Molecular Aging of North Atlantic right whales (*Eubalaena glacialis*)

by

Amanda M. Lee

A Thesis Submitted to
Saint Mary's University, Halifax, Nova Scotia
In Partial Fulfillment of the Requirements for
the Degree of Bachelor of Science with Honours

April 2021, Halifax, Nova Scotia

Approved: Dr. Timothy Frasier
Supervisor

Approved: Dr. David Chiasson
Reader

Date Submitted: April 30, 2021

Table of Contents

List of Figures and Tables

The numbers corresponding to each Figure/Table indicate the page number

Molecular Aging of North Atlantic Right Whales (*Eubalaena glacialis)*

By Amanda M. Lee

Abstract

Knowing the age of individuals within a population provides a wealth of information that is important for understanding aspects of their biology, including many aspects that are important for conservation. Some of these include the relative rates of birth, survival, immigration, and emigration for each age class. This gives useful information for estimating trends over time and estimating extinction probabilities. Cetaceans (whales, dolphins, and porpoises) are challenging to study because they spend the vast majority of their lives below the surface of the water and their habitats are widespread across oceans. Current methods for estimating age in baleen whales are limited. Recent studies have found that methylation patterns associated with certain genes change consistently over time, and therefore provide a "molecular clock" that can be used to estimate the age of individuals. This study investigated whether the age-related methylation patterns exist at CpG sites in the *GRIA2, KLF14*, and *TET2* genes in North Atlantic right whales (*Eubalaena glacialis*). Specifically, methylation patterns were examined for 40 known-age individuals, representing males and females from throughout the spectrum of known ages, to understand the relationship between CpG methylation and age at these age-related CpG sites. The results showed that the levels of methylation at these sites correspond with the age of the individuals, but that methylation patterns were also influenced by how long the tissue samples were in storage. Therefore, this approach can be used to molecularly estimate age in North Atlantic right whales, but the age of the samples must also be taken into account. Future studies should investigate how methylation patterns degrade and what storage solutions are best to prevent degradation.

April 30, 2021

Acknowledgements

I would first like to thank my supervisor Dr. Timothy Frasier for all his guidance, patience, and support throughout the entire research process. I would also like to thank other members of the Frasier lab: Dr. Brenna Frasier, Carla Crossman, Erin Schormans, and Leah Springate for all of their encouragement, guidance, and support. A special thank you to all my family and friends for their endless support and encouraging words. Finally, I would like to thank Fisheries and Oceans Canada and the BRC-Seq Sequencing core at the University of British Columbia for making this research project possible.

Introduction

1.1 The Importance of Age Estimation

Age Estimation within Conservation

Understanding the age structure of a population provides information on the relative rates of birth, survival, immigration, and emigration for each age class. This provides useful information for estimating trends over time and estimating extinction probabilities, such as through population viability analyses (Morris & Doak, 2002). This is often performed using life tables, which help to visualize and understand a population's mortality rate and survival rate at various ages (Tanabe et al., 2020).

Often age structure and the trends it shows are used to predict future trends for many populations, including humans. The information gained through analyses of patterns in age structure is often a key component in estimating how the human population will change in the near future, which has vast implications for public policy (Cohen, 2003). One illustrative example of the utility of knowing the age-structure in wildlife populations involves the Steller sea lion (*Eumetopias jubatus*) in the North Pacific. From 1956 through to 1998 the numbers of Stellar sea lions declined by ~85% due to unknown factors (Sease & Loughlin, 1999). Through age structure analyses, it was discovered that the decline was due to two distinct factors at different periods: beginning in the 1980s the decline was due to a low rate of juvenile survival, whereas during the 1990s the decline was due to low fecundity (Holmes & York, 2003). Thus, despite the external patterns being the same (a decline) the underlying causes differed, with vastly different

implications for appropriate conservation/mitigation actions. Without having the age-related knowledge in this population, these differing factors causing their decline would have been missed. As such, having this information allowed for a better understanding of the factors driving the decline and provided crucial information for the appropriate placement of conservation efforts.

Age Estimation in Whales

Whales belong to a group called cetaceans; this group encompasses whales, dolphins, and porpoises. Cetaceans are divided into two suborders, odontocetes, referring to those with teeth, and mysticetes, referring to those with baleen (Beal et al., 2019; Lubetkin et al., 2008). Odontocetes include dolphins and larger toothed whales, such as killer whales (*Orcinus orca*) and sperm whales (*Physeter macrocephalus)*. While mysticetes include whales that have baleen, such as humpback whales *(Megaptera novaeangliae),* fin whales *(Balaenoptera physalus),* and our study species North Atlantic right whales (*Eubalaena glacialis*). Baleen are keratin-based sheets that hang from the roof of their mouths and have inter-connecting bristles that act to strain food from the water.

Cetaceans are challenging to study because they spend the vast majority of their time below the surface and their habitats are widespread across oceans (Beal et al., 2019). Despite these challenges, much has been learned about many species of whales (Schick et al., 2013; Waldick et al., 2002). One of the most widely-used tools for studying cetaceans is photo-identification. This process takes advantage of the fact that, in most species, each individual has distinguishing physical features that differ from other individuals but are stable

throughout the life of each individual (Hamilton et al., 2007). Taking photographs of these features allows individuals to be identified throughout their life, an example is illustrated in **Figure 1**. This ability to track individuals throughout their lives provides a wealth of data for population monitoring and can be used for abundance estimation, ecological studies, and to better understand population demography (Hamilton et al., 2007).

Since individuals can be identified, the age of some individuals will also be known if they are identified in the year in which they are born. In this way, photo-identification is currently the most widely-used method of age-estimation in cetaceans. However, the lifespan of many of the large whales is comparable to that of humans, or longer (Polanowski et al., 2014). This means that it will take decades before the age of most individuals within a population is known because most of the individuals alive now (or in the earlier decades of a study) will have been born before the study began, and therefore be of unknown age. Because of this problem, several other methods of age estimation have been developed and investigated.

1.2 Other Methods for Age Estimation

Growth Layers in Teeth

The teeth in many mammals, including the toothed whales (odontocetes), grow at different rates throughout the year (Beal et al., 2019). These different growth rates leave rings in the dentin of odontocete teeth (also called “growth layer groups”). Growth-layer groups within the dentin of marine mammals have been a widely used aging method. This is performed by removing a tooth under

anesthesia or from a deceased individual (Beal et al., 2019). The tooth is then cut in half, and the quantification of dentin layers are used to estimate age, and the number of layers equals the number of years—much like counting the rings within a tree trunk (Beal et al., 2019). Although this method provides a means for age estimation, it is not possible for many toothed whale species, such as those that cannot be easily captured temporarily and/or that do not strand regularly (and therefore, there is not often access to dead individuals). Moreover, the extensive capturing and handling of an individual that is required is often not desirable for species that are of conservation concern (Beal et al., 2019). Additionally, this method is limited to the toothed whales, whereas many whales (mysticetes) do not have teeth.

Growth Layers in Earplugs

For baleen whales, one method of age estimation that has been developed is assessing ear-plug growth. Many large baleen whales are known to collect earwax within their ear canal (Trumble et al., 2013). This collection begins to accumulate in layers from birth; this accumulation occurs continuously, allowing for the age of these whales to be estimated from the growth layers. Specifically, this is performed by counting the growth layers that appear within the earplugs, similar to counting the growth rings in teeth. (Tanabe et al., 2020). Although this method appears to give reliable age estimates, it can only be performed on individuals that have died. As a result, it is not useful for monitoring the age structure of wild populations.

Telomere Length

Telomeres are repetitive sequences of DNA that are found at either end of each chromosome (Shammas, 2011). They protect the genome from degradation and intrachromosomal fusion during replication. Chromosomal telomere length analyses have been used for several years as an aging technique. Over time, telomeres may lose repeat units with each cell division; and researchers have found that, in some cases, specific tissues lose a certain number of base pairs per year (Jiang et al., 2008). However, age is not the only factor that can cause telomeres to shorten; stress and environmental factors can also influence an individual's telomere length (Polanowski et al., 2014). Environmental factors and diet have strong influences on telomere length, thus causing wide variation in individuals of similar age. Additionally, telomere length tends to vary at birth, causing difficulty in understanding the changes to their length over time. As a result, telomere length does not seem to be a viable option for age estimation for many long-lived species.

Current Methods have Limitations

Estimating age can be quite challenging, especially in species that are large, mobile, and legally protected, such as cetaceans (Beal et al., 2019). Combined, the information above indicates that there is still a lack of viable methods for accurate age estimation for free-swimming cetaceans. However, methods for estimating age in some species have recently been developed based on epigenetic changes that occur over time: specifically changes associated with DNA methylation patterns, and these method looks promising for use in whales.

1.3 Age Estimation from Patterns in DNA Methylation

DNA Methylation

DNA methylation is an epigenetic change that occurs by the addition of a methyl group to the fifth carbon on a cytosine ring, resulting in a 5-methyl cytosine (5-mc) (Polanowski et al., 2014). This methylation process occurs at CpG sites; where a guanine nucleotide follows a cytosine nucleotide in the linear DNA sequence. Interestingly, CpG sites occur much less frequently than expected throughout most of the genome, but are then concentrated in “CpG islands” in the promoter regions of genes (Jung & Pfeifer, 2015). The methylation patterns of CpG sites influences gene transcription, and is therefore a key mechanism by which transcription and gene expression can be controlled. Specifically, increased methylation is associated with reduced transcription factor binding and therefore a reduction in gene expression (Harris et al., 2020; Moore, Le, & Fan, 2013). Indeed, the addition of a methyl group to a single CpG site within the promoter region can be enough to completely eliminate transcription factor binding for some genes (Harris et al., 2020). However, it is thought that multiple CpG sites within promoter regions allows for more fine-scale control of gene expression than simple “on” and “off”.

Gene expression varies depending on a gene's function and the tissue that is being examined. For example, some gene products are essential for life, and therefore the expression of these genes (and their methylation patterns) remain constant throughout life and within tissue type (Horvath, 2013). On the other hand, the expression of some genes is known to change in response to stress or

other external factors (Bell et al., 2019). Therefore, there are changes in methylation patterns of these genes in response to stress. Some genes, however, are known to change in expression at a consistent rate throughout our lives. It is the methylation patterns of these genes that could be useful for molecular age estimation. Changes in methylation at specific CpG sites have been linked to age in mice and humans (Polanowski et al., 2014), where the degree of methylation is linearly correlated with chronological age (Huang et al., 2015).

DNA Methylation in Humans and Other Species

One area where methylation-based molecular ageing is being aggressively developed is in human forensics. Currently, DNA from a crime scene is primarily used to obtain a genotype at multiple microsatellite loci that can be used for individual identification (Allen, 2010). However, it would be incredibly advantageous if that same DNA could also be used to identify the age of the perpetrator – greatly narrowing the range of suspects. Many studies have now demonstrated that DNA methylation patterns can provide fairly accurate and precise estimates of human age based on a range of bodily fluids (blood, saliva, semen, and vaginal secretions) (Lee et al., 2012; Park et al., 2014).

DNA methylation patterns have also been developed for a range of other species, such as mice and bats (Harris et al., 2020; Horvath, 2013; Wright et al., 2018). These studies found that there was a consistent change in DNA methylation at a subset of genes as age progressed. Throughout many mouse studies, the focus was on DNA methylation within specific tissues, such as heart

or liver tissues (Harris et al., 2020; Horvath, 2013). Through this they evaluated methylation at specific genes within these varying tissues. While within bats, the focus involved skin biopsy sampling, taking a small section of the wing to evaluate the methylation patterns at specific genes (Wright et al., 2018). This approach is what is currently performed in many methylation studies, particular of living individuals because skin represents the only tissue retrievable from many species of free-living organisms.

DNA methylation has also recently been used to estimate age in two whale species: humpback whales (*Megaptera novaeangliae*) and Antarctic minke whales (*Balaenoptera bonaerensis*). These studies examined loci that had been previously shown to have a linear correlation to age in humans and mice (Polanowski et al., 2014; Tanabe et al., 2020). They found that methylation within age-related CpG sites had a linear relationship with age within their study species, showing that this technique for age estimation is a viable option for whale species.

1.4 North Atlantic Right Whales

Aging in Whale Populations

From the eleventh century through to the early twentieth century, North Atlantic right whales (*Eubalaena glacialis)* were hunted and killed at an extremely high rate (Kraus and Rolland, 2007). They were named right whales because they were the “right” whale to kill. Their thick layer of blubber, which caused them to float when dead – which made them easier to transport than other whales and resulted in high volumes of oil - was the primary reason for which they were

initially hunted (Kraus and Rolland, 2007). This, along with being slow swimmers and living close to coastal areas, made North Atlantic right whales prime targets for whalers. This high degree of exploitation caused populations to decline rapidly. North Atlantic right whales were given international protection by the League of Nations in 1935, which lead to right whale hunting being banned throughout the world by the International Whaling Commission (Kraus and Rolland, 2007). However, despite being a protected species since 1935, their population has not been recovering.

North Atlantic right whales have continued to have a high mortality rate and low reproductive rate since the end of their exploitation, leading to this lack of recovery (Kraus et al. 2007). In recent years, instead of intentionally killing them, their high mortality rate stems from their interactions with industrial activities of humans within the ocean, mainly through entanglement in fishing gear and ship strikes (Knowlton et al. 2012; Kraus et al. 2005). Their reproductive rate is also three-times lower than their known potential (Frasier et al. 2007). This reproductive problem manifests itself in two ways. First, there is extremely large variation in the number of calves produced each year, much more so than is expected due to stochasticity. Second, this wide variation exists on top of an overall mean rate that is three-times lower than their known potential. The large year-to-year variation in calf production has been linked, at least in part, to environmental conditions and food availability (Kraus et al. 2007). The factors limiting the overall reduced mean rate are not known, but the main hypothesis is inbreeding, due to that fact that reproductive rates are very stable within

individual females, whereas they differ vastly across females, suggesting that there is an intrinsic component influencing each female's reproductive capacity (Kraus et al. 2007).

Understanding the population dynamics of right whales is becoming increasingly important as the number of whales within the wild continues to diminish (Fujiwara & Caswell, 2001). Conservation efforts for the protection of right whales depend on understanding the age distribution within these populations. Aging in whale populations became extremely important when whales were first subject to commercial fisheries to monitor their population status (Polanowski et al., 2014). The importance of understanding the age distribution has not changed; it continues to be essential for monitoring the recovery of whale populations from past disturbances.

Currently, photo-identification is the most widely used and best technique for the age estimation of North Atlantic right whales. However, the use of this technique causes limitations in understanding the age distributions of these populations. North Atlantic right whales have been a focus of research for 40 years. This is quite a long time for a continuous wildlife study; however, they are a long-lived species, and in comparison to their assumed lifespan, this is a short time frame. The use of photo-identification requires a calf to be seen within its first year and re-identified throughout its life. Since the most prominent way to age right whales is through photo-identification, the lifespan of North Atlantic right whales is largely unknown.

1.5 Objectives

This study aims to develop a molecular method for age estimation in North Atlantic right whales by determining the relationship between the percentage of methylation in the promotor region of age-related genes and known age of North Atlantic right whales. To estimate this relationship, I examined four loci previously identified to have methylation patterns that correlated well with age across other species. Methylation patterns at these loci were quantified across 40 known-age whales, where the age of these individuals ranged from <1 to 33 years old.

I hypothesize that methylation patterns on age-related CpG sites within the promoter regions of these genes will change consistently with age. Recent studies have shown that age-related CpG methylation correlated significantly with age in humpback and minke whales (Jarman et al., 2015; Polanowski et al., 2014). Therefore, this study aims to identify if such a relationship is also present in North Atlantic right whales. This study is part of a larger project that aims to establish a new method of aging via DNA methylation across all cetaceans.

Methods

2.1 Sample Collection

Estimating the relationship between age and percent of DNA methylation requires the analyses of whales of known age. The sample-set consisted of DNA from 40 known-age North Atlantic right whales. These samples were collected over more than a thirty-year period by various research groups associated with the North Atlantic Right Whale Consortium and the New England Aquarium. Samples were chosen in an attempt to represent an even distribution of age across the known-

age individuals. The ranged from less than one-year-old to thirty-three years old. There were 18 samples from males and 22 samples from females. The characteristics of the samples used are shown in **Table 1**. DNA from these samples had previously been extracted, using standard phenol:chloroform procedures (Sambrook & Russell, 2001) as part of routine genetic analyses that our laboratory conducts on this species.

2.2 Selection of Age-Responsive Loci

Other members of the laboratory had previously conducted a literature review to identify candidate loci, as part of a study to develop an epigenetic method to estimate age in another whale species: bottlenose whales (*Hyperoodon ampullatus*). These analyses identified four candidate loci for pursuing in whales: *GRIA2*, *ITGA2B*, *KLF14*, and *TET2* (**Table 2**).

2.3 Sodium Bisulfite Conversion

To age individuals based on the methylation patterns within their DNA, these patterns must be evaluated, and the percentage of DNA methylation must be quantified. To do this a sodium bisulfite conversion is performed. This technique converts all unmethylated cytosines into uracil which then become thymine after amplification. This allows methylation patterns to be converted into differences in DNA sequences, which can be detected using sequencing procedures. Thus, allowing the methylated cytosines to be quantified and for a percentage of methylation to be determined. This percentage of DNA methylation in known-age individuals is then used to estimate age.

Prior to PCR amplification, sodium bisulfite conversions were conducted on the 40 known-age North Atlantic right whale samples. For this process, Qiagen Epitect ® 96 Bisulfite Kits were used (Qiagen, 2020). It is important to note that sodium bisulfite conversions only work with single-stranded DNA. This reaction does not occur for methylated cytosine or any other bases (A, G, or T).

The sodium bisulfite reactions were prepared in 200 µl PCR tubes containing; up to 20 µl of DNA solution, RNase-Free water to bring the DNA volume up to 20 µl, 85 µl of Bisulfite mix which contains high bisulfite salt concentrations to perform the conversion, and 35 µl of DNA Protect Buffer which helps to prevent fragmentation of the DNA during the conversion. This gave an overall volume of 140 µl. These tubes were then placed in the VeritiPro Thermal Cycler and run through a series of denaturing and incubation steps. First the samples were denatured for 5 minutes at 95℃, then incubated for 25 minutes at 60℃, the denaturing step was repeated, incubation was then repeated for 85 minutes, the samples were once again denatured, and lastly, they were incubated for 175 minutes. These thermal cycling conditions provide an optimized series of incubation steps that are necessary for DNA denaturing and sulfonation. Sodium bisulfite conversions must be performed on single stranded DNA; thus, the multiple denaturing steps ensure the DNA remains singled stranded throughout the process. These conditions are designed to enable high cytosine conversion rates (Qiagen, 2020). When single-stranded DNA is exposed to sodium bisulfite ($NaHSO_3$) at high concentrations, the bisulfite will attack the double bond between the 5’ and 6’ carbon atoms, adding a sulfite molecule to the 6’ carbon

atom (Hayatsu et al., 1970; Shapiro et al., 1973), producing cytosine sulphonate (**Figure 2A**).

After 5 hours in the thermal cycler, the samples were transferred to new 1.5 ml tubes. 560 µl of Buffer BL was added to each sample and spun on the centrifuge at 15,000xg for one minute. This solution caused the NH_2 attached to the 4’ carbon atom to deaminate and replaces it with double-bonded oxygen; this produces uracil sulphonate (**Figure 2B**). The mixture was then transferred into Epitect spin columns; these columns were then spun on the centrifuge at 15,000xg for one minute. The flow-through was discarded, and the columns were placed back into the collection tubes. Then a wash step was conducted by adding 500 µl of Buffer BW to each column, which were then spun on the centrifuge at 15,000xg for one minute.

Once again, the flow-through was discarded, and the columns were placed back into the collection tubes. Next, 500 µl of Buffer BD was added to each column, which desulphonated each molecule, resulting in uracil (**Figure 2C**). The samples were then spun on the centrifuge at 15,000xg for one minute. The flow-through was discarded, and the columns were placed back into the collection tubes. Two more wash steps were conducted by adding 500 µl of Buffer BW to each column, which were then spun on the centrifuge at 15,000xg for one minute. The flow-through was discarded, and the columns were placed back into the collection tubes. To remove and residual liquid, the columns were placed into 2 ml collection tubes and then spun on the centrifuge at 15,000xg for one minute. Then the columns were placed in new 1.5 ml tubes and incubated at

room temperature for 5 minutes, to evaporate off any residual ethanol. Lastly an elution buffer was added, 20 µl of Buffer EB was added to the center of each column to remove the DNA from the column, and then the columns were spun on the centrifuge at 15,000xg for one minute. The collection tubes were closed and placed in the freezer at -20.

2.4 Sample Amplification

Previous work conducted in our laboratory had developed primers for the amplification of the promoter region of four genes for use in molecular aging of bottlenose whales. The same four loci were used in this study, they are indicated in **Table 2**. After the North Atlantic right whale samples had undergone a sodium bisulfite conversion, the primers for these four loci were tested for amplification in this species. This was performed by amplifying four test samples from each locus. To ensure that the bisulfite conversion and amplification were successful, the test samples were separated by gel electrophoresis in order to view their band sizes (**Figure 3**). The remaining samples were then amplified and run through gel electrophoresis to view their band sizes as well.

This bisulfite-treated DNA was used as a template polymerase chain reaction (PCR) to amplify each of the four target loci using forward and reverse primers. This reaction uses KAPA HotStart Uracil+ ReadyMix (Roche Sequencing), it is a ready-to-use cocktail containing all components required for PCR, such as dNTPs, $MgCl_2$, as well as a KAPA polymerase that is designed specifically to amplify Uracil. Each amplification reaction had a volume of 30 µl. The desired components for the amplification reactions consisted of 1x KAPA

reagent mix, 0.2 µM forward primer and 0.2 µM reverse primer, 2 µl of bisulfite-treated template DNA, and water was added to volume. The VeritiPro Thermal Cycler was then used for the amplification. Thermocycling conditions were dependent on which locus was being amplified; conditions for each are shown in **Table 3**.

2.5 Bead Clean-up

After amplification, the samples must go through a bead clean-up. Excess primers, nucleotides, salts, and enzymes are removed during this step, leaving purified PCR product. This process used AMPure XP beads to clean the PCR product prior to sequencing (Ampure, 2016). AMPure XP are polystyrene-magnetic beads that are coated in a layer of negatively charged carboxyl groups. Under strict conditions, DNA will bind to the carboxyl groups on the beads. Since the beads are magnetic, placing them on a magnetic stand (ThermoFisher) pulls the beads–with the DNA attached–to the magnetic posts. This allows for the remaining PCR cocktail to be removed.

First, 25 µl of the amplified DNA is transferred to a 96-well plate. The AMPure XP beads are brought to room temperature and then vortexed for 30 seconds. AMPure XP beads preferentially bind to larger fragments, due to this a bead to sample-volume ratio must be chosen that is best suited for each locus. Each locus had a different bead to sample ratio. For *GRIA2* and *TET2* the ratio was 0.8:1, meaning 20 µl of beads were added to each sample. For *KLF14* the desired ratio was 1:1, 25 µl of beads were added to each sample. The sequences of *GRIA2* and *TET2* are larger than those of *KLF14*, therefore, a larger bead to

sample ratio is needed to properly purify the PCR product of the *KLF14* samples. The desired ratio of beads was then added to each sample, mixed, and left to incubate for 5 minutes to allow the DNA fragments to bind to the paramagnetic beads.

The 96-well plate was then placed on a magnetic stand, once the supernatant was clear, it was removed and discarded. The beads were then washed with 200 µl of 80% ethanol, which is incubated for 30 seconds and then removed. This step was repeated. The ethanol removes any contaminates that are bound to the beads or DNA. The plate was then left to dry for 10 minutes to ensure that the remaining ethanol had been evaporated. The beads were then washed with the desired volume of Tris-HCl, which has a pH of 8.5, breaking the bonds between the DNA and carboxyl groups, thus allowing for the cleaned PCR product to be eluted. The desired volume varied for each sample; the elusion volumes were chosen based on the band-size of the sample prior to the bead clean-up. If the band size was greater than or equal to 400 bp then 52.5 µl of Tris-HCl was added, if the band-size was between 100 bp to 400 bp 27.5 µl was added, and lastly if the band-size was less than 100 bp then 17.5 µl was added.

2.6 Illumina Sequencing

Next-generation sequencing (NGS) techniques obtain a sequence for every molecule of DNA in a sample. This differs from Sanger sequencing, where the results are one consensus sequence for each sample (Behjati & Tarpey, 2013). However, with Next Generation Sequencing technologies, you get one sequence per molecule, which will allow you to detect the percent methylation of each site

across a range of cells. The percent methylation found at each age-related site is then used to estimate age.

Additionally, NGS can capture a much broader spectrum of mutations than Sanger sequencing. The increased sensitivity of NGS allows for the detection of small variants, which are only present in a few percent of cells. At the moment, the most widely used approach for NGS is that based on technology developed by the company Illumina, and this is the approach that I used.

Library Preparation

To prepare the samples for sequencing, a library must be created. This is a preparation step for sending the PCR product for Illumina Sequencing. There are three steps to this process: adding the oligonucleotide adapter sequences, adding the overhang adapters, and adding a barcode to identify each sample (Illumina, 2018). These steps are illustrated in **Figure 4**. The library was added using the Nextera XT Index Kit v2 Set A (Illumina, 2018).

Each of the primers used in the amplification step were specifically designed for Next-Generation Sequencing. They have overhangs that allow the library adapters to be attached to the DNA strand (**Figure 5**). The overhang adapter sections of the library are complementary to those of the sequencing primer overhangs. This allows for the attachment of the barcode and adapter sequences.

First, 5 µl of PCR product for each sample was transferred to a new 96-well plate. Then for both Nextera XT Index Primer 1 (N1XX) and Nextera XT

Index Primer 2 (S5XX), 5 µl was added to each well. These index primers contain the adapters that are added to each strand, as well an individual barcode is attached that is indicated by the ‘N’ or ‘S’ number. Barcodes are short 8 bp sequences that can be combined into many unique combinations. Each sample has two barcodes, one on either side of the strand. This ensures that each sample has a unique pair of barcodes.

Next, 25 µl of KAPA and 10 µl of water were added to each well. KAPA HotStart Uracil+ ReadyMix (Roche Sequencing), it is a ready-to-use cocktail containing all components required for PCR, such as dNTPs, $MgCl_2$, as well as a KAPA polymerase that is designed specifically to amplify Uracil. The mixtures within the wells were mixed and the plate was sealed. The plate was then put into the centrifuge at 1000xg for 1 minute. Lastly, the samples were amplified to attach the adapters onto the PCR products. First the samples were heated to 95℃ for 3 minutes. Then they were denatured for 30 seconds at 95℃, annealed for 30 seconds at 55℃, and elongated for 30 seconds at 72℃, these steps were repeated for 8 cycles. Lastly, the final elongation step was at 72℃ for 5 minutes.

Next, a bead clean-up was performed following the same protocol as detailed above. This ensured that the samples were purified and that there was a clean template for sequencing. Next-Generation Sequencing is very sensitive, therefore, it is important that the template is clean, so the unincorporated reaction mixture does not interfere with the results. After the bead clean-up, a subset of samples for each locus were run on an agarose gel to ensure that these were added properly (**Figure 6**). For each locus, the band size should be

approximately 100 bp larger than the band size previously viewed after the amplification, due to the addition of the library.

The samples were then sent to BRC-Seq Sequencing Core at the University of British Columbia (UBC) to be quantified, pooled, and sequenced on an Illumina MiSeq (Illumina, 2018).

2.7 Bioinformatic Analyses

Organizing the Sequences by Locus

The BRC-Seq Sequencing Core returned `fastq` files where the reads had been organized based on their barcodes (where each pair of barcodes represented a different sample). `Fastq` files are text files that contain the sequence data from the Illumina Sequencing. Each file contains the sequence identifier, the reads for each sample, and the quality scores.

Each individual sample had the same barcode at the different loci; therefore, I first had to separate the reads for each sample based on which locus they were from. This is performed by using the program `cutadapt` (Martin, 2011). This is a Python-based program designed to identify sequences containing particular primer or adapter sequences to separate the identified sequences into a different folder. The primer sequences for each locus are illustrated in **Table 4**.

Quality Filtering and Paired Reads

I then filtered the reads for each sample at each locus based on their quality scores. The criteria used were as follows. First, any read where the 5'-end had a

quality score lower than 20 was removed. Next a sliding window was used, the window size was 5 bases and a required average quality score for that window of 20. Then reads where the average quality score across as bases was less than 30 were removed. Lastly, the minimum length of the read was set to 36 bases, those that did not meet this requirement were also removed. This was performed using the `trimmomatic` program (Bolger et al., 2014).

After the reads had been filter by quality the next step was to identify the reads that had a complimentary read that was sequenced in the other direction, and therefore that represented “paired reads”. This step was also performed using the `trimmomatic` program (Bolger et al., 2014). Once the paired reads have been identified, the program `FLASH` was used to generate consensus sequences from the paired reads (Magoč & Salzberg, 2011). This tool extends the length of short reads by finding the correct overlap between paired-end reads and stitching them together.

Alignment with BWA and SAMtools

The paired sequences were then aligned with the reference sequence for each locus. The reference sequences first needed to be indexed so that each position on the reference sequence has a call location and can be mapped. This ensures that the reads will be aligned properly, and the same positions will be called for each read. The alignment was then conducted. Each read was aligned against the reference sequence so that each CpG sites was assigned the same position across reads. These alignments were performed with the program `BWA` (Li &

Durbin, 2009). The data were then converted to BAM files using `samtools` (Li et al., 2009).

R-Analysis

These aligned sequences were then imported into R (R Core Team, 2020), where the C/T ratio for each methylated site (CpG sites) was counted for each read of each individual at each locus. A multiple regression analysis was performed where the age of each individual was the predicted variable and the percent methylation at each site were the predictor variables. Based on these analyses, we then used the estimate coefficients from the regression to predict the age of each whale based on the percent methylation of each site. These predicted ages were compared to the true ages to evaluate the performance of our approach.

Results

3.1 Sample Amplification

Originally, I attempted to amplify all four of the loci previously used in bottlenose whales. However, one locus did not amplify well in North Atlantic right whales and therefore had to be dropped. Moving forward, three loci were included; they are underlined in **Table 2**. These loci were *GRIA2, KLF14*, and *TET2*. The primer sequences for these loci are indicated in **Table 4**.

3.2 Read Depth

We obtained approximately 350,000 reads per sample from the Illumina sequencing, this includes all three loci. When divided by locus there were

approximately 100,000 reads/sample for *GRIA2*, 100,000 for *TET2*, and 120,000 for *KLF14*. After filtering based on quality the number of reads/sample were approximately 20,000 for *GRIA2,* 80,000 for *TET2*, and 90,000 for *KLF14.*

3.3 Correlation Between Methylation and Chronological Age

Based on my original multiple regression analysis, there was not a strong relationship between predicted age (based on percent methylation) and actual age of North Atlantic right whales (**Figure 7A**). Specifically, all samples showed a mean predicted age of ~10 years old regardless of their true age. This was unexpected because these genes showed a relationship between percentage of methylation and age in other organisms. These results suggested that methylation in these genes did not correlate with age in North Atlantic right whales.

3.4 The Effect of Time in Storage on Methylation

The original analyses using multiple linear regression showed that the results did not follow the expected trend. Due to this discrepancy, we tried to think of other factors that may have interfered with the signal between methylation patterns and age. We began by reviewing every step throughout the process and methods to ensure that controls worked properly, and we did not miss any discrepancies between expected and observed outcomes of each step. Through this process, it was possible to determine that each step in the techniques used had performed as expected, with the appropriate results being obtained. Since the results were unexpected and did not correlate to what is found in the literature, we began to

troubleshoot possible issues. Through this, it was noticed that many of the samples used had been in storage for quite a few years, some dating back to 1989. This is more than a 30-year period; therefore, it was thought that the length of time that the samples had been in storage might affect methylation patterns. Although the samples are stored in a solution known to prevent DNA degradation over time, it is not known if, or how, methylation patterns degrade over time in different storage conditions. Therefore, we thought it might be possible that methylation may degrade over time under these storage conditions, and therefore that the amount of time a sample has been in storage may impact the relationship between percent methylation and age. This was tested by adding the "time in storage" to the model as another predictor variable. After this, it became clear that there was an association between percent methylation and age in North Atlantic right whales (**Figure 7B**).

Multiple linear regression was then performed comparing the actual age of each whale to their estimated age, while considering methylation patterns at each site, as well as how long each sample has been in storage (**Figure 7B**). When “time in storage” was added to the model, the results changed and showed that percent methylation at all sites was now strongly informative of whale age (**Figure 7B**). Moreover, with this new model it was possible to estimate whale age based on methylation patterns fairly accurately (± 5 years). Additionally, this model shows that younger individuals will tend to be below the expected 1:1 relationship while older individuals will often be over the expected relationship (**Figure 7B**). This suggests that younger individuals will tend to have their age

overestimated, while older individuals are more likely to have their age underestimated.

The contribution of sex to the regression model is shown in **Figure 8**. There was no difference shown in data points for females and males when comparing age and predicted age. The distribution of males and females across the regression is quite similar; this suggests that the relationship between percent methylation and age is not affected by sex.

3.5 CpG Methylation at Age-Responsive Genes

Overall, a total of thirty-two CpG sites were sequenced over three loci (*TET2*, n = 13; *KLF14*, n = 2; *GR1A2*, n = 17). All of these CpG sites were assayed for correspondence between methylation levels and age (**Figure 9**). All CpG sites showed a strong relationship between methylation and age.

Discussion

4.1 Aging-Responsive CpG Sites Display a Relationship Between Levels of Methylation and Age

Initially, we did not find a relationship between percent methylation and age in North Atlantic right whales (**Figure 7A**). This was rather surprising because other studies that used similar techniques found a relationship between methylation and age at these same sites in other organisms. Since the results were unexpected and did not correlate to what is found in the literature, we began to troubleshoot possible issues. Through this, it was noticed that many of the samples used had been in storage for quite a few years, some dating back to

1989. This is more than a 30-year period; therefore, it was thought that the length of time that the samples had been in storage might affect methylation patterns. Multiple linear regression was then performed comparing the actual age of each whale to their estimated age, while considering methylation patterns at each site, as well as how long each sample has been in storage

These results showed that age could be predicted in North Atlantic right whales using this model with an estimated precision of ± 5 years (**Figure 7B**). While this range is much larger than expected, it is still a great approximation from what current aging methods can estimate. Additionally, a study using similar techniques in humpback whales found that their model can estimate age with a precision of approximately ± 3 years (Polanowski et al., 2014). However, within our data set there were few samples available in the older age range. Since this data set only encompasses a small window of possible ages, the regression is only based on the percent methylation at those ages. This means it can be expected that a model can be evaluated in right whales that is more precise if a more extensive age range is used.

Thirty-two CpG sites were found to have a relationship between methylation levels and age (**Figure 9**). Most of which were strong relationships. This means that these sites can be used as predictors of age in North Atlantic right whales. Three of these CpG sites, one from each locus, were plotted separately (**Figure 8**). Each site showed a very similar trend, a high correlation between methylation and age, once the age of the sample was taken into consideration.

4.2 Sex-Specific Differences and the Effect of Over/Under Estimation

In previous studies, it has been suggested that sex-specific traits may contribute to a difference in methylation as organisms reach sexual maturity (Beal et al., 2019). Mammals display many critical differences in gene regulation between males and females (Horvath et al., 2016). Due to the potential increasing contribution of sex-specific traits as age increases, this is thought to cause a wider spread of data points around the regression line in older individuals.

In the present study, there were no large differences between males and females with respect to percent methylation and age (**Figure 8**). This finding was expected as the genes used in this study were specifically chosen because their methylation patterns do not change with sex, they are only known to change with age (Polanowski et al., 2014; Weidner et al., 2014).

The change in methylation patterns due to age has been found to have some limiting factors when it comes to the underlying relationship between chronological age and the proxy markers for age (Polanowski et al., 2014). Meaning that there is a limit to how predictive proxy markers, such as methylation, can be. In a study by Hannum et al. (2013), using 70,387 age-related CpG sites in humans, the results had a standard deviation in age prediction that was approximately 5% of the human life span. When compared to the current study, which used 32 CpG sites, this is large number of sites. Therefore, this study shows that even with an extensive array of CpG sites, the estimation of age could only be so accurate. This indicates that the ability of the

regression used in this study to predict age is comparable to those found in other studies.

This underlying relationship between chronological age and the proxy markers for age means that it is likely that those on the extremes will have their age overestimated and underestimated. This is what was found in the current study. The model used shows that individuals below the age of 12 are likely to have their age overestimated. This is shown as the younger individuals tend to reside below the expected 1:1 relationship in the model, while individuals over the age of 12 will likely have their age underestimated (**Figure 7B**). This is shown from the older individuals residing above the expected relationship. However, much like with sex, this relationship may be difficult to see on this data set. Since the age range is quite small, the distinction between younger and older individuals is not as clear as it would be if the samples encompassed a larger age-range. Therefore, while this phenomenon can be seen within this model, its effects on aging are not quite as clear as they are in other studies.

4.3 Methylation Patterns Degrade Over Time

The model within this study differs from previous research because it includes the length of time that each sample is in storage as a predictor variable. The addition of this variable and how it changed the age predictions gave some great insight into the factors influencing methylation patterns. These results indicate that methylation patterns degrade over time. This finding could have massive implications for epigenetic studies that involve the use of DNA methylation: any

models that include samples that could have degraded methylation patterns would be affected.

The mechanisms behind the degradation of methylation patterns could be very similar to those that cause DNA to degrade. DNA degradation and how to prevent it was the focal point of many research groups for several years (Dahm, 2007). Understanding this process was a huge finding and allowed for proper storage conditions to be used. DNA is subject to hydrolytic attack and oxidative damage that causes bonds within the structure to be broken (Lindahl, 1993). It was found by Lindahl (1993) that in mammalian cells, cytosines that are methylated make up 10% of the cytosines that undergo hydrolytic deamination. Since this was found to be a factor in DNA degradation, this may also be an issue in the degradation of methylation patterns.

4.4 Implications

There are a few implications of the results found. Most are directly caused by the impact that “time in storage” had on the model. Firstly, while this approach did show that molecular age can be estimated in North Atlantic right whales using this technique, it also suggested that this technique will perform best on fresh samples. This will make this approach less useful for studies trying to estimate the age of individuals based on archived samples. For many species, tissue banks have been collected through decades of collaborative research by a large number of groups. The requirements of “fresh” samples will make this approach less useful for many of these studies.

Second, these results show that methylation patterns degrade over time, even when in a storage solution designed to prevent DNA degradation. This result is extremely important because this issue has not been discussed in epigenetic studies. Therefore, the length of time that samples are in storage will impact analyses of methylation patterns and are relevant for a wide range of epigenetic studies. This situation is fascinating because a vast amount of research has been conducted on what storage solutions and conditions are best for preserving DNA. However, such studies are now needed to assess what conditions and solutions are best for preserving methylation patterns.

4.5 Limitations and Recommendations

Determining the age of individual North Atlantic right whales is incredibly difficult if the date of birth is not known. Since researchers have been studying right whales for 40 years, the only known-aged individuals in this species are those that have been born within this time frame. Thus, the oldest known-aged individual in our study was approximately 40 years old (Kraus and Rolland, 2007). Given that right whales are thought to live for ~70 years (Corkeron et al., 2018), our current data set only captures a small window of possible ages. Therefore, it will be important to continue to test and improve this approach over time, as a larger number of known-aged individuals become older.

Additionally, future research needs to focus on methylation patterns and how they degrade over time. Discovering the best storage solutions to prevent the degradation of methylation patterns is essential for future epigenetic work.

This would be useful for future work on museum specimens and many archival tissues, which have not been kept in conditions to preserve methylation.

4.6 Conclusion

Like many other marine mammals, North Atlantic right whales show very few visible signs of aging (Beal et al., 2019). Therefore, alternative methods for aging are required. Currently, the methods that are used to estimate age do not encompass the entire age range of cetaceans and have many limitations. Many can only be performed on those that are deceased, and those that can be used on live individuals have not been proven useful for population studies (Dunshea et al., 2011).

Currently, there is ongoing research to create more effective molecular aging tools. This study investigates the relationship between percent methylation and age at thirty-two age-related CpG sites in North Atlantic right whales. We identified that CpG sites on *GRIA2, TET2*, and *KLF14* genes could be used to predict age in this species. However, the length of time that the skin samples have been in storage must be added to the model as a predictor variable. This suggests that methylation patterns degrade over time, and therefore future research should focus on the degradation of methylation patterns and finding ways to prevent this.

Figures and Tables

Figure 1: Shipboard photo of male whale 1227. The callosity and its unique pattern are visible in white. Photograph by Moira W. Brown and used with permission by the New England Aquarium.

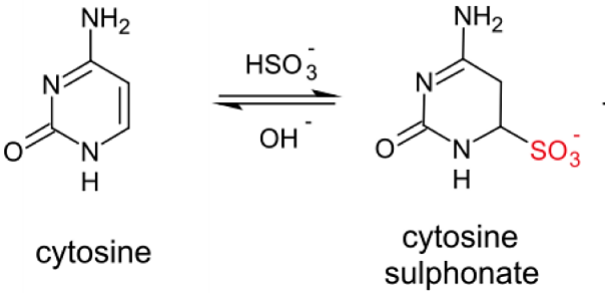
Table 1: Characteristics of known-age samples used from North Atlantic right whales for method development. Samples were collected by research groups associated with the New England Aquarium and the North Atlantic Right Whale Consortium. The samples used were retrieved through skin-biopsy sampling. Whales were aged by the North Atlantic Right Whale Consortium through photo-identification.

Age (years)	Sex of Individual
0	0 female, 1 male
1	1 female, 1 male
2	1 female, 1 male
3	1 female, 0 male
4	1 female, 1 male
5	1 female, 1 male
6	1 female, 1 male
7	1 female, 1 male
8	1 female, 1 male
9	1 female, 1 male
10	1 female, 1 male
11	1 female, 1 male
12	1 female, 1 male
14	1 female, 1 male
15	1 female, 1 male
16	0 female, 1 male
17	1 female, 0 male
18	1 female, 0 male
19	1 female, 0 male
20	1 female, 0 male
21	1 female, 0 male
27	1 female, 0 male

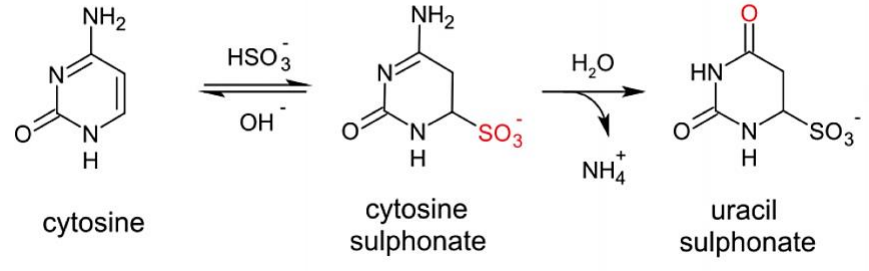
Table 2: Candidate loci identified via a literature review. The samples that amplified after bisulfite treatment in North Atlantic right whales are underlined.

Locus	Species Tested	References
GRIA2	Humpback whales, Minke whales, Bechstein's bats	Polanowski et al. (2014); Tanabe et al. (2019); Wright et al. (2018)
ITGA2B	Humans	Freire-Aradas et al. (2016); Huang et al. (2015); Weidner et al. (2014)
KLF14	Humans	Cho et al. (2017); Jung et al. (2019); Zbiec-Piekarska et al. (2015)
TET2	Humans, humpback whales, Bechstein's bats	Gronniger et al. (2010); Polanowski et al. (2014); Tanabe et al. (2019); Wright et al. (2018)

A)



B)



C)

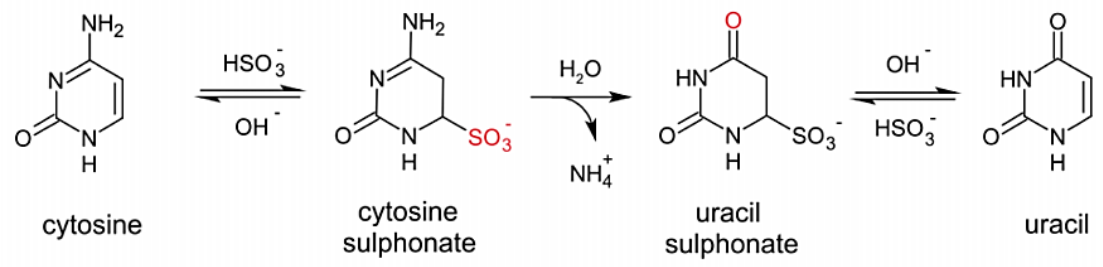


Figure 2: Illustration of the Sodium Bisulfite Conversion reaction. **A)** Sodium bisulfite reacting with cytosine to create cytosine sulphonate. The sodium bisulfite attacks the double bone between the 5' and 6' carbon to add a sulfite molecule to the 6' carbon atom. **B)** Reaction of cytosine sulphonate to uracil sulphonate through deamination. The NH_2 attached to the 4' carbon atom is then replaced with double-bonded oxygen. **C)** Complete reaction of converting cytosine to uracil through a bisulfite treatment by placing it in Buffer BD, a basic solution, to desulphonate the molecule. Image from (*FFPE Bisulfite Conversion*, 2016)

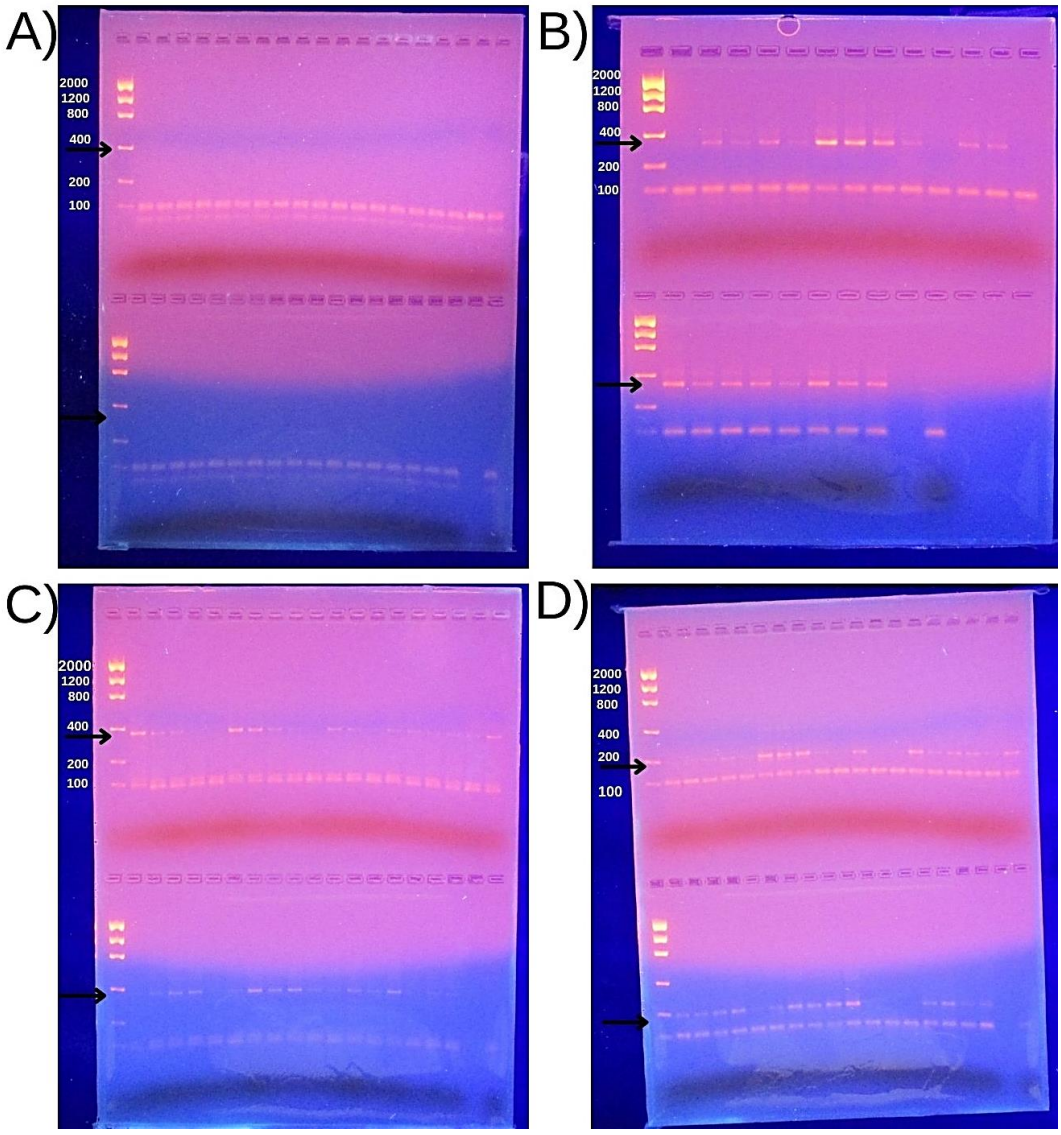


Figure 3: 1.5% Agarose gel showing the sodium bisulfite treated samples for each locus after the amplification step. The ladder fragment sizes are shown on each gel in white. The arrows point to where there should be bands on each gel. **A)** *ITGA2B* band size should be ~310 bp. **B)** *TET2* band size is ~340 bp. **C)** *GRIA2* band size is ~365 bp. **D)** *KLF14* band size is ~195 bp

Table 3: Thermocycling conditions for bisulfite treated DNA for each locus.

Locus	GRIA2	ITGA2B	KLF14	TET2
Conditions	5 min @ 94℃	5 min @ 94℃	5 min @ 94℃	5 min @ 94℃
	1 min @ 94℃ 30 sec @ 60℃ 1 min @ 72℃ X 40 cycles	1 min @ 94℃ 30 sec @ 50℃ 1 min @ 72℃ X 40 cycles	1 min @ 94℃ 30 sec @ 55℃ 1 min @ 72℃ X 35 cycles	1 min @ 94℃ 30 sec @ 55℃ 1 min @ 72℃ X40 cycles
	10 min @ 72℃	10 min @ 72℃	10 min @ 72℃	10 min @ 72℃

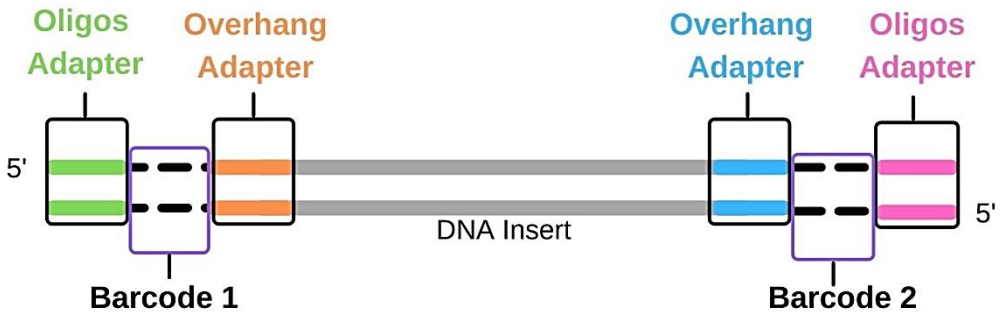


Figure 4: The general components required for sequencing, the addition of which takes place during library preparation. The overhang adapters are complementary to the overhangs that were added on to the fragments by primers. The barcodes are 8 bp sequences, each sample has a unique pair of barcodes. Oligos adapters are complementary to the sequences that are on the flow cell, they attach the fragments onto the flow cell for sequencing.

Forward overhang	5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-[locus-specific forward primer]
Reverse overhang	5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG-[locus-specific reverse primer]

Figure 5: Known sequences that are added to each fragment as primer overhangs that can be used for downstream index and adapter sequence addition

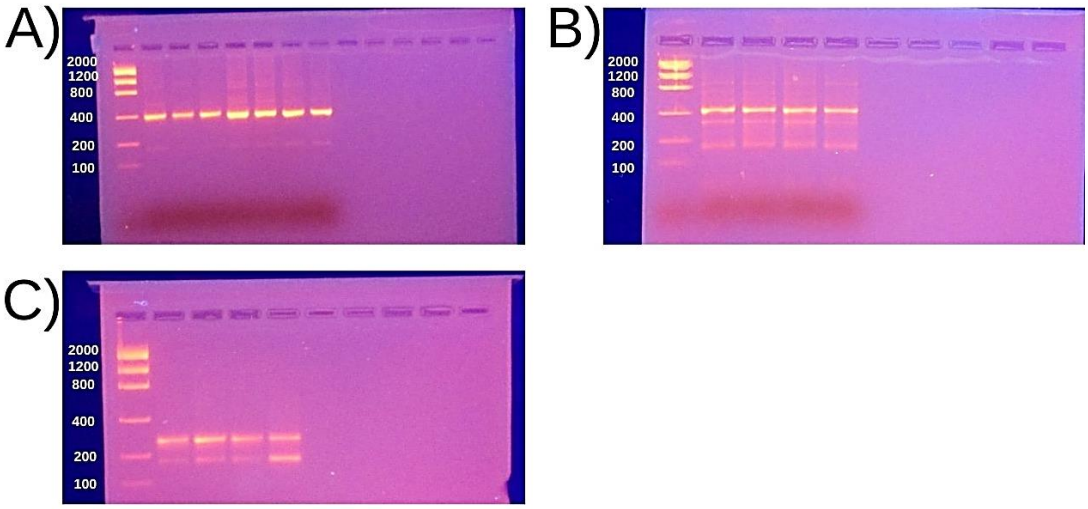


Figure 6: 1.5% Agarose gel showing the samples for each locus after the library indexes have been added to the DNA strands. The ladder fragment sizes are indicated in white. **A)** *TET2* **B)** *GRIA2* **C)** *KLF14*

Table 4: Forward primer sequence and the reverse complement of the reverse primer sequence for each locus.

Locus	Forward/Reverse	Primer Sequence
GRIA2	Forward	5'-GTG TGT GAG TGT ATG GG-3'
GRIA2	Reverse	5'-TAA AAC ATC CAC AAA ATA CCC-3'
KLF14	Forward	5'-AAG TYG GTA GGT TGT TTA GAA GTT A-3'
KLF14	Reverse	5'-CCC ACC GAA CTA AAT CAT TTT TAA C-3'
TET2	Forward	5'-TAA ATT TAA GTA TTT GAA AGT GTA G-3'
TET2	Reverse	5'-TCA TCT CAC TCA ACA AAA ACA C-3'

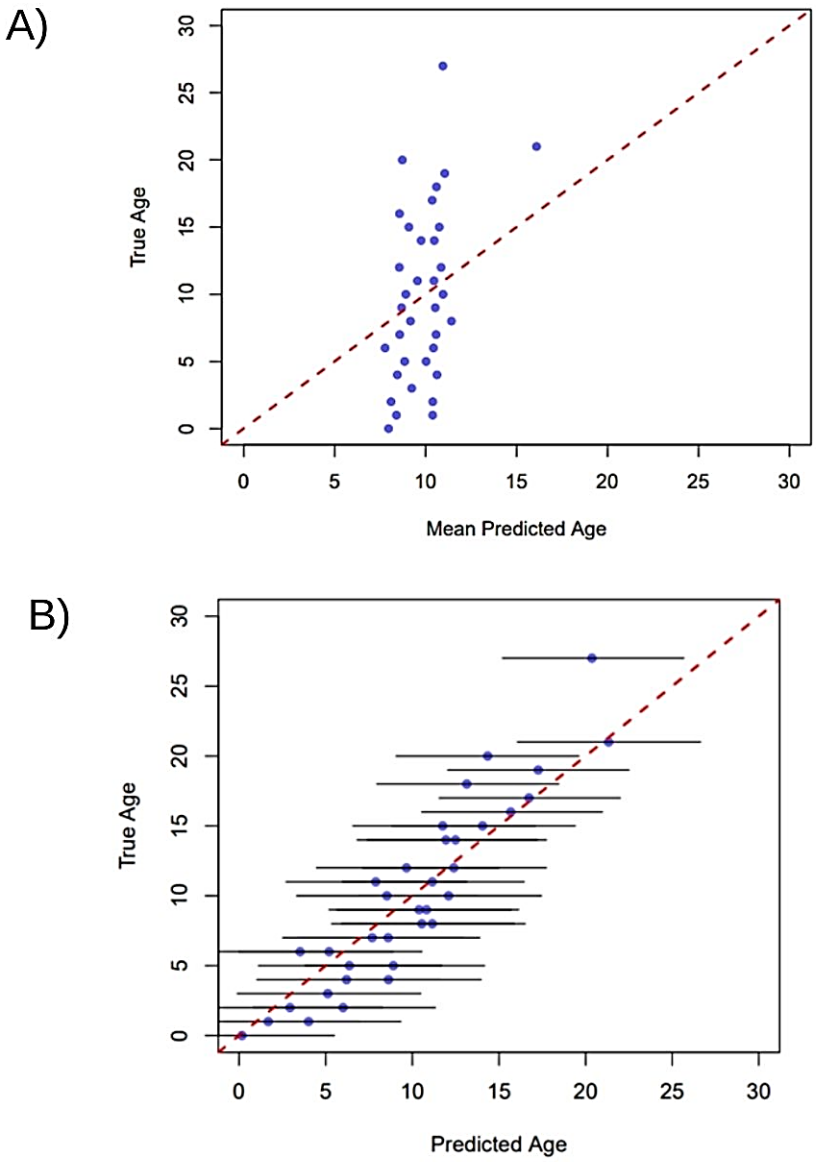


Figure 7: Multiple linear regression plot of the predicted age of each individual based on the methylation patterns across all sites, compared with the actual ages. The red dashed line represents the expected 1:1 between true age and estimated age. **A)** Plot of initial analysis of North Atlantic right whales showing the mean predicted age of each individual relative to their actual age. Variation around each mean predicted age (error bars) are not shown because they are so wide, they would span the whole plot and make it difficult to see. **B)** The same relationships, but now how long each sample has been in storage has been included in the model. The error bars indicated the estimated precision of the regression, the standard deviation is ± 5 years.

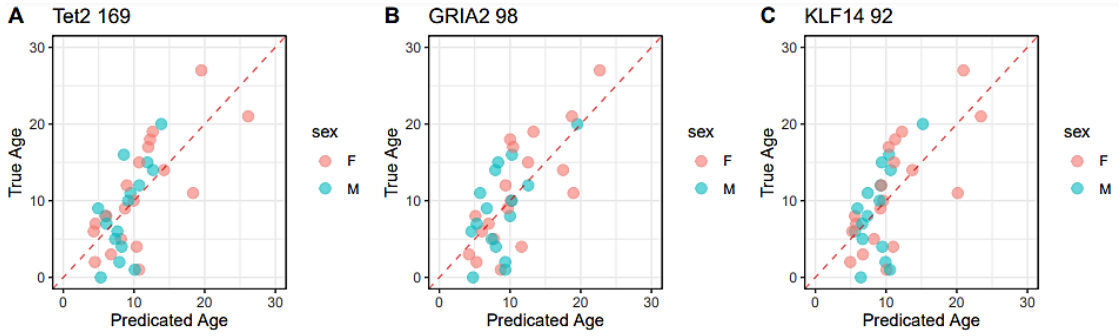


Figure 8: Linear regression model of the mean predicted age of each individual based on the methylation patterns at specific CpG sites compared with the actual ages. The red dashed line represents the expected 1:1 between true age and estimated age. This model compares the female and male samples.

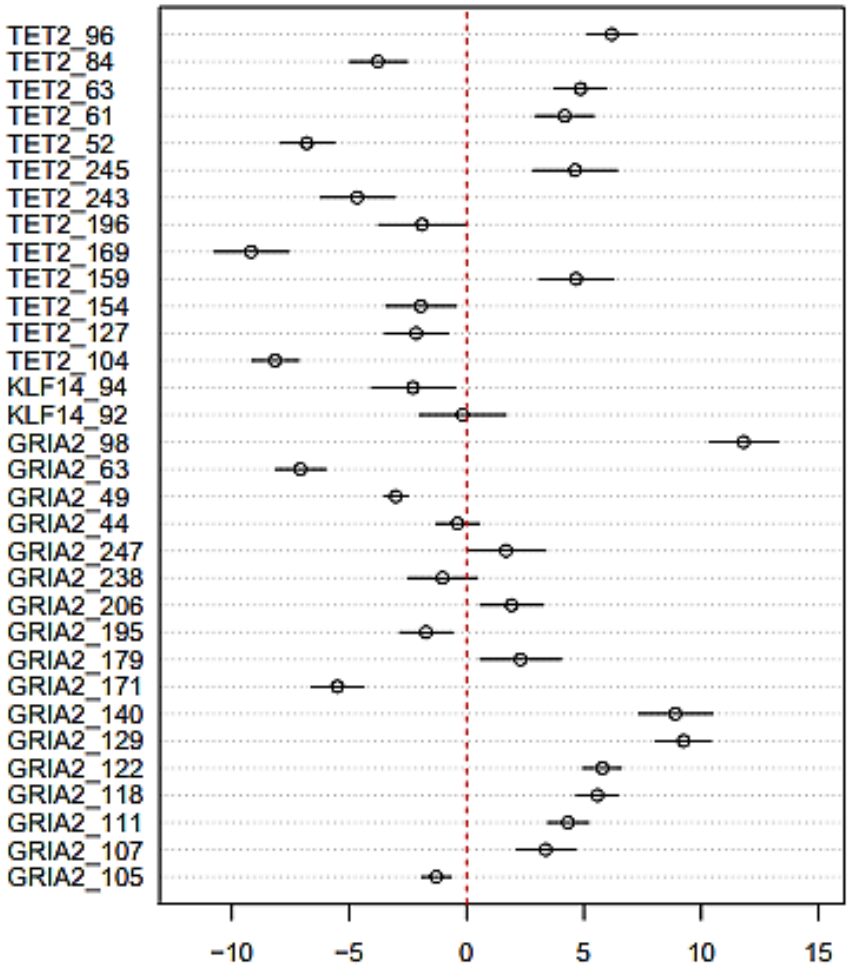


Figure 9: Effects of percent methylation of each site on whale age. It can be seen that some sites have a positive relationship between percent methylation and age, whereas others have a negative relationship. The Y-axis shows each CpG site, and the X-axis shows how percent methylation affects each site. The error bars for each site indicates the predicted error.

References

Allen, R. W. (2010). Identification Through DNA Analysis in Criminal and Family-Relatedness Investigations. In *Molecular Diagnostics* (pp. 381–398). Elsevier Inc. doi: 10.1016/B978-0-12-369428-7.00031-8

Ampure, A. (2016). *Instructions For Use*. Retrieved from www.beckmancoulter.com/customersupport/support.

Beal, A. P., Kiszka, J. J., Wells, R. S., & Eirin-Lopez, J. M. (2019). The bottlenose dolphin epigenetic aging tool (BEAT): A molecular age estimation tool for small cetaceans. *Frontiers in Marine Science*, *6*(SEP), 1–10. doi: 10.3389/fmars.2019.00561

Behjati, S., & Tarpey, P. S. (2013). What is next generation sequencing? *Archives of Disease in Childhood: Education and Practice Edition*, *98*(6), 236–238. doi: 10.1136/archdischild-2013-304340

Bell, C. G., Lowe, R., Adams, P. D., Baccarelli, A. A., Beck, S., Bell, J. T., Christensen, B. C., Gladyshev, V. N., Heijmans, B. T., Horvath, S., Ideker, T., Issa, J. J., Kelsey, K. T., Marioni, R. E., Reik, W., Relton, C. L., Schalkwyk, L. C., Teschendorff, A. E., Wagner, W., Zhang, K., & Rakyan, V. K. (2019). DNA methylation aging clocks: Challenges and recommendations. *Genome Biology*, *20*(1), 1–24. doi: 10.1186/s13059-019-1824-y

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. doi: 10.1093/bioinformatics/btu170

Cohen, J. E. (2003). Human population: the next half century. *Science*, 302, 1172-1175.

Corkeron, P., Hamilton, P., Bannister, J., Best, P., Charlton, C., Groch, K. R., Findlay, K., Rowntree, V., Vermeulen, E., & Pace, R. M. (2018). The recovery of North Atlantic right whales, Eubalaena glacialis, has been constrained by human-caused mortality. *Royal Society Open Science*, *5*(11). doi: 10.1098/rsos.180892

Cho, S., Jung, S.E., Hong, S.R., Lee, E.H., Lee, J.H., Lee, S.D., & Lee, H.Y. (2017). Independent validation of DNA-based approaches for age prediction in blood. *Forensic Science International: Genetics*, 29, 250–256.

Dahm, R. (2008). Discovering DNA: Friedrich Miescher and the early years of nucleic acid research. *Human Genetics*, Vol. 122, pp. 565–581. Springer. doi: 10.1007/s00439-007-0433-0

Dunshea, G., Duffield, D., Gales, N., Hindell, M., Wells, R.S., & Jarman, S.N. (2011). Telomeres as age markers in vertebrate molecular ecology. *Molecular Ecology Resources*, 11, 225–235.

FFPE Bisulfite Conversion. (2016). Retrieved from www.activemotif.com

Frasier, T. R., McLeod, B. A., Gillett, R. M., Brown, M. W., & White, B. N. (2007). 'Right Whales Past and Present as Revealed by Their Genes', in Kraus, S.M., and Rolland R.M. (ed.) The Urban Whale: North Atlantic Right Whales at the Crossroads. Cambridge, MA: Harvard University Press, pp. 200-231.

Freire-Aradas, A., Phillips, C., Mosquera-Miguel, A., Girón-Santamaría, L., Gómez-Tato, A., Casares de Cal, M., Alvarez-Dios, J., Ansede-Bermejo, J., Torres-Espanol, M., Schneider, P.M., Pospiech, E., Branicki, W., Carracedo, A., & Lareu, M.V. (2016). Development of a methylation marker set for forensic age estimation using analysis of public methylation data and the Agena Bioscience EpiTYPER system. *Forensic Science International: Genetics*, 24, 65–74.

Fujiwara, M., & Caswell, H. (2001). Demography of the endangered North Atlantic right whale. *Nature*, *414*(6863), 537–541. doi: 10.1038/35107054

Grönniger, E., Weber, B., Heil, O., Peters, N., Stäb, F., Wenck, H., Korn, B., Winnefeld, M., & Lyko, F. (2010). Aging and chronic sun exposure cause distinct epigenetic changes in human skin. *PLoS Genetics*, *6*(5), 6. doi: 10.1371/journal.pgen.1000971

Hamilton, P. K., Knowlton, A. R., & Marx, M. K., (2007). 'Right Whales Tell Their Own Stories: The Photo-Identification Catalog, in Kraus, S.M., and Rolland R.M. (ed.) The Urban Whale: North Atlantic Right Whales at the Crossroads. Cambridge, MA: *Harvard University Press*, pp. 75-104.

Hannum, G., Guinney, J., Zhao, L., Zhang, L., Hughes, G., Sadda, S. V., Klotzle, B., Bibikova, M., Fan, J.B., Gao, Y., Deconde, R., Chen, M., Rajapakse, I., Friend, S., Ideker, T., & Zhang, K. (2013). Genome-wide Methylation Profiles Reveal Quantitative Views of Human Aging Rates. *Molecular Cell*, *49*(2), 359–367. doi: 10.1016/j.molcel.2012.10.016

Harris, C. J., Davis, B. A., Zweig, J. A., Nevonen, K. A., Quinn, J. F., Carbone, L., & Gray, N. E. (2020). Age-Associated DNA Methylation Patterns Are Shared Between the Hippocampus and Peripheral Blood Cells. *Frontiers in Genetics*, 11, 1. doi: 10.3389/fgene.2020.00111

Hayatsu, H., Wataya, Y., Kai, K., & Iida, S. (1970). Reaction of Sodium Bisulfite with Uracil, Cytosine, and Their Derivatives. *Biochemistry*, *9*(14), 2858–2865. doi: 10.1021/bi00816a016

Holmes, E. E., & York, A. E. (2003). Using age structure to detect impacts on threatened populations: a case study with Stellar sea lions. *Conservation Biology*, *17*(6), 1794-1806.

Horvath, S. (2013). DNA methylation age of human tissues and cell types. *Genome Biology*, *14*(10), R115. doi: 10.1186/gb-2013-14-10-r115

Horvath, S., Gurven, M., Levine, M. E., Trumble, B. C., Kaplan, H., Allayee, H., Ritz, B.R., Chen, B., Lu, A.T., Rickabaugh, T.M., Jamieson, B.D., Sun, D., Li, S., Chen, W., Quintana-Murci, L., Fagny, M., Kobor, M.S., Tsao, P.S., Reiner, A.P., Edlefsen, K.L., Absher, D., & Assimes, T. L. (2016). An epigenetic clock analysis of race/ethnicity, sex, and coronary heart disease. *Genome Biology*, *17*(1), 171. doi: 10.1186/s13059-016-1030-0

Huang, Y., Yan, J., Hou, J., Fu, X., Li, L., & Hou, Y. (2015). Developing a DNA methylation assay for human age prediction in blood and bloodstain. *Forensic Science International: Genetics*, *17*, 129–136. doi: 10.1016/j.fsigen.2015.05.007

Illumina. (2018). *Nextera XT DNA Library Prep Kit Reference Guide For Research Use Only. Not for use in diagnostic procedures*. Retrieved from www.illumina.com/company/legal.html.

Illumina. (2018). *MiSeq System Guide*. Retrieved from www.illumina.com/company/legal.html.

Jarman, S. N., Polanowski, A. M., Faux, C. E., Robbins, J., de Paoli-Iseppi, R., Bravington, M., & Deagle, B. E. (2015). Molecular biomarkers for chronological age in animal ecology. *Molecular Ecology*, 24(19), 4826–4847. doi: 10.1111/mec.13357

Jiang, H., Schiffer, E., Song, Z., Wang, J., Zürbig, P., Thedieck, K., Moes, S., Bantel, H., Saal, N., Jantos, J., Brecht, M., Jeno, P., Hall, M.N., Hager, K., Manns, M.P., Hecker, H., Ganser, A., Dohner, K., Bartke, A., Meissner, C., Mischak, H., Ju, Z., & Rudolph, K. L. (2008). Proteins induced by telomere dysfunction and DNA damage represent biomarkers of human aging and disease. *Proceedings of the National Academy of Sciences of the United States of America*, 105(32), 11299–11304. doi: 10.1073/pnas.0801457105

Jung, M., & Pfeifer, G. P. (2015). Aging and DNA methylation. *BMC Biology*, 13(1), 1–8. doi: 10.1186/s12915-015-0118-4

Knowlton, A.R., Hamilton, P.K., Marx, M.K., Pettis, H.M., & Kraus, S.D. (2012). Monitoring North Atlantic right whale *Eubalaena glacialis* entanglement rates: A 30 yr retrospective. *Marine Ecology Progress Series* 466, 293–302.

Kraus, S.D., Brown, M.D., Caswell, H., Clark, C.W., Fujiwara, M., Hamilton, P.K., Kenney, R.D., Knowlton, A.R., Landry, S., Mayo, C.A., McLellan, W.A., Moore, M.J., Nowacek, D.P., Pabst, D.A., Read, A.J., & Rolland, R.M. (2005). North Atlantic right whales in crisis. *Science 309*, 561–562.

Kraus, S.D., Pace, R.M., & Frasier, T.R. (2007). 'High Investment, Low Return: The Strange Case of Reproduction in *Eubalaena glacialis'*, in Kraus, S.M., and Rolland R.M. (ed.) The Urban Whale: North Atlantic Right Whales at the Crossroads. Cambridge, MA: *Harvard University Press*, pp. 1-38.

Kraus, S. D., & Rolland, R. M. (2007). 'Right Whales in the Urban Ocean', in Kraus, S.M., and Rolland R.M. (ed.) The Urban Whale: North Atlantic Right Whales at the Crossroads. Cambridge, MA: *Harvard University Press*, pp. 1-38.

Lee, H. Y., Park, M. J., Choi, A., An, J. H., Yang, W. I., & Shin, K. J. (2012). Potential forensic application of DNA methylation profiling to body fluid identification. *International Journal of Legal Medicine*, *126*(1), 55–62. doi: 10.1007/s00414-011-0569-2

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. doi: 10.1093/bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. doi: 10.1093/bioinformatics/btp352Lindahl, T. (1993). Instability and decay of the primary structure of DNA. *Nature*, Vol. 362, pp. 709–715. Nature Publishing Group. doi: 10.1038/362709a0

Lubetkin, S. C., Zeh, J. E., Rosa, C., & George, J. C. (2008). *Age estimation for young bowhead whales (Balaena mysticetus) using annual baleen growth increments*. doi: 10.1139/Z08-028

Magoč, T., & Salzberg, S. L. (2011). FLASH: Fast length adjustment of short reads to improve genome assemblies. *Bioinformatics*, *27*(21), 2957–2963. doi: 10.1093/bioinformatics/btr507

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal*, *17*(1), 10. doi: 10.14806/ej.17.1.200

Moore, L. D., Le, T., & Fan, G. (2013). DNA methylation and its basic function. *Neuropsychopharmacology*, *38*(1), 23–38. doi: 10.1038/npp.2012.112

Morris, W. F., & Doak, D. F. (2002) *Quantitative Conservation Biology: Theory and Practice of Population Viability Analysis*. Sinauer Associates Inc., Sunderland, MA.

Park, J. L., Kwon, O. H., Kim, J. H., Yoo, H. S., Lee, H. C., Woo, K. M., Kim, S.Y., Lee, S.H., & Kim, Y. S. (2014). Identification of body fluid-specific DNA methylation markers for use in forensic science. *Forensic Science International: Genetics*, *13*, 147–153. doi: 10.1016/j.fsigen.2014.07.011

Polanowski, A. M., Robbins, J., Chandler, D., & Jarman, S. N. (2014). Epigenetic estimation of age in humpback whales. *Molecular Ecology Resources*, *14*(5), 976–987. doi: 10.1111/1755-0998.12247

Qiagen. (2020). *EpiTect ® Bisulfite Kit: Handbook*. Germantown, MD: Author.

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL. https://www.R-project.org/.

Sambrook, J., Russell, D.W. (2001). *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbour Laboratory Press. Cold Spring Harbour, MD.

Schick, R. S., Kraus, S. D., Rolland, R. M., Knowlton, A. R., Hamilton, P. K., Pettis, H. M., Kenney, R.D., & Clark, J. S. (2013). Using Hierarchical Bayes to Understand Movement, Health, and Survival in the Endangered North Atlantic Right Whale. *PLoS ONE*, *8*(6), 1–14. doi: 10.1371/journal.pone.0064166

Sease, J.L., & Loughlin, T. R. (1999). Aerial and land-based surveys of Stellar sea lions (*Eumetopias jubatus*) in Alaska, June and July 1997 and 1998. Technical memorandum NMFS-AFSC-100. U.S. Department of Commerce, Seattle

Shammas, M. A. (2011). Telomeres, lifestyle, cancer, and aging. *Current Opinion in Clinical Nutrition and Metabolic Care*, 14(1), 28–34. doi: 10.1097/MCO.0b013e32834121b1

Shapiro, R., Braverman, B., Louis, J. B., & Servis, R. E. (1973). Nucleic acid reactivity and conformation. II. Reaction of cytosine and uracil with sodium bisulfite. *Journal of Biological Chemistry*, *248*(11), 4060–4064. Retrieved from https://europepmc.org/article/med/4736082

Tanabe, A., Shimizu, R., Osawa, Y., Suzuki, M., Ito, S., Goto, M., Pastene, L.A., Fujise, Y., & Sahara, H. (2020). Age estimation by DNA methylation in the Antarctic minke whale. *Fisheries Science*, *86*(1), 35–41. doi: 10.1007/s12562-019-01371-7

Trumble, S. J., Robinson, E. M., Berman-Kowalewski, M., Potter, C. W., & Usenko, S. (2013). Blue whale earplug reveals lifetime contaminant exposure and hormone profiles. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(42), 16922–16926. doi: 10.1073/pnas.1311418110

Waldick, R. C., Kraus, S., Brown, M., & White, B. N. (2002). Evaluating the effects of historic bottleneck events: An assessment of microsatellite variability in the endangered, North Atlantic right whale. *Molecular Ecology*, *11*(11), 2241–2249. doi: 10.1046/j.1365-294X.2002.01605.x

Weidner, C., Lin, Q., Koch, C., Eisele, L., Beier, F., Ziegler, P., Bauerschlag, D.O., Jockel, K.H., Erbel, R., Muhleisen, T.W., Zenke, M., Brummendorf, T.H., & Wagner, W. (2014). Aging of blood can be tracked by DNA methylation changes at just three CpG sites. *Genome Biology*, 15, R24–12.

Wright, P. G. R., Mathews, F., Schofield, H., Morris, C., Burrage, J., Smith, A., Dempster, E.L., & Hamilton, P. B. (2018). Application of a novel molecular method to age free-living wild Bechstein's bats. *Molecular Ecology Resources*, *18*(6), 1374–1380. doi: 10.1111/1755-0998.12925

Zbieć-Piekarska, R., Spólnicka, M., Kupiec, T., Parys-Proszek, A., Makowska, Ż., Pałeczka, A., Kucharczyk, K., Ploski, R., & Branichki, W. (2015). Development of a forensically useful age prediction method based on DNA methylation analysis. Forensic Science International: *Genetics*, 17,173–179.