_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAGTATCTGGCACATTTCTACTACA
HiC_scaffold_16_92129728_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAGGACTGTAAATCGGATTGCGAA
HiC_scaffold_16_92129728_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCAAACACTAGAATGGAAGAGGCG
HiC_scaffold_16_108099592_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGGGGAGAGAAAGATCCTTGGATT
HiC_scaffold_16_108099592_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCTGTTGCCATGAACAAGTTTGGA
HiC_scaffold_16_110460632_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCTACTGTGACCGCCTCAGAATAA
HiC_scaffold_16_110460632_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCACAGTCCATCCCTCAAGTGTAT
HiC_scaffold_16_115020813_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGTCTAATAGCCTCGCCTTCATGA
HiC_scaffold_16_115020813_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGATTCTTTCACGCACACATTCAGG

Primer Name	*Sequence including Illumina tails (5′-3′)*
HiC_scaffold_16_131639550_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTCTGGACTGAGAAGAGTAGAGGG
HiC_scaffold_16_131639550_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTAGCAGCAGTTCTATTTGCCCAT
HiC_scaffold_16_134739704_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGCCTATTAACTAGCTTCCGGTCT
HiC_scaffold_16_134739704_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGATTCCAATCCACTCCCAGTTCT
HiC_scaffold_16_135628552_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTAGAGATTTCAGGAGCTCCAGAC
HiC_scaffold_16_135628552_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTGGTTGGTTATGGAGCTTCAGAA
HiC_scaffold_16_140033199_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCTTACTGCTTCATTGTGTCACGG
HiC_scaffold_16_140033199_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTATGCTTCACCGAATCAGAGTCA
HiC_scaffold_16_156146022_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGACTCAGGGTAGTGGTCAGTACAT
HiC_scaffold_16_156146022_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCCTGAACACTATCCCAAGTACGT
HiC_scaffold_17_1070986_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGCTGTTTAAATCCCATCACCACT
HiC_scaffold_17_1070986_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGCTCATTCAGCTAATAGGATTCTTA CC
HiC_scaffold_17_3074322_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCGGGAAAACTCAGACGTATTGAC
HiC_scaffold_17_3074322_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGAAGCTCTTACAGGTGACAAGGA
HiC_scaffold_17_3191026_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTGGATGCAATGGACAGGTAAAAG
HiC_scaffold_17_3191026_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGATCTAAAGACCTGCTCTAGCTGC
HiC_scaffold_17_4591041_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGGAGTTACTTAGTGGTGCCTCAT
HiC_scaffold_17_4591041_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGACCCATCTTTGAATTTGGCTTAGA
HiC_scaffold_17_5236427_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTAGGAATTCTAGTCCAAGTGGCC
HiC_scaffold_17_5236427_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCCTAGGGCACAAGGATTATTCCA
HiC_scaffold_17_6361666_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTGTCCCAGCGTAAGAAAGAGATT
HiC_scaffold_17_6361666_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTTACAGGCTGGGCGAAGAATTAT
HiC_scaffold_17_9753819_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTGGGAGATGCTTCTAGAGTAACG
HiC_scaffold_17_9753819_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAGGATCTTATTGGCCACTGGTAG
HiC_scaffold_17_13465162_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCTGGATTACAACTTGGCTTTGGG
HiC_scaffold_17_13465162_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCATAAGGGGGAAGAAAGGTAGCA
HiC_scaffold_17_16999892_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCATACAGTATTCACGCATGTGCA
HiC_scaffold_17_16999892_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCTATGGTTTCTGCCCTTCGGATA
HiC_scaffold_17_19421329_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCAAAGGAATAGCACAATCAGGC
HiC_scaffold_17_19421329_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAGCTGGTTTGTCTACTTGACCTT
HiC_scaffold_17_19473865_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAATATAAGGAGCTTACCTCGAGC
HiC_scaffold_17_19473865_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGGTCTTGTTTCTCCCAGGATTTT
HiC_scaffold_17_20793907_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGTACTATGTCACTTCTCCCGCTC
HiC_scaffold_17_20793907_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGTAGGTTGTGAGAGTGCCTAAGA
HiC_scaffold_17_23893749_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCTGACCAGTATGTACGAGTCAGG
HiC_scaffold_17_23893749_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTGATGTAAGGGGAGTGTTGTTGA
HiC_scaffold_17_33477773_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCATCCATTTAAAGGGGCCTTTG
HiC_scaffold_17_33477773_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCCACTTGGGGCATACGAATTAAC
HiC_scaffold_17_34942520_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTCAGTTCCAACATTAGAAAGGGGT
HiC_scaffold_17_34942520_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTGCTCCTCTTCTACGTTTCTCAC

Primer Name	*Sequence including Illumina tails (5′-3′)*
HiC_scaffold_17_43078089_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGGCATTGAGTCTATGGCATTGAC
HiC_scaffold_17_43078089_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGAGAAAAGGCATGTAAGAGTCGG
HiC_scaffold_17_49823713_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGAAAGTTTACAGATGGCCCGTGA
HiC_scaffold_17_49823713_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTTTTTGAGAAGGGGGATGCTGTA
HiC_scaffold_17_51894984_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTCTCCCCCACTTGTTAATGTCAG
HiC_scaffold_17_51894984_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGACTAGCTACCGAAACAGTTTGGA
HiC_scaffold_17_66445112_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGTCCCTTCTCTCACTTGAAGTCA
HiC_scaffold_17_66445112_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGTTCAAAGCCAAACTCAGGAGC
HiC_scaffold_18_4888896_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGACTTTGGGAGGTTATCAACATAATCA
HiC_scaffold_18_4888896_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTGTTCTCTGGTTTGGGGAGAATT
HiC_scaffold_18_14524086_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCAGGACAGTTCTCTGAATTCCGT
HiC_scaffold_18_14524086_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGAATGTCTTAGGTGTGGCATCCC
HiC_scaffold_18_16988672_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGATCCCTGTATCCCTTGTTCAACC
HiC_scaffold_18_16988672_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCCGTATGTTTCCTACTCTGAGCA
HiC_scaffold_18_36810931_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGGTTGGGGAGGAGAACAGAAATA
HiC_scaffold_18_36810931_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGATCTTAATTGGGCCCCTTTAGG
HiC_scaffold_18_66766479_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAGTGGGCAAAAATGCAAAAATTGT
HiC_scaffold_18_66766479_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCCTGGATAGAACATCAGTCTGGT
HiC_scaffold_18_69539773_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTTTGACCAGTGGCTAAGTGATCA
HiC_scaffold_18_69539773_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTCTCACCATGTGTTCACGTTAGT
HiC_scaffold_18_69908851_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTCTATGAACTGCTGTCTCTAGGT
HiC_scaffold_18_69908851_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCTTCTTCCAGGGGACAGTCAAG
HiC_scaffold_18_70510508_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGTAAAACACCACTCATCATCGGG
HiC_scaffold_18_70510508_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTCCACAGAATTCTGATTCCCAGG
HiC_scaffold_19_11836040_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGCACGTTACAGCATAGTTGATACA
HiC_scaffold_19_11836040_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTCACGCTGTCTGATTTCTTACCC
HiC_scaffold_19_44102747_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTTCTCTCAGAACCAACCTTTCCC
HiC_scaffold_19_44102747_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCACATCTGTCCAACAGTACCTCA
HiC_scaffold_19_46052688_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTATGCCTCCAATTTTCAGTGTGC
HiC_scaffold_19_46052688_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGGTGGGGGAGATTCATATTTCCA
HiC_scaffold_19_64937679_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGGTGTGTAAAACCACTGGTAGCT
HiC_scaffold_19_64937679_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTTTGGGCAGATCTGTTTTGAACC
HiC_scaffold_19_76125194_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAGCATCCTAAGAGTGGGGGATTA
HiC_scaffold_19_76125194_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGACCTCTGTTTTTGTCTGGTGACT
HiC_scaffold_19_79408530_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGATCATTCAGTTTCTAGCTTGGCG
HiC_scaffold_19_79408530_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCTTCTCACCTCATCCAAAGACCA
HiC_scaffold_19_80623136_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGACACTTGTCCTGTTTAAGCTTCA
HiC_scaffold_19_80623136_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTCTTTTGTGGGTGGTATGTCCTT
HiC_scaffold_20_8253393_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAACAGTACTGGTTTGGGAATCCT
HiC_scaffold_20_8253393_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCAGCTTAAGAATGCCAGCAATGA
HiC_scaffold_20_9281594_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAAGCACACATCAAAGTGGTCAAG

Primer Name	*Sequence including Illumina tails (5′-3′)*
HiC_scaffold_20_9281594_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTTCTAGGATGCACTGAAACCAGG
HiC_scaffold_20_13900917_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTTGCAGATTCCAAAGATTACTGTT
HiC_scaffold_20_13900917_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTCACACTGTAAATGCCAGATTGT
HiC_scaffold_20_14372141_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGGAAAACATTCCAAAGCTCCACT
HiC_scaffold_20_14372141_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCCCCCAGATTCTCCATCATGAAT
HiC_scaffold_20_14912846_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGATCTGTTCTTTAAGCCTGGCTCA
HiC_scaffold_20_14912846_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGACCCAAAGTGTTAAGGCTATAGGG
HiC_scaffold_20_24059228_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCGGCACAAGGTCTCAGTCATATA
HiC_scaffold_20_24059228_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGATCCGCTAACTCTGGTGGAAAAA
HiC_scaffold_20_41475624_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGATAGTGGACAGCAGGATGTGAAG
HiC_scaffold_20_41475624_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAATCACCTAGAAAACTGCCAAGC
HiC_scaffold_20_51474071_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAGAGCAATGTTACTTCTTTTCACCT
HiC_scaffold_20_51474071_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCCAAGAGAGACAGATATGCCGAG
HiC_scaffold_20_54327979_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCTTGGGAAAGAGCATGAGTTTC
HiC_scaffold_20_54327979_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTCTCTTCACCTTTGTCAGTTCTCA
HiC_scaffold_20_77321398_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAGACTCTTAAGCAAGATACCCCA
HiC_scaffold_20_77321398_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGACCTCTGTCTTCCCAGTAGAAA
HiC_scaffold_20_88821537_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTTCTGGGATCCATCTGTCAGAGT
HiC_scaffold_20_88821537_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTTATACCCACCCACGCCGATTC
HiC_scaffold_20_113386523_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCAGTTGATTCTGTGGACCTGTA
HiC_scaffold_20_113386523_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGATCATCTCACAAGCTCCCTTTCA
HiC_scaffold_20_122840973_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCTTCCAGCTTGTGAAATGGACTG
HiC_scaffold_20_122840973_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTCTGTAGAGCTTCCTTCATTGCA
HiC_scaffold_20_140362925_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAAGCCCTTTTCTGCTACTGATCA
HiC_scaffold_20_140362925_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCTGAGATGCCAGCCCTATTGAG
HiC_scaffold_20_146771138_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAGATGGAAGAGTAAGGGAACAAGT
HiC_scaffold_20_146771138_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCTCTCTCCTTCCCCAACTTTACC
HiC_scaffold_20_148258812_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGGGCTTGTGGCTATACAGGTTAA
HiC_scaffold_20_148258812_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTTTTGTTCTGCATTACTCTGCCC
HiC_scaffold_20_152111609_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAACAATGCACGCTGTGAATTCTT
HiC_scaffold_20_152111609_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGAGGTGGTAGGGTTTTGGTAACT
HiC_scaffold_20_152191631_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCTTGAAAATTTGGGGGCACTACC
HiC_scaffold_20_152191631_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCTATGGCTCCTGCTGACTTTAGT
HiC_scaffold_20_152728745_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTCCAAGAACCAAATAAGCTCCCG
HiC_scaffold_20_152728745_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGAGGAAACAAGCTGACTTTCGTC
HiC_scaffold_20_162792851_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGGCAATGGGAGATGGTCTCTAAA
HiC_scaffold_20_162792851_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAAACTGCAGGTGATTTGGTTCAG
HiC_scaffold_20_170026716_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGGATCCCCAAATTCCTCAGCTTA
HiC_scaffold_20_170026716_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTGAGAATTATCTGGATTGCGAAGG
HiC_scaffold_20_174828289_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTATGTTTGAGTTTGGGCACGTTG
HiC_scaffold_20_174828289_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTCTGATGAAATCCGTCAAACACG

Primer Name	*Sequence including Illumina tails (5′-3′)*
HiC_scaffold_20_176340421_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCCCTACCTTCCACTTAGTGTTT
HiC_scaffold_20_176340421_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGACAAACATACCACCCCTACGTTC
HiC_scaffold_20_178736191_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGGGTCCTGCTTTACCAAGATGTA
HiC_scaffold_20_178736191_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTTTCACAGTAAGGTTCGAGCCTC
HiC_scaffold_20_184153692_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGACTTTGGGGGAAGGGTAACATAC
HiC_scaffold_20_184153692_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCCAAGGTACTGGTGGCATTACAT
HiC_scaffold_21_3093001_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGTGTTGCTTGTTACATTATGGGGT
HiC_scaffold_21_3093001_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTGGGAACTCAGTCTATGCTTGAC
HiC_scaffold_21_8078044_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTGCTCTTCTTAGCCATTTCCTCA
HiC_scaffold_21_8078044_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCACAAGCACAAGATTTGGCTGTA
HiC_scaffold_21_15265666_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTCTAGCAATTGTCCTAACGGTGT
HiC_scaffold_21_15265666_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCCAGATAGCCTTGCCAAAACATT
HiC_scaffold_21_17550331_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTGCATCCAGTTACCTACTCAACG
HiC_scaffold_21_17550331_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTAAAAGCCTTTCCCCATAACCCA
NARWmtDNA_A_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTCACAGTACTATGTCAGTATTA
NARWmtDNA_A_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGATAATTGAATGCACGATTATAC
NARWmtDNA_B_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTCACTACGGGAAGTTAAAGCTC
NARWmtDNA_B_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCCATAAGATTCAGTTGACTTGA
SRY_Richard1994_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCATTGTGTGGTCTCGTGATC
SRY_Einfelt_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAGTCTCTGTGCCTCCTCGAAGAAT
CetSex94_longer_for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGAGAGCCACAAGCTGACC
CetSex94_longer_rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCCATTTTGTGAGTAAACAAAGCC