Gamete Compatibility and Reproductive Success in the North Atlantic Right Whale

By
Leah D. Springate

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

August, 2017, Halifax, Nova Scotia

Copyright Leah Springate, 2017

Approved:	Dr. Timothy Frasier Supervisor
Approved:	Dr. Zhongmin Dong Supervisor Committee
Approved:	Dr. Christa Brosseau Supervisor Committee
Approved:	Dr. Trevor Pitcher External Examiner
Date:	August 2017

Abstract

Gamete Compatibility and Reproductive Success in the North Atlantic Right Whale

by Leah Springate

The goal of this project was to assess the influence that genetic characteristics at gamete compatibility genes have on the reduced reproductive rates in the endangered North Atlantic right whale. To do this candidate genes were identified that play a role in fertilization, and primers were developed for the amplification of the putative functional sites of these genes. Mother-father-calf triads were sequenced to test for non-random mating patterns and non-Mendelian inheritance patterns, which would be indicative of mate choice and/or biased fertilization based on the characteristics of these genes. Overall there was low variability across individuals, but a slight bias in mate choice for mates with similar genotypes within loci, but differing across loci. One locus also showed signs of biased fertilization patterns, with successful fertilizations resulting when offspring inherit the same allele from both parents.

August 2017

Acknowledgements

First, I would like to thank my supervisor Tim Frasier for allowing me to work on this project, and for all his help and patience the last two years. He has increased my knowledge and understanding of genetics, as well as challenged me to think outside the box. I would also like to thank my committee members, Dr. Christa Brosseau and Dr. Zhongmin Dong for their guidance and support over the last two years. In addition, thank you to the other members of the Frasier Lab, including Suchinta, Christine, Brenna, Alysha, and Kiera for helping me out whenever I needed it.

Thank you to the lab of Dr. Bradley N. White, Trent University, as well as the North Atlantic Right Whale Consortium for processing the right whale samples and for providing access to the data.

Last, but not least, thank you to all my Halifax friends, especially Taylor Witty, Nat Drouin, and Megan Fitzpatrick for helping me outside of the lab by keeping me sane over the last two years!

Table of Contents

Abstract…………………………………………………………………………...2

Chapter 1: Overview ………………………………………………………………10

Chapter 2: Introduction……..……………………………………………………13

Chapter 3: Gamete compatibility genes in mammals: Candidates, applications and a potential path forward…………………………………………………………..37

3.1 Introduction ………………………………………………………………38

3.2 Gamete Structure ………………………………………………………...41

3.3 Fertilization ………………………………………………………………42

3.4 Potential Gamete Compatibility Genes …………………………………48

3.5 Applications ………………………………………………………………56

Chapter 4: Primer development for gamete compatibility genes in cetaceans ………85

4.1 Introduction ………………………………………………………………86

4.2 Methods …………………………………………………………………...90

4.3 Results and Discussion …………………………………………………...97

Chapter 5: Gamete compatibility and reproductive success in the North Atlantic right whale ……………………………………………………………………………...111

5.1 Introduction ……………………………………………………………..112

5.2 Methods …………………………………………………………………116

5.3 Statistical Analyses ……………………………………………………...119

5.4 Results …………………………………………………………………...121

5.5 Discussion ………………………………………………………………140

Chapter 6: Conclusion and Moving Forward…...…………………………………157

List of Figures

3.1 Overview of egg structure ……………………………………………………………..79

3.2 Overview of major structures, sites, and transport/guiding processes involved in fertilization ……………………………………………………………………………...80

3.3 Sperm structure and capacitation ………………………………………………81

3.4 Changes to spermatozoa during the acrosome reaction…………………………...82

3.5 Major steps in fertilization ………………………………………………………83

4.1 Testing annealing temperatures in fin whales (*Balaenoptera physalus*) to determine optimal PCR conditions ……………………………………………………………...92

4.2 Gel Image of Loci Cd9a (row 1) and PKDREJ-S1a (row 2) ………………………98

5.1 Observed vs. Expected Counts of Mating Pairs for Variable Site 448 at locus Cd9……………………………………………………………………………………...123

5.2 Observed vs. Expected Counts of Mating Pairs for Variable Site 185 at locus PKDREJ-S1b …………………………………………………………………………….124

5.3 Interaction effects for different combinations of mating pairs at locus PKDREJ-S1b site 185 …………………………………………………………………………….125

5.4 Observed vs. Expected Counts of Mating Pairs for Variable Site 464 at locus PKDREJ-S1b …………………………………………………………………………….126

5.5 Observed versus expected counts of mating pairs when comparing between loci……………………………………………………………………………………127

5.6 Interaction effects of Cd9a vs. PKDREJ-S1b site 185 ……………………………128

5.7 Interaction effects of Cd9a vs. PKDREJ-S1b site 464……………………………129

5.8 Interaction effects of PKDREJ-S1b site 185 vs. PKDREJ-S1b site 464…………130

5.9 Observed vs. Expected values for PKDREJ-S1b Site 185 ………………………131

5.10 Interaction Values for PKDREJ-S1b Site 185 and Sex …………………………132

5.11 Observed vs. Expected values for PKDREJ-S1b Site 464 ……………………...133

5.12 Interaction Values for PKDREJ-S1b Site 464 and Sex …………………………133

5.13 Observed vs. Expected values for Cd9 …………………………………………..134

5.14 Interaction Values for Cd9 and Sex ……………………………………………..135

5.15 Pedigrees for the North Atlantic right whale based on genotype ………………136

5.16 Probability of an offspring inheriting an allele that is the same as the other parent - being homozygous at locus PKDREJ-S1b site 464 …………………………………137

5.17 Probability of an offspring inheriting an allele that is the same as the other parent - being homozygous at locus Cd9a ……………………………………………………138

5.18 Probability of an offspring inheriting an allele that is the same as the other parent - being homozygous at locus PKDREJ-S1b site 185 …………………………………139

Appendix 1.1 Diagram representation of model used for Bayesian alternative to a contingency test ……………………………………………………………………..146

Appendix 2.1 Diagram representation of the Bayesian model used to test for biased inheritance patterns ……………………………………………………………...147

List of Tables

4.1 Accession numbers from GenBank for mRNA and genome sequences …………92

4.2 Primer sets and PCR conditions ……………………………………………………95

5.1 SNPs found in the North Atlantic right whale ……………………………………121

5.2 Variation counts for locus Cd9a ………………………………………………...121

5.3 Variation counts for locus PKDREJ-S1b: Site 185 ………………………………122

5.4 Variation counts for locus PKDREJ-S1b: Site 464 ………………………………122

List of Acronyms

SNP ……………………………………………………..single nucleotide polymorphism

ZP…………………………………………………………………………...zona pellucida

PCR …………………………………………………………...polymerase chain reaction

SAG…………………………………………………………………..surface active group

EBR1…………………………………………………………..egg bindin receptor 1

VE……………………………………………………….................vitelline envelope

VERL……………………………………………..vitelline envelope receptor for lysin

CRISP1…………………………………………………cysteine-rich secretory protein 1

CRISP2…………………………………………………cysteine-rich secretory protein 2

PKDREJ ……………polycystic kidney disease and receptor for egg jelly related gene

HFC………………………………………………..heterozygosity fitness correlations

GWAs…………………………………………………...genome wide association studies

dNTP………………………………………………….deoxyribonucleotide triphosphate

Chapter 1:

Overview

1. Objectives

The North Atlantic right whale is considered the world's most endangered large whale. A reduced reproductive rate is just one issue that impacts this species' recovery, thought to be caused by a lack of genetically dissimilar mates. However, a previous study showed that offspring inherit paternal alleles that differ from their inherited maternal alleles more often than expected. This results in offspring that are more heterozygous than expected based on Mendelian inheritance. Therefore, there is some mechanism that is promoting heterozygosity in this population. It is hypothesized that females select for genetically dissimilar gametes based on genes expressed on the surface of the sperm. These genes, referred to as gamete compatibility genes, have mainly been studied in marine invertebrates. By sequencing interactive sites at candidate gamete compatibility genes in mother-father-calf triads, it could be possible to identify the mechanism(s) driving these biased fertilization patterns. Specifically: (1) are certain combinations of mating pairs over- or under-represented; (2) do inheritance patterns at these genes differ from Mendelian inheritance patterns; and (3) is there a bias in which sex possesses the variable site? If there is a bias in mating pairs, this would suggest there is a mechanism for either pre- or post-copulatory mate choice in this population, and biased inheritance patterns would support the hypothesis of cryptic female choice, in that the female's egg is selecting for specific sperm.

2. Scope

Similar biased fertilization patterns are seen throughout nature, and identifying the genes and mechanisms behind them is considered a high priority. This research could

provide insight into how gamete compatibility can affect genetic variation in small populations, as well as their effect on individual's reproductive success. This research could also be applied in conservation genetics to potentially identify the mechanism behind low reproduction in populations with low genetic diversity, as well as address still unidentified processes involved in reproductive success, speciation, and the maintenance of genetic variation within small populations.

3. Outline

Chapter 2 provides a literature review of research relevant to the North Atlantic right whale, as well as an introduction into gamete compatibility in fertilization and speciation, and an overview of the importance of genetics in conservation biology. Chapter 3 is a review on the candidate genes and the justification of using these specific genes in this study, as well as a review of fertilization, and gamete compatibility in mammals. Chapter 4 discusses the development of lab protocols for sequencing these candidate genes in cetaceans, and the potential future uses. Chapter 5 contains the sequencing results, as well as statistical analyses and a discussion of the results. Chapter 6 is a summary of the main findings and suggestions for future research.

Chapter 2:

Introduction

1. The North Atlantic Right Whale

The North Atlantic right whale is considered the world's most endangered large whale (Perry *et al.* 1999). The North Atlantic right whale is a large baleen whale, with adults averaging 14 meters in length. Right whales are characterized by their large body, black coloration, lack of a dorsal fin, callosities on the head region, and a large strongly bowed lower jaw (Allen 1908; Andrews 1908). Callosity patterns are made from patches of cornified skin on the top of the head, along the jaw, and above the eyes and lips. The patterns are unique to each individual and are used to identify individuals in the field. Callosities appear white in color due to being infested by whale lice, are formed approximately 7-10 months after birth, and the placement usually remains stable throughout their lives (Payne *et al.* 1983; Kraus *et al.* 1986). The range of the North Atlantic right whale extends from Florida in the South to the Gulf of Saint Lawrence, Newfoundland and Greenland in the North (Kraus & Rolland 2007). However, in the last few years, the distribution of right whales appears to be shifting to new, unknown habitats (Pettis & Hamilton 2016). The whales move along the eastern coasts of Canada and the United States, putting them in harms way due to numerous major shipping lanes and areas of intensive fishing. In spring and early summer, right whales are found off the coast of New England. In the summer and fall, the whales move further north into Canadian waters, such as the Bay of Fundy and Brown's Bank. In the fall and early winter, pregnant females and also some juveniles, start to move south to the calving grounds off of the southeastern United States. It is still unknown where the remainder of the population goes during much of the winter (Perry *et al.* 1999; Kraus & Rolland 2007).

1.1 History

Industrial whaling began in the North Atlantic and was one of the longest and most intensive campaigns of wildlife exploitation in human history (Kraus & Rolland 2007). Right whales were one of the first species of whales to be commercially hunted (Perry *et al.* 1999). Right whales earned their name because they were considered the "right" whales to hunt, due to them being slow swimmers; they also yield high levels of oil and baleen, and usually float after death (Greene & Pershing 2004). Although conventional wisdom had historically been that this long period of extensive whaling is what reduced this species to its current small size, recent genetic analysis of bones from historic whaling sites are raising questions about this interpretation. First, these data show that early Basque whalers hunted the closely related bowhead whale (*Balaena mysticetus*) rather than the right whale (Rastogi *et al.* 2004; McLeod *et al.* 2008). Additionally, genetic analysis of historic right whale bones indicates that their genetic diversity was reduced *prior to* industrial whaling (McLeod *et al.* 2010). As a result, current understanding of right whale history, and the subsequent implications on their current status and recovery potential are in a state of revision.

1.2 Current Status

The North Atlantic right whale is considered the most endangered large whale, with an estimated 500 individuals remaining (Perry *et al.* 1999). Entanglement in fishing gear and ship strikes are the leading cause of death (Kraus 1990; Clapham *et al.* 1999; Perry *et al.* 1999; Kraus *et al.* 2005). Despite international protection since 1935, the species has shown little sign of recovery. Specifically, abundance has increased at an average of just

~2% per year, with substantial year-to-year variation (Fujiwara & Caswell 2001; Kraus *et al.* 2005). This is in contrast to many South Atlantic right whale populations, which were also intensively whaled, that are increasing at annual rates of ~7% (Best *et al.* 2001). However, anthropogenic mortalities are not the only cause of the low recovery rate in the North Atlantic right whale. This species also suffers from a reduced reproductive rate that is three times lower than their known potential (Frasier *et al.* 2007).

2. Mating System and Reproduction

2.1 Mating System

Right whale mating groups consist of one female and multiple males (anywhere from two to forty) (Mate *et al.* 2005; Frasier *et al.* 2007). These groups are called surface active groups, or SAGs. The female rolls onto her back, making copulation difficult, and the males compete for the spot closest to the female. When the female rolls over to breathe, the closest male is able to mate with her (Kraus & Hatch 2001). Through this mating group, females can promote competition throughout the males, and selectively mate with males of higher fitness, since these are the males that are able to maintain the close position. On average, one female may mate 60 times in a single SAG (Frasier *et al.* 2007). There is even a documented case of a single female mating simultaneously with two males (Mate *et al.* 2005).

Right whales also compete indirectly through sperm competition, and the right whale mating system is considered one of the most extreme examples of sperm competition in mammals (Frasier *et al.* 2007). In species with sperm competition, males tend to have

larger than expected testes (Ginsberg & Huck 1989, Preston *et al.* 2003), and right whales have the highest testes-weight to body-weight ratio in mammals, and one of the highest penis-length to body-length ratios in baleen whales (Frasier *et al.* 2007). The presence of sperm from multiple males in the female reproductive tract creates an environment that promotes sperm competition. This is an ideal situation for the female to select for advantageous sperm, which is known as cryptic female choice (Birkhead & Pizzari 2002). Thus, the right whale mating system has multiple levels of competition, precopulatory direct competition for the desired spot in a SAG, to post-copulatory sperm competition, and also cryptic female choice.

2.2 Reproductive Issues

The reduced recovery rate of the North Atlantic right whale is not solely attributed to anthropogenic causes (Kraus *et al.* 2005). Instead, North Atlantic right whales also suffer from a reduced reproductive rate that is three times lower than their potential. Based on the number of females in the current population, approximately 30-35 calves should be born per year, but the birth rate has an average of about 13 calves per year (Frasier *et al.* 2007). However, yearly counts can vary from as low as 1 calf up to 31. In 2016, only 13 calves were born, and this past year (2017) only 4 calves were born (Pettis & Hamilton 2016). A large amount of variation is also seen when assessing patterns of inter-birth intervals over time. Based on comparisons to the South Atlantic right whale, a normal interval would be 3 years: 1 year for pregnancy, 1 year for birth, lactation, and weaning, and 1 year for recovery. The average inter-birth interval for North Atlantic right whale females is 3.67 years, but fluctuated between 3 and 5 years until 2005 (Kraus & Rolland 2007; Browning

et al. 2010). By 2016, the average calving interval had increased to 6.6 years (Pettis & Hamilton 2016). There are also some adult females that have never given birth (12% in 2005) (Kraus & Rolland 2007; Browning *et al.* 2010). There is no one obvious cause for low fertility, but some hypotheses include disease events, environmental changes impacting food availability, toxins, and genetic factors (Browning *et al.* 2010, Kraus & Rolland 2007).

2.3 Biased Fertilization

Right whales have among of the lowest levels of genetic variability found in any wild population. However, previous studies suggest that biased fertilization processes are resulting in offspring with higher heterozygosity than expected given the genotypes of the identified parents. Specifically, Frasier et al. (2013) has shown that successful fertilizations are biased towards gametes that are genetically dissimilar. This resulted in an overall trend of increasing heterozygosity in calves born throughout the 25 years of data used for their study. The mating pairs were not more or less related than expected based on random mating, suggesting that this pattern was not due to mates preferentially mating with genetically dissimilar mates. Instead, it was found that offspring inherited paternal alleles that differed from the inherited maternal alleles more often than expected given Mendelian inheritance. Therefore, a postcopulatory mechanism, such as cryptic female choice for genetically dissimilar sperm, appears to be influencing fertilization patterns. This pattern could also be caused by differential mortality of zygotes. Female choice, in the sense of being able to identify gametes that are genetically dissimilar, would be an advantage from an evolutionary perspective, since losing a zygote most likely means the female will not become pregnant again that same year (Frasier *et al.* 2013).

3. Patterns of Biased Fertilization in Other Species

Biased fertilization patterns, like the ones seen in right whales, are widespread across nature, throughout different species (Olsson *et al.* 1996; Kempenaers *et al.* 1996; Tregenza & Wedell 2000; Foerster *et al.* 2003; Firman & Simmons 2008; Dziminski *et al.* 2008; Frasier *et al.* 2013). A few illustrative examples are provided below.

The Scandinavian wolf, *Canis lupus*, is highly inbred, and suffers from inbreeding depression. However, reproducing wolves that had higher levels of heterozygosity produced offspring that were more heterozygous than offspring of wolves that were randomly selected from the population. Heterozygosity is therefore maintained in this population by heterozygous wolves being preferentially selected as the breeders. Despite an increase in inbreeding over the years, there is a long-term mechanism for maintaining heterozygosity (Bensch *et al.* 2006).

In blue tits, *Parus caeruleus*, if the parents were more genetically similar they would have a higher proportion of unhatched eggs than compared to a mating pair that was genetically dissimilar. The females also pursue copulations with males that are different than their mates to increase heterozygosity in their offspring (Kempenaers *et al.* 1996).

Female Swedish sand lizards, *Lacerta agilis*, are promiscuous, mating with multiple males, which can include close relatives. It was shown that males who were less genetically related to the female sired more offspring than males that were more closely related to the female (Olsson *et al.* 1996).

In house mice, *Mus domesticus*, successful fertilizations were biased towards non-sibling mates when a female mated with a sibling and non-sibling. This result was seen regardless of mating order (Firman & Simmons 2008).

In the frog, *Crinia georgiana*, a breeding program showed that offspring fitness depended on specific combinations of parental haplotypes. Egg surface proteins could be selecting for the combination of compatible genotypes during polyandrous matings. The polyandrous females produced offspring that had the highest embryo survival, and the resulting tadpoles had the fastest development speeds (Dziminski *et al.* 2008).

Therefore, it seems there may be a mechanism, or mechanisms, that exist throughout nature that influence fertilization patterns of the gamete level, resulting in these biased fertilization patterns. The species discussed above all possess a mechanism for maintaining heterozygosity despite their mating systems, or lack of genetically dissimilar mates.

4. Polyandry and Cryptic Female Choice

Genetic compatibility may be one of the driving factors in the evolution of polyandry where females gain genetic benefits by mating with more than one male (Tregenza & Wedell 2000). Numerous studies have indicated that polyandrous females have increased fitness due to larger numbers of offspring and higher fitness of those offspring. Such effects appear to be due, at least in part, to a reduction in the chance of an egg being fertilized by a genetically incompatible male, and/or increase in the chance of obtaining a favorable genetic combination. Polyandry also creates the opportunity to select

for sperm from a specific male (cryptic female choice) (Ivy:2007kl; Madsen & Shine 1992; Zeh 1997; Tregenza & Wedell 1998, 2000; Simmons 2001; Dziminski *et al.* 2008).

Cryptic female choice is a generic term describing a range of mechanisms through which females can manipulate which sperm successfully fertilize their eggs. It is considered cryptic since the effects are concealed within the female reproductive tract, and as a result the specific mechanisms involved are not clear in most circumstances. Due to cryptic female choice being hidden, the effects are often masked by male-driven processes and more difficult to detect. Also, in species with internal fertilization, post-copulatory mechanisms are difficult to study. However, the effects are seen as biased fertilization patterns in offspring that are born (Birkhead & Pizzari 2002).

Cryptic female choice can result in either directional or non-directional sexual selection. There are only a few examples of directional cryptic female choice, which is when females bias sperm use in favor of the male phenotypes that are favored in mate choice. For instance, *Gallus gallus domesticus*, feral fowl, females mate with both socially dominant and non-dominant males, but prefer fertilization from socially dominant males. The females are capable of expelling semen after copulation, and the probability of this is negatively correlated with the social status of the male. In non-directional cryptic female choice, females are predicted to favor sperm from males with compatible genotypes regardless of phenotype. How the females would identify the genotype from the sperm is still unknown, but it is hypothesized that they can select for sperm based on non-recombining regions of the genome that are expressed on the surface of the sperm (Birkhead & Pizzari 2002).

5. Gamete Compatibility

The interaction of gametes is mediated through proteins on the surface of the cells (Vacquier 1998). The proteins that are involved in fertilization must have a species-specific structure to prevent cross-species fertilization. Genes that mediate fertilization tend to be more divergent than others in the genome, presumably to maintain this species-specific barrier (Swanson & Vacquier 2002). These reproductive proteins have been studied most extensively in sea urchins and abalones, due to the large availability and ease of study of gametes from species with external fertilization (Turner & Hoekstra 2008; Aagaard *et al.* 2009, 2013; Pujolar & Pogson 2011; Vacquier & Swanson 2011; Wilburn & Swanson 2015). Bindin was the first fertilization protein to be identified, and is found on the surface of sea urchin sperm (Wilburn & Swanson 2015). The main role of bindin is sperm-egg recognition. It attaches the sperm to the egg surface, aiding in the fusion of the sperm and egg membranes (Pujolar & Pogson 2011). Bindin's receptor is EBR1, a large protein found on the vitelline envelope (VE), which is the protective outer egg coat (Kamei & Glabe 2003; Wilburn & Swanson 2015). Like most proteins involved in gamete recognition, there is variation in bindin between species. The main role of bindin is sperm-egg recognition, and the female bindin genotype is a good indicator of the sperm genotype chosen for fertilization. The eggs "choose" sperm with whom they share the highest number of bindin alleles. Therefore, there is no "best" bindin genotype, instead, a bindin allele only has an advantage in the presence of a certain female genotype (Palumbi 1999).

Abalone, a marine gastropod, have two main sperm surface proteins: lysin and sp18. Lysin evolves extremely rapidly (Galindo *et al.* 2002), and is a species-specific protein that

is able to create a hole for sperm to pass through in the vitelline envelope of the egg (Aagaard *et al.* 2009, 2013; Vacquier & Swanson 2011). Lysin's receptor on the VE is a glycoprotein named VERL (Vitelline Envelope Receptor for Lysin). Lysin and VERL were the only pair of animal gamete recognition proteins known to bind to each other with high affinity (Galindo *et al.* 2002), until the identification of Juno and Izumo in mammals (Bianchi *et al.* 2014). These interactions result in non-random fertilization patterns, and can have large impacts on patterns of an individual's reproductive success within populations, and on the development of reproductive barriers during speciation (Turner & Hoekstra 2008). A similar mechanism to the one in marine invertebrates may exist in mammals (Swanson & Vacquier 2002). However, this has not been explored, and only one interacting pair of surface proteins has been identified in mammals (Bianchi *et al.* 2014). Gamete compatibility in mammals, and the candidate genes, are discussed in Chapter 3.

6. Genetics and Conservation Biology

The World Conservation Union has recommended genetic variation as one of the three biodiversity levels for conservation (Reed & Frankham 2003). Genetic diversity is required for populations to respond to environmental changes, which could include climate change, disease, pests and parasites, or predators (Frankham 2003). A unique example is the Tasmanian devil (*Sarcophilus harrisii*), which suffers from a contagious form of facial tumor disease, which is passed between individuals through biting (Hamede *et al.* 2012). Low diversity at genes of the Major Histocompatibility Complex (MHC) (immune system antigens that play a role in pathogen, tumor and graft recognition, and self/non-self

recognition) has allowed the tumor cells to pass from one individual to another, due to the cancer cells not being distinguished as non-self (Siddle *et al.* 2007).

Low heterozygosity levels, which can arise as a result of inbreeding, are directly linked to reduced population fitness (Reed & Frankham 2003). Across species, inbreeding can have negative consequences including reduced fecundity and lower survival of inbred individuals. One example is a population of song sparrows (*Melospiza melodia*) on Mandarte Island in British Columbia, in which inbred individuals are more likely to die in harsh winters than more outbred individuals (Keller *et al*. 1994).

Ignoring genetic factors in the management of endangered species can have unintended consequences. For example, in Illinois, the habitat of the greater prairie chicken was restored, but the population still did not recover. It was not until the population was outbred with birds from other states that the population recovered, with an increase in fertility and offspring (Frankham 2003; Frankham 2005). However, using a population for reintroduction regardless of the genetic diversity of the group can cause issues. This was the case in Australia when koalas were reintroduced. The founders were an island population that consisted of only two or three individuals. Due to such a small founding population, the current population suffers from inbreeding, low genetic diversity, and reproductive issues caused by testicular abnormalities (Seymour *et al.* 2001).

In Canada, specifically, there are issues with recovery of endangered species. A lack of implementation of the Species at Risk Act results in species being unable to recover. Species that are listed become illegal to kill/harm, and a critical habitat must be identified and protected. However, improvement in status (i.e., listed as special concern, threatened,

endangered) only occurred in 5.7% of the species between 1977 and 2013. Species on the list usually either decrease in status level (i.e. threatened to endangered) or remain at the same level, suggesting a possible failure in the legislation and/or implementation and enforcement (Favaro *et al.* 2014).

However, if the mechanisms behind genetic incompatibility can maintain heterozygosity even in fragmented populations, this shows that species may possess an intrinsic mechanism that can minimize the loss of genetic diversity. However, this mechanism and its implications are still generally overlooked in conservation biology (Frasier *et al.* 2013).

7. Gamete Compatibility and Sympatric Speciation

In addition to their important role in small populations and conservation, gamete compatibility genes also likely play an important role in the speciation process. Speciation can arise in one of two ways: through allopatry or sympatry. Allopatric speciation occurs when a population is divided geographically, and then is reproductively isolated. During this time of separation, biological reproductive barriers evolve that prevent reproduction between these groups once they are brought together again. Sympatric speciation, on the other hand, occurs when reproductive isolation between groups of individuals evolves within one population (i.e., in the absence of geographic separation) (Elmer & Meyer 2010).

One prevalent theory of sympatric speciation is speciation through sexual conflict. Since cooperation is needed between partners in sexual reproduction, the evolution of

sexual traits in one sex normally causes an evolutionary response in the other. Sexual conflict occurs when characteristics that boost the fitness of one sex reduce the fitness of the other (Birkhead & Pizzari 2002; Gavrilets & Waxman 2002). This creates a process of inter-sexual coevolution and also antagonistic coevolution between certain individuals that induces inter-sexual specialization (Birkhead & Pizzari 2002). If females selectively avoid mating with specific males, this would emphasize the incompatibilities, and in some cases, it can result in females genetically diversifying into two distinct groups. The males would then need to split and "chase" different female groups to avoid being stuck in the middle with low mating success (Gavrilets & Waxman 2002; Tregenza & Wedell 2000). This can result in reproductive isolation between populations, causing pre-zygotic isolation, population divergence, and finally, speciation (Birkhead & Pizzari 2002).

A global time tree of life shows consistent time to speciation across taxonomic groups. This suggests that a time-based acquisition of genetic incompatibilities, and not adaptive change, is driving reproductive isolation (Hedges *et al.* 2015). Under such a scenario, the raw material needed to cause reproductive isolation (different gamete compatibility alleles, with different affinities for one another) would already be segregating within species, and reproductive isolation could be attained through divergence in frequencies of numerous pre-existing, polymorphic, small-effect compatibility alleles (Corbett-Detig *et al.* 2013).

8. Research Objective

Although the North Atlantic right whale is negatively affected by many anthropogenic factors, the reduced reproductive rate is a main problem. However, the fact

that heterozygosity is maintained at higher levels than expected shows that there is some mechanism in play that helps to maintain genetic diversity in this population (Frasier *et al.* 2013). By looking for variation at the interactive sites of candidate gamete compatibility genes, it could be possible to identify the mechanism(s) that drives these biased fertilization patterns. If gamete compatibility plays a role through cryptic female choice for dissimilar gametes, biased fertilization patterns should be evident in the offspring. Specifically, by sequencing mother-father-calf triads, and testing for bias in mating pairs and in inheritance patterns, this research can shed light on whether variation at these gamete compatibility genes affect mate choice, and/or influence cryptic female choice and individual reproductive success in this population.

In many species, the mechanism that drives biased fertilization patterns remains unknown, so identifying the genes and mechanisms is considered a high priority (Tregenza & Wedell 2000; Birkhead & Pizzari 2002; Mays Jr & Hill 2004). Thus, this research could provide some insight as to how gamete compatibility maintains genetic variation in small populations, and how it affects reproductive success. It can also be applied to conservation genetics to potentially identify the mechanism that causes reduced fecundity in populations with low genetic diversity. This analysis also has the potential to address still unidentified processes involved in reproductive success, speciation, and the maintenance of genetic variation in small populations.

References

Aagaard JE, Springer SA, Soelberg SD, Swanson WJ (2013) Duplicate Abalone Egg Coat Proteins Bind Sperm Lysin Similarly, but Evolve Oppositely, Consistent with Molecular Mimicry at Fertilization. *PLoS Genetics*, **9**, e1003287–11.

Aagaard JE, Vacquier VD, MacCoss MJ, Swanson WJ (2009) ZP Domain Proteins in the Abalone Egg Coat Include a Paralog of VERL under Positive Selection That Binds Lysin and 18-kDa Sperm Proteins. *Molecular Biology and Evolution*, **27**, 193–203.

Allen JA (1908) The North Atlantic right whale and its near allies. *Bulletin of the American Museum of Natural History*, **24**, 227–329.

Amos W, Balmford A (2001) When does conservation genetics matter? *Heredity*, **87**, 257–265.

Andrews RC (1908) Notes upon the external and internal anatomy of Balaena glacialis Bonn. *Bulletin of the American Museum of Natural History*, **24**, 171–182.

Bensch S, Andrén H, Hansson B *et al.* (2006) Selection for heterozygosity gives hope to a wild population of inbred wolves. *PLOS ONE*, **1**, e72.

Best PB, Brandão A, Butterworth D (2001) Demographic parameters of southern right whales off South Africa. *Journal Cetacean Resource Management (Special Issue)*, 161–169.

Bianchi E, Doe B, Goulding D, Wright GJ (2014) Juno is the egg Izumo receptor and is essential for mammalian fertilization. *Nature*, **508**, 483–487.

Birkhead TR, Pizzari T (2002) Postcopulatory Sexual Selection. *Nature Reviews Genetics*, **3**, 262–273.

Browning CL, Rolland RM, Kraus SD (2010) Estimated calf and perinatal mortality in western North Atlantic right whales (Eubalaena glacialis). *Marine Mammal Science*, **26**, 648–662.

Clapham PJ, Young SB, Brownell RL (1999) Baleen whales: conservation issues and the status of the most endangered populations. *Mammal review*, **29**, 37–62.

Corbett-Detig RB, Zhou J, Clark AG, Hartl DL, Ayroles JF (2013) Genetic incompatibilities are widespread within species. *Nature*, **504**, 135–137.

Dziminski MA, Roberts JD, Simmons LW (2008) Fitness consequences of parental compatibility in the frog Crinia georgiana. *Evolution*, **62**, 879–886.

Elmer KR, Meyer A (2010) Sympatric speciation without borders? *Molecular Ecology*, **19**, 1991–1993.

Favaro B, Claar DC, Fox CH *et al.* (2014) Trends in Extinction Risk for Imperiled Species in Canada. *PLOS ONE*, **9**, e113118–8.

Firman RC, Simmons LW (2008) Polyandry Facilitates Postcopulatory Inbreeding Avoidance in House Mice. *Evolution*, **62**, 603–611.

Foerster K, Delhey K, Johnsen A, Lifjeld JT (2003) Females increase offspring heterozygosity and fitness through extra-pair matings. *Nature*, **425**, 714–717.

Frankham R (2003) Genetics and conservation biology. *Comptes Rendus Biologies*, **326**, 22–29.

Frankham R (2005) Genetics and extinction. *Biological Conservation*, **126**, 131–140.

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, Hamilton PK, Brown MW *et al.* (2007) Patterns of male reproductive success in a highly promiscuous whale species: the endangered North Atlantic right whale. *Molecular Ecology*, **16**, 5277–5293.

Fujiwara M, Caswell H (2001) Demography of the endangered North Atlantic right whale. *Nature*, **414**, 537–541.

Galindo BE, Moy GW, Swanson WJ, Vacquier VD (2002) Full-length sequence of VERL, the egg vitelline envelope receptor for abalone sperm lysin. *Gene*, **288**, 111–117.

Gavrilets S, Waxman D (2002) Sympatric speciation by sexual conflict. *Proceedings of the National Academy of Sciences*, **99**, 10533–10538.

Ginsberg JR, Huck UW (1989) Sperm competition in mammals. *Trends in Ecology & Evolution*, **4**, 74–79.

Greene CH, Pershing AJ (2004) Climate and the conservation biology of North Atlantic right whales: the right whale at the wrong time? *Frontiers in Ecology and the Environment*, **2**, 29–34.

Hamede RK, McCallum H, Jones M (2012) Biting injuries and transmission of Tasmanian devil facial tumour disease. *Journal of Animal Ecology*, **82**, 182–190.

Hedges SB, Marin J, Suleski M, Paymer M, Kumar S (2015) Tree of life reveals clock-like speciation and diversification. *Molecular Biology and Evolution*, **32**, 835–845.

Kamei N, Glabe CG (2003) The species-specific egg receptor for sea urchin sperm adhesion is EBR1, a novel ADAMTS protein. *Genes & development*, **17**, 2502–2507.

Keller LF, Arcese P, Smith J *et al*. (1994) Selection against inbred song sparrows during a natural population bottleneck. *Nature*, **372**, 356-357.

Kempenaers B, Adriaensen F, Van Noordwijk AJ, Dhondt AA (1996) Genetic Similarity, Inbreeding and Hatching Failure in Blue Tits: Are Unhatched Eggs Infertile? *Proceedings of the Royal Society B: Biological Sciences*, **263**, 179–185.

Kraus SD (1990) Rates and Potential Causes of Mortality in North Atlantic Right Whales (*Eubalaena Glacialis*). *Marine Mammal Science*, **6**, 278–291.

Kraus SD, Hatch JJ (2001) Mating strategies in the North Atlantic right whale (*Eubalaena glacialis*). *Journal of Cetacean Research and Management Special Issue*, **Special Issue 2**, 237–244.

Kraus SD, Rolland RM (Eds.) (2007) *The Urban Whale*. Harvard University Press.

Kraus SD, Brown MW, Caswell H *et al.* (2005) North Atlantic Right Whales in Crisis. *Science*, **309**, 561–562.

Kraus SD, Moore KE, Price CA *et al.* (1986) The use of photographs to identify individual North Atlantic right whales (*Eubalaena glacialis*). *Report of the International Whaling Commission*, **Special Issue 10**, 145–151.

Madsen T, Shine R (1992) Determinants of reproductive success in female adders, *Vipera berus*. *Oecologia*, **92**, 40–47.

Mate B, Duley P, Lagerquist B (2005) Observations of a female north atlantic right whale (*Eubalaena glacialis*) in simultaneous copulation with two males: Supporting evidence for sperm competition. *Aquatic Mammals*, **31**, 157–160.

Mays Jr HL, Hill GE (2004) Choosing mates: good genes versus genes that are a good fit. *Trends in Ecology & Evolution*, **19**, 554–559.

McLeod B, Brown M, Frasier T, White B (2010) DNA profile of a sixteenth century western North Atlantic right whale (*Eubalaena glacialis*). *Conservation Genetics*, **11**, 339-345.

McLeod B, Brown M, Moore M *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.

Olsson M, Shine R, Madsen T, Gullberg A, Tegelstrom H (1996) Sperm selection by females. *Nature*, **383**, 585.

Palumbi SR (1999) All males are not created equal: fertility differences depend on gamete recognition polymorphisms in sea urchins. *Proceedings of the National Academy of Sciences*, **96**, 12632–12637.

Payne R, Brazier O, Dorsey EM *et al.* (1983) External features in southern right whales (*Eubalaena australis*) and their use in identifying individuals. In: *Communication and behavior of whales* (ed Payne R), pp. 371–445. Boulder, CO.

Perry SL, DeMaster DP, Silber GK (1999) The Great Whales: History and Status of Six Species Listed as Endangered Under the U.S. Endangered Species Act of 1973. *Marine Fisheries Review*, **61**, 1–74.

Pettis HM, Hamilton PK (2016) North Atlantic Right Whale Consortium 2016 Annual Report Card. *Report to the North Atlantic Right Whale Consortium*, 1–14.

Preston BT, Stevenson IR, Pemberton JM, Coltman DW, Wilson K (2003) Overt and covert competition in a promiscuous mammal: the importance of weaponry and testes size to male reproductive success. *Proceedings of the Royal Society B: Biological Sciences*, **270**, 633–640.

Pujolar JM, Pogson GH (2011) Positive Darwinian selection in gamete recognition proteins of Strongylocentrotus sea urchins. *Molecular Ecology*, **20**, 4968–4982.

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.

Reed DH, Frankham R (2003) Correlation between fitness and genetic diversity. *Conservation biology*, **17**, 230–237.

Seymour AM, Montgomery ME, Costello BH *et al.* (2001) High effective inbreeding coefficients correlate with morphological abnormalities in populations of South Australian koalas (*Phascolarctos cinereus*). *Animal Conservation*, **4**, 211–219.

Siddle HV, Kreiss A, Eldridge MDB *et al.* (2007) Transmission of a fatal clonal tumor by biting occurs due to depleted MHC diversity in a threatened carnivorous marsupial. *Proceedings of the National Academy of Sciences*, **104**, 16221–16226.

Simmons LW (2001) The evolution of polyandry: an examination of the genetic incompatibility and good-sperm hypotheses. *Journal of Evolutionary Biology*, **14**, 585–594.

Swanson WJ, Vacquier VD (2002) The Rapid Evolution of Reproductive Proteins. *Nature Reviews Genetics*, **3**, 137–144.

Tregenza T, Wedell N (1998) Benefits of multiple mates in the cricket *Gryllus bimaculatus*. *Evolution*, **52**, 1726.

Tregenza T, Wedell N (2000) Genetic compatibility, mate choice and patterns of parentage: invited review. *Molecular Ecology*, **9**, 1013–1027.

Turner LM, Hoekstra HE (2008) Causes and consequences of the evolution of reproductive proteins. *The International Journal of Developmental Biology*, **52**, 769–780.

Vacquier VD (1998) Evolution of gamete recognition proteins. *Science*, **281**, 1995–1998.

Vacquier VD, Swanson WJ (2011) Selection in the Rapid Evolution of Gamete Recognition Proteins in Marine Invertebrates. *Cold Spring Harbor Perspectives in Biology*, **3**, a002931–a002931.

Wilburn DB, Swanson WJ (2015) From molecules to mating: Rapid evolution and biochemical studies of reproductive proteins. *Journal of proteomics*.

Zeh JA (1997) Polyandry and enhanced reproductive success in the harlequin-beetle-riding pseudoscorpion. *Behavioral Ecology and Sociobiology*, **40**, 111–118.

Chapter 3:

Gamete compatibility genes in mammals: Candidates, applications and a potential path forward

(Formatted for Royal Society Open Science)

Gamete compatibility genes in mammals: Candidates, applications and a potential path forward

Leah Springate and Timothy R. Frasier

Department of Biology, Saint Mary's University, 923 Robie St. Halifax, Nova Scotia B3H 3C3, Canada

Corresponding Author:

Timothy R. Frasier

E-mail: timothy.frasier@smu.ca

Tel: 1-902-491-6382

Abstract

Fertilization represents a critical stage in biology, where successful alleles of a previous generation are shuffled into new arrangements and subjected to the forces of selection in the next generation. Although much research has been conducted on how variation in morphological and behavioural traits lead to variation in fertilization patterns, surprisingly little is known about fertilization at a molecular level, and specifically about how genes expressed on the sperm and egg themselves influence fertilization patterns. In mammals, several genes have been identified whose products are expressed on either the sperm or the egg, and which influence the fertilization process, but the specific mechanisms are not yet known. Additionally, in 2014 an interacting pair of proteins was identified: "Izumo" on the sperm, and "Juno" on the egg. With the identification of these genes comes the first opportunity to understand the molecular aspects of fertilization in mammals, and to identify how the genetic characteristics of these genes influence fertilization patterns. Here, we review recent progress in our understanding of fertilization and gamete compatibility in mammals, which should provide a helpful guide to researchers interested in untangling the molecular mechanisms of fertilization and the resulting impacts on population biology and evolutionary processes.

Keywords: cryptic female choice, gamete compatibility, mate compatibility, post-copulatory sexual selection

1. Introduction

The process of fertilization represents one of the most important steps in population biology and evolution: being the stage at which successful alleles of a previous generation are shuffled into new combinations and packaged as individuals to be subjected to the forces of selection in the next generation. Factors that influence patterns of fertilization have subsequent impacts on the variance in reproductive performance across individuals, which ultimately influences the reproductive and growth potential of a population [1]. Moreover, factors that divide a population into groups of individuals where intragroup fertilization occurs more readily than intergroup fertilization can ultimately lead to speciation [2]. Given these wide-ranging effects, improving our understanding of the forces shaping patterns of fertilization is a goal of biologists across a broad range of specialties. Although much research has been conducted on how variation in morphological and behavioural traits lead to variation in fertilization patterns [3–5], surprisingly little is known about how genes expressed on the sperm and the eggs themselves - and which represent the crucial "locks and keys" needed for successful fertilization - influence fertilization success and fitness.

Fertilization is mediated by a complex series of interactions between the sperm and the egg, key steps of which require complimentary interactions between proteins expressed on the surface of each gamete [6–8]. The properties of these proteins influence how compatible a sperm and egg are, in terms of potential for successful fertilization, which is often referred to as their “gametic compatibility”. Until recently, the complimentary genes controlling this compatibility had only been well-described and studied in a few marine invertebrates, most notably abalone [9–12] and urchins [13–16]. In both taxa, the genotype of the receptor on the egg directly determines which sperm genotypes are capable of

fertilization [9,12,14]. These sperm-egg interactions, and resulting non-random fertilization patterns, scale-up to having large impacts on patterns of individual reproductive success within populations [14], and on the development of reproductive barriers during speciation [16–19]. The complimentary nature of these genes means that there is not one “best” genotype, but rather what genotype is “good” depends on the genotype of the corresponding gamete. This sort of epistasis, or non-additive interaction between alleles, provides an important mechanism for maintaining genetic diversity within populations and for providing the raw material to drive reproductive isolation and speciation through the presence of segregating incompatible alleles [20]. Although the benefits of genes influencing gamete compatibility are clear in species with external fertilization such as these, to ensure eggs are fertilized by sperm of the correct species, much evidence exists that such genes are important in organisms with internal fertilization as well, such as mammals.

Genes involved in reproduction, and gamete compatibility in particular, have proven to be among the fastest evolving genes in organisms yet studied (along with those of the immune system) [11,21,22]. There are three primary hypotheses, which are not mutually exclusive, regarding the underlying selection pressures [21]. The first is sperm competition, where the genetic compliment of a sperm could influence its chances of fertilization success at many stages throughout the fertilization process, and intense competition among sperm could then lead to rapid evolution of the associated genes. Second is sexual selection, where particular sperm-egg combinations have higher success rates than others, leading to the continual co-evolution of genes expressed on the gametes of both sexes. Lastly is sexual conflict, where selection on eggs to block polyspermy, and intense competition among sperm, provide conflicting selection pressures on the gametes

(on eggs to make multiple fertilization difficult, and on sperm to more rapidly fertilize the egg). This process would lead to a continual co-evolution of the genes involved in such strategies. Related to the 'sexual selection' hypothesis is the rapid divergence in gamete compatibility genes often found between recently diverged species [17,19,23]. Selection against cross-species fertilization could lead to rapid divergence of gamete compatibility genes, relative to other parts of the genome, and thus lead to effective reproductive boundaries between taxa.

In mammals, many genes have been identified whose products are expressed on either the sperm or the egg, and are somewhat involved in gametic compatibility, but the specific interactions and mechanisms are not yet known [22,24–28]. However, this changed in 2014 when a pair of genes (called “Izumo” for the sperm surface protein, and “Juno” for the complimentary egg receptor) was identified with a specific ligand-receptor relationship [29,30]. With the identification of these genes comes the first opportunity to understand the details and mechanisms of gamete compatibilities in mammals, to identify how genetic variation at these genes influence fertilization patterns and fitness, and to assess the subsequent implications for the development of reproductive barriers and speciation.

Given this recent progress, it seems timely to review our state of knowledge of these candidate genes and the processes that they influence. Here, we provide a brief overview of the structure of mammalian gametes and the fertilization process in mammals, review what is known about key candidate genes involved, and provide a brief review of the key areas where the analysis of such genes may be fruitful. The hope is that such a review will provide motivation, as well as a guide, for researchers interested in untangling the mechanisms of gamete compatibility and the resulting impacts on population biology and evolutionary processes.

2. Gamete Structure

2.1 The Egg

An ovulated mammalian egg is surrounded by two key layers (**Figure 1**). The first is an outer layer of cumulus cells that are contained in an extracellular matrix composed mainly of hyaluronic acid (also called the cumulus oophorus) [8,31]. Cumulus cells promote oocyte growth and development, and secrete progesterone, which is likely one of the chemoattractants---attracting the sperm to the egg [32,33]. The cumulus oophorus has many soluble factors and hormones that affect the egg and sperm in several ways, such as coordination of oocyte maturation and transport, and stimulating sperm motility [31]. The cumulus cells also act as the first of several barriers to spermatozoa, representing a dense mass through which only sperm that have undergone proper initial steps (see below) can pass. Interestingly, the cumulus layer is shed shortly after ovulation in marsupials and monotremes, and is therefore only a key player in fertilization within eutherian mammals [34,35]

The second major layer is a thick glycoprotein layer called the zona pellucida (ZP) [7]. It often serves as a species-selective barrier for sperm, and the binding of sperm to the ZP represents the first (of two) major interactions between the sperm and the egg. The ZP is composed of three different glycoproteins in most mammals: ZP1, ZP2, and ZP3; but humans and other primates have an additional glycoprotein, ZP4 [36,37]. Alterations to the ZP after fertilization prevent polyspermy and protect the early-stage embryo [35].

Interior to the ZP is the perivitelline space, which separates the egg proper from the zona pellucida, creating an area of protection [38]. Lastly, the egg is surrounded by a plasma membrane, to which spermatozoa bind during fertilization.

2.2 The Sperm

Mammalian sperm can be divided into three main sections: the head, midpiece, and tail, with lengths and characteristics that vary across species [39]. The sperm head contains the nucleus and the acrosome. The nucleus contains the haploid genome needed for fertilization. The acrosome is a secretory organelle that covers the first two-thirds of the sperm head, and is key to the binding of spermatozoa to the egg (**Figure 3**) [39,40]. The midpiece is the central segment that connects the tail to the sperm head. It contains a central filamentous core surrounded by a large number of mitochondria as energy suppliers for the spermatozoa. The tail, or flagellum, is the longest part of the sperm, and is responsible for propulsion to the site of the egg [39].

3. Fertilization

Fertilization is an extremely complex, multi-step process of which many details remain poorly understood. For our purposes, we will consider the processes involved in fertilization that occur post copulation. This delineation point is arbitrary, and prior aspects of reproduction such as the structure of reproductive organs, and mating systems and behaviour, obviously influence patterns of fertilization. However, contrary to the wide

variation in these characteristics across mammals, there is much similarity in the process once sperm have entered the female reproductive tract. Therefore, this stage serves as a suitable starting point for examining the context and processes associated with the molecular aspects of fertilization.

When sperm first enter the reproductive tract, two main obstacles can be envisioned: evading the female immune system, and targeting movement toward the egg. The importance of the former can be seen from data where seminal fluid triggers an invasion of antisperm antibodies and white blood cells into the vagina [41–43], that can proceed to break down the spermatozoa (motile sperm) [44,45]. Indeed, it is thought that avoiding such an immune response is why many species evolved genitalia capable of depositing spermatozoa directly into the uterus, or at least close to the cervix where they can then quickly be moved through the cervix into the uterus [46]. Components within the seminal plasma also appear to provide at least some additional protection from phagocytosis [47]. This immune response, and the media of the cervical mucus, are thought to limit the progress of a large portion of the sperm (including those that are malformed or damaged), whereas a small portion of morphologically normal sperm may proceed rapidly into the oviduct.

Numerous factors appear to aid the movement of sperm towards the egg, and the relative importance of each may vary across species. In general, four main factors are key [46,48,49]. First, uterine contractions can efficiently move large numbers of sperm through the initial components of the reproductive tract (vagina, cervix, uterus). Second, folds present in the tissues may serve as pathways directing sperm through the cervix and uterus, and towards the oviduct. Third, thermotaxis - or the movement of spermatozoa along a temperature gradient - helps guide sperm down the fallopian tube toward the site of

fertilization, which is one to two degrees warmer than the entrance of the fallopian tube [49,50]. The fourth, and perhaps most interesting, factor is chemoattraction: where sperm are attracted by chemical signals released from the egg [48,51]. Such a process is prevalent in the animal kingdom, but less well understood in mammals. For example, the specific chemicals used have been identified for many non-mammalian species, but have yet to be identified in mammals [49,51] (**Figure 2**).

3.1 Modification of Sperm Prior to Fertilization

Successful movement of the gametes towards one another is not the only hurdle to overcome for fertilization to be successful. Instead, early studies showed that spermatozoa cannot fertilize eggs immediately after ejaculation, but rather require an incubation time in the female reproductive tract before acquiring this potential [52–55]. These studies provided early indications of the complexity of fertilization, and the important role of environmental conditions within the female reproductive tract [52,55]. Although such an incubation time is required across mammals, the necessary length of time varies across species, ranging from ~1 hour in humans to ~5 hours in rabbits and cows [56–58]. These data demonstrated that the sperm must undergo modifications, that are triggered by the environment, in order to interact properly with the egg. Two such transitions are now known to occur: capacitation and the acrosome reaction.

3.1.1 Capacitation

As spermatozoa make their way through the uterus and into the oviductal isthmus, they become reversibly bound to the oviductal epithelium [51]. This is the stage at which

spermatozoa become capacitated (**Figure 2**). One of the key steps that appears to trigger this process is the removal (*in vitro*) or dilution (*in vivo*) of the seminal fluid associated with the spermatozoa, which is a known inhibitor of capacitation [55,59,60]. Not all spermatozoa undergo capacitation at the same time, however, and at any given time only a small portion (~10%) are capacitated, with a relatively high turnover rate of which sperm are capacitated and which are not [55].

The processes that take place during capacitation have two major effects on the fertilization abilities of the sperm. First, it is at this time when sperm become "hyperactivated". Hyperactivation usually involves increased amplitude and asymmetry in flagellar beating patterns, and appears necessary for the spermatozoa to break free from their bonds with the oviductal epithelium, complete their journey toward the egg, and penetrate the outer layers of the egg [61,62]. Second, it is during capacitation when the proteins needed for sperm-egg interactions become "unmasked" due to the removal and/or changes in the proteins present on the plasma membrane on the head of the sperm [55,63,64] (**Figure 3**). Thus, it is at this stage when the first proteins involved in sperm-egg interaction are exposed, and the genes underlying such proteins should be a key target in investigations into the molecular aspects of gametic compatibility.

After capacitation, spermatozoa move through the fallopian tube towards the egg, likely guided by a combination of thermotaxis, chemotaxis, and oviductal contractions. Only capacitated sperm can make their way through the cumulus cells surrounding the egg [55].

3.1.2 The Acrosome Reaction

The second major transition that must take place in the spermatozoa for fertilization to be successful is the acrosome reaction (AR). The acrosome is a secretory vessicle in the head of mammalian spermatozoa that is enclosed by a continuous acrosomal membrane. The membrane can be further divided into the inner acrosomal membrane, which is in close proximity to the nuclear membrane, and the outer acrosomal membrane, which is under the plasma membrane that covers the acrosome [65,66] (**Figure 4**). During the AR the plasma membrane and outer acrosomal membrane fuse, and the acrosomal contents are released. This process uncovers a new set of proteins that will interact with the plasma membrane of the egg during fertilization [66,67] (**Figure 3**). However, once the spermatozoa makes its way through the ZP and reaches the plasma membrane of the egg, the point of contact with the egg is not the tip of the spermatozoa, but rather the equatorial region on either side [8] (**Figure 4**). Thus, it is proteins expressed on these regions, after the acrosome reaction, that are likely key to gamete compatibility at this stage of fertilization.

It has historically been thought that the acrosome reaction is triggered when the proteins on the head of the spermatozoa that were exposed during capacitation interact with those on the ZP. Indeed, several studies have shown that the ZP, and ZP3 in particular, have sperm-binding capabilities and can also trigger the AR [68,69]. Additionally, during the AR enzymes are released that can dissolve the ZP, creating a hole through which spermatozoa can pass [70]. However, recent studies have shown that - at least in mice - this is not necessarily the case, and that spermatozoa can undergo the acrosome reaction prior to interaction with the ZP, and even prior to encountering the cumulus cells surrounding the egg [71,72]. Therefore, at this time the trigger(s) for the acrosome reaction, and the

exact location where it takes place, are not known. A role for an interaction with the ZP still seems likely, but what that role is, and how essential it is, are now unclear. One possibility is that interaction with the ZP3 may facilitate the completion of the AR, rather than being a key aspect of AR initiation [71].

3.2 Summary of Gamete Interaction Stages

In summary, there are two major stages where the proteins of the gametes interact with one another, and thus where the characteristics of these proteins may influence fertilization patterns. First is when the proteins on the head of the spermatozoa interact with those on the egg's zona pellucida. However, as stated above, it was originally thought that it was the proteins exposed during capacitation that interact with the ZP, triggering the acrosome reaction. However, it is now clear that many of the spermatozoa that bind to the ZP have already undergone the AR, and therefore it is likely proteins exposed on the head post-AR that are key to sperm-ZP interactions. Second is when those proteins exposed on the equatorial region of the spermatozoa during the acrosome reaction interact with those on the egg's plasma membrane (**Figure 4**). By considering which proteins are expressed when, and in what locations, it is therefore possible to identify a suite of potentially interacting candidate genes influencing gametic compatibility.

4. Potential Gamete Compatibility Genes

Below is a brief description of the genes that, at the time of this writing, have the most potential for being key players in gamete compatibility. It is divided into those found on each gamete. We re-iterate that only one interacting pair is currently known in mammals: Izumo on the sperm with Juno on the egg. However, the other genes described are known to influence gamete compatibility in some way, even though the details have not yet been worked out. We caution readers that this list represents many of the likely candidates given our current understanding; however, it is not completely exhaustive, and our understanding is still in its infancy. Therefore, some genes that may prove key in the future may not be included here, and some included here may be of limited use.

4.1 Sperm

4.1.1 Izumo1

The Izumo1 protein (named after a Japanese marriage shrine) has a large extracellular region, a single transmembrane region and a short cytoplasmic tail [74,75]. During the acrosome reaction, Izumo1 shifts from the anterior head of the sperm to the equatorial segment where fusion takes place [76]. Mice that lack the Izumo1 protein produce normal sperm that are capable of binding to, and penetrating, the zona pellucida, but which are unable to fuse with eggs. The sperm instead built up in the perivitelline space (the space in between the ZP and the plasma membrane of the egg) [77]. An inhibitory antibody bound to this section inhibited sperm-egg fusion but it did not affect sperm motility or egg binding

[74]. Binding is a necessary step where the sperm is attached to the egg, before fusion can take place (Figure 5) [31]. This suggests the inhibitory effect occurs during the sperm-egg fusion [74]. The putative functional sites where Izumo1 interacts with Juno have been identified, with amino acids 148-163 being particularly important [78,79].

4.1.2 CRISP1

CRISP1, also known as DE (due to showing up on non-denaturing gels as two bands – called proteins D and E [80]) is one member of the Cysteine-Rich Secretory Protein (CRISP) family [81]. Members of the CRISP family vary in their biological functions, are found in different mammalian tissues, and can even be found in the venom of snakes, lizards and snails [82,83]. The proteins are characterized by sixteen conserved cystein residues, with ten being clustered in the C-terminal domain [27]. CRISP1 is unique among the candidate genes considered here, in that it appears to be involved in both stages of sperm-egg interaction. There are two "populations" of CRISP1 expressed on spermatozoa. One loosely bound population that is involved in the initial binding of sperm to the ZP (and which are subsequently released from the sperm during the acrosome reaction); and a second, tightly bound, population that migrates to the equatorial region of the sperm head after the acrosome reaction and is subsequently involved in egg membrane binding [27]. Mice with a mutated CRISP1 gene were still fertile, but had decreased fusion ability in an environment that promoted sperm competition with healthy sperm [84]. Additionally, the masking of CRISP1 resulted in a significantly lower ability to fertilize eggs that had the cumulus cells and zona pellucida removed. Thus, sperm lacking CRISP1 have a disadvantage in their capacity to both interact with the zona pellucida and fuse with the egg

[84]. The egg-binding ability of CRISP1 is located in a specific 12 amino acid sequence known as Signature 2 [85]. However, another member of the CRISP family may compensate in sperm that are lacking CRISP1. CRISP2 may interact with common binding sites on the egg as CRISP1 [86], and CRISP2's Signature 2 region only differs from CRISP1 by two amino acids [84].

4.1.3 CRISP2

CRISP2, also called Tpx-1, is expressed exclusively in male haploid germ cells, and shows high homology (69%) to CRISP1 [86]. Unlike CRISP1, CRISP2 is not involved in ZP binding, and is only associated with binding to the plasma membrane of the egg. CRISP2 is expressed on the equatorial section of the sperm after the acrosome reaction, and experimental studies have found that an inhibitory CRISP2 antibody reduces the percentage of fertilized eggs, with sperm accumulating in the perivitelline space [86]. The antibody had no effect on ZP penetration, sperm motility, or the acrosome reaction [86]. A potential functional site was seen in human males: a polymorphism in exon 9 of CRISP2 resulted in sterility [87].

4.1.4 PKDREJ

PKDREJ is a large, intron-less gene that codes for an ~8-kb transcript in humans [88]. Its name is derived from the fact that it has high homology to two different types of genes: the PKD family and the REJ gene. The PKD family of genes code for membrane-bound proteins that form calcium ion channels and are involved in cell-cell and cell-extracellular

matrix interactions [89]. A region of PKDREJ is also homologous with the sea urchin REJ gene, which is involved in sperm-egg interaction [90]. The PKDREJ protein is located on the acrosome of the sperm head, suggesting that PKDREJ is involved in ZP binding [91]. Experimental evidence indicates that although PKDREJ is involved in ZP binding, it is not essential [92]. For example, male mice homozygous for a mutated PKDREJ allele could still fertilize eggs, but had lower fertilization success when in a competitive environment with normal sperm. This reduction was due to an increase in the amount of time needed for the acrosome reaction to occur [92]. The likely location of the functionally important segment of the PKDREJ gene is the region homologous to the REJ gene, which corresponds approximately to amino acids 280-800 in humans [25].

4.1.5 PH-20

PH-20 is a plasma membrane protein located on the sperm head as well as on the inner acrosomal membrane, the latter of which appears to be released during the acrosome reaction [93–95]. For many years it was thought that PH-20 is required for sperm binding to the ZP, and this requirement led to investigations of using medicinal blockage of PH-20 as a form of male contraception [93,95]. However, more recent studies suggest that this may not be universal because PH-20-null mice are still fertile [96]. PH-20 appears to have a dual role in fertilization [94,97]. First, PH-20 has enzymatic activity and these proteins covering the head of the sperm are important for penetrating the cumulus layer of cells surrounding the egg. Second, it has a non-enzymatic role in the secondary binding of spermatozoa to the ZP after the acrosome reaction [94,97,98]. The active site of PH-20 required for hyaluronic acid binding, a step in the ability to penetrate the cumulus layer,

has been identified at amino acid sites 205-235 [99], but the site required for secondary binding to the ZP has yet to be identified.

4.1.6 Zonadhesin

Zonadhesin is an acrosomal protein that is unique in its ability to bind to the zona pellucida in a species-specific manner [100]. It is localized on the outer acrosomal membrane and exposed during capacitation [101–103]. It differs between species due to rapid evolution and also domain duplication, mRNA splice variation, and processing heterogeneity during the functional maturation of the protein [102]. Sperm adhesion to the ZP, or sperm penetration was decreased when sperm cells were exposed to a zonadhesin antibody [103]. Additionally, mice that lack zonadhesin are fertile, but lost the species specificity of sperm-ZP fusion. This loss has not been seen with knock-out individuals of other sperm proteins [102]. A potential binding region is an exposed fragment of 30 amino acids in MAM (meprin/A5 antigen/mu receptor tyrosine phosphatase) domain 3 in mice. This section is characterized by a substantially increased rate of positively selected amino acid sites and exhibits high variability in predicted post translation modifications [104].

4.2 Eggs

4.2.1 The Zona Pellucida

The zona pellucida (ZP) is composed of three different glycoproteins: ZP1, ZP2, and ZP3. In humans and other primates, there is an additional glycoprotein, ZP4 [36]. ZP1 is

necessary for forming and maintaining the structural integrity of the zona pellucida [105]. Mice lacking ZP1 still have a zona pellucida, but it is thinner than normal, and has a poorly defined border. This disfiguration can lead to granulosa cells (which make up cumulus cells) accumulating in the perivitelline space, causing functional disorganization within the egg, and resulting in diffusion of the zona matrix. In functional studies of ZP1, ZP1-null mice had reduced fertilization rates (80% of ZP1-null mice were sterile), and those where fertilization was successful had litter sizes that were half those of normal mice [106]. The fact that ZP1-null mice can still be fertile indicates that ZP1 is not essential for proper sperm-egg interaction and fertilization.

In humans, ZP2 is responsible for secondary binding of the acrosome-reacted sperm (following initial binding with ZP3) [107]. ZP2 also provides an effective block to polyspermy. After fertilization, ZP2 is cleaved from the zona pellucida so that additional sperm are unable to bind to the early embryo, ensuring monospermic fertilization [107]. Mice without ZP2 are able to form a thin zona pellucida comprised of ZP1 and ZP3. However, the resulting ZP is not sustainable, and the resulting eggs are ZP free [105]. The absence of the zona pellucida has a negative effect on the development of the egg, resulting in sterility of that female [109].

Although, as stated above, the role of sperm binding with the ZP has been revised with respect to triggering the acrosome reaction, sperm-ZP binding is still an important step in fertilization, regardless of its role in the AR. For example, mice that lack ZP3 form oocytes without a zona pellucida, which results in sterility [105]. ZP3 is also thought to be responsible for the species-specific binding of sperm to the egg [70]. Although some studies have found indications of which specific regions directly influence sperm binding [110,111], other studies have obtained conflicting results [112,113], and therefore the key

regions involved remain unknown. ZP3 polypeptides do not appear to interact with the sperm directly, but rather do so via oligosaccharides that bind to the ZP3 polypeptides [70]. Thus, variation within the gene itself, as well as in the associated oligosaccharides, is responsible for the subsequent effects on fertilization. Indeed, previous studies have shown that this gene is under strong selection, causing rapid divergence between species [114].

The role of the human ZP4 is not yet well understood, and requires more research. Since it is structurally similar to ZP1, it has been assumed that ZP4 also plays a role in maintaining the structural integrity of the human zona pellucida [105].

4.2.2 Juno

Juno is the egg receptor for Izumo1 [29]. Previously called Folr4, this gene was renamed Juno after the Roman goddess of fertility and marriage, once it was recognized as the paired receptor for Izumo. Female mice that lack Juno are completely sterile. Juno is also rapidly shed after fertilization, which could provide an additional block to polyspermy [29]. The shedding of Juno creates a layer of "fake" eggs that could attract and bind acrosome-reacted sperm, preventing them from reaching the already fertilized egg. The interaction of Izumo1 and Juno is a necessary event for adhesion between acrosome-reacted sperm and the egg membrane [29]. Adhesion is the sustained interaction of sperm cells with the egg extracellular matrix that should lead to fertilization with normal sperm [31]. However, these proteins do not facilitate the following step, fusion, which is vital for successful fertilization [29]. Juno has a folate-binding pocket at amino acids 60-175; however, the binding site for Izumo1 has been identified as the surface behind this binding pocket, and specifically within amino acid sites 44-91 and 145-191 [78,79].

4.2.3 CD9

Another putative type of gene for moderating fusion on the egg surface is the tetraspanin family [75]. Tetraspanins are small transmembrane proteins that are thought to affect cell adhesion, motility, proliferation, differentiation and signalling. CD9 is a necessary tetraspanin for gamete fusion [115]. Knock-out mice that lack CD9 have severely reduced fertility. The sperm is able to penetrate the ZP and bind to the egg membrane, but the membranes are unable to fuse. The exact role of CD9 in sperm fusion is still unknown. Research suggests that it does not interact directly with a complementary protein on the sperm, but rather that it binds with another "egg fusion protein", causing a change in conformation, that then interacts with the sperm [75,115]. The functional sites of CD9 have been identified as part of the large extracellular loop 2, amino acids 173-175. A mutation at these amino acids results in eggs without fusion ability [115].

The genes discussed here primarily code for proteins thought to interact directly with complementary proteins on the other gamete. However, the products of some gamete compatibility genes, such as ZP3, bind to sugar molecules and it is this combined glycoprotein that is involved in gamete interactions [116–118]. This greatly increases the complexity of gamete interactions, with variation in the proteins themselves, the sugar molecules, and in the post-translational modification (glycosylation), potentially impacting gamete compatibility. However, we currently know little about the role of glycoproteins in mammalian gamete compatibility, outside of those involved with the ZP, but, given their importance in other taxonomic groups [117], it seems likely that they play an important role, the details of which remain to be discovered.

5. Applications

Understanding how characteristics at gamete compatibility genes influence patterns of fertilization has implications for a broad range of fields, ranging from reproductive biology to evolutionary and conservation genetics, to speciation. Here, we will briefly summarize some of these applications, highlighting how they can fill key gaps in our understanding.

In terms of reproductive biology and evolutionary genetics, patterns of non-random fertilization are widespread in nature [119–125]. These are often studied in the context of post-copulatory mate choice, where females are able to "choose" which sperm fertilize their eggs. In species where laboratory experiments are possible, this "choice" appears due to differential fertilization success of different types of sperm relative to the characteristics of each egg [124,124,125]. Despite the widespread nature of these patterns, however, the mechanisms involved have remained elusive. Thus, identifying the underlying genes and mechanisms has long-been regarded as a high priority [120,121,123]. Genes involved in gamete compatibility are clearly the most likely candidate genes influencing these non-random fertilization patterns [14], and their analyses will therefore shed much needed light on the issues of post-copulatory sexual selection, female choice, and evolutionary genetics.

Understanding these patterns also has large implications for the fields of conservation biology and conservation genetics. In many of the species where non-random fertilization patterns have been found, fertilizations are biased towards gametes that are genetically dissimilar [119–121,123,125–127]. The result is offspring with higher levels of heterozygosity than expected from a similar-sized random-mating population. In this way, this process can not only slow the decline of heterozygosity expected from genetic drift, but can also maintain heterozygosity at higher levels than expected in small populations.

Thus, these biased fertilization patterns can significantly counter the effects of genetic drift, and act to maintain genetic diversity in small populations [128,129]. Moreover, the resulting benefits (primarily offspring with high heterozygosity) have been proposed as one of the main driving forces behind the evolution of polyandry [130–133]. Obtaining a better understanding of how the characteristics of gamete compatibility genes shape fertilization patterns can therefore lead to a better understanding of the mechanisms through which patterns of genetic diversity influence reproductive performance and recovery potential in endangered species, and lead to a more thorough understanding of the evolution of different mating systems and strategies.

The process of speciation involves the evolution of reproductive barriers between closely related groups of individuals. Although the development of geographic barriers (resulting in allopatric populations) is often thought to be the trigger for the subsequent development of "biological" barriers, it is the presence of these biological barriers that often underlies where species lines are drawn [134,135]. Gamete compatibility genes are likely candidates for the initial development of reproductive incompatibilities between closely related groups of individuals [17–19,136–138]. Indeed, Gavrilets & Waxman (2002) [17] showed that if segregating alleles within a population result in females differing in their compatibility to different males, this will lead to different "groups" of reproductive (compatible) individuals. Over time, this can result in the sympatric evolution of reproductively isolated groups. Thus, gamete compatibility genes are likely a key factor in the evolution of biological reproductive barriers, and provide a clear path through which one species can sympatrically be split into two based solely on different fertilization patterns among existing alleles. A similar process is likely also important in many cases of allopatric speciation with gene flow, where differentiation at gamete compatibility genes

underlies the development of reproductive barriers. In this way, the analysis of these genes represents a promising approach for improving our understanding of how genetic characteristics influence the speciation process.

When trying to identify loci influencing specific traits, two approaches are generally used: the candidate gene approach where specific genes or loci are targeted for sequencing and analysis; and genome-wide association studies (GWAS) where tens of thousands of loci (generally single nucleotide polymorphisms, or SNPs) are analyzed to screen the genome for regions or loci showing an appropriate signature. The rapid evolution, and decreasing cost, of methods to characterize and genotype individuals at tens of thousands of SNPs has led to the rapid growth of our understanding of how genotype influences fitness and phenotype based on the GWAS approach [139–141]. However, such studies often involve two stages: the first involving the large-scale genome screening to identify loci with an appropriate signature, and then a second stage of further sequencing and characterization of the area around the SNP originally identified. Therefore, the candidate gene approach may still be a more efficient option in cases where putative candidate genes have been identified [142]. Much research has been conducted on potential gamete compatibility genes, and their likely roles in the fertilization process. Therefore, the goal of this review was to bring this wealth of literature together into one cohesive paper and framework, and to create a list of candidate genes that hold the most potential for success, and therefore serve as a guide for future studies. Moreover, given the broad range, and importance, of processes influenced by gamete compatibility genes, we hope that his paper will serve as motivation for more researchers to pursue this line of inquiry.

Ethics Statement

This study did not involve the use of humans or animals, and therefore did not require any ethical or animal care permits, permissions, licenses, or approval.

Data Accessibility

As a review, this paper did not generate new data of its own. Therefore, the data deposition and accessibility requirements for this journal are not applicable.

Competing Interests

We do not have any competing interests to declare.

Author's Contributions

TF developed the idea for this review paper as part of LS's M.Sc. thesis. Both TF and LS conducted the literature review, worked together to develop its structure and main points, and shared the duties of writing the manuscript.

Acknowledgements

We thank two anonymous reviewers whose comments greatly improved the manuscript.

Funding Statement

This work was funded by a National Sciences and Engineering Research Council of Canada (NSERC) Discovery Development Grant to TF.

References

1.López-Sepulcre A, Norris K, Kokko H. 2009 Reproductive conflict delays the recovery of an endangered social species. *Journal of Animal Ecology* **78**, 219–225.

2.Schluter D. 2001 Ecology and the origin of species. *Trends in Ecology & Evolution* **16**, 372–380.

3.Rolshausen G, Segelbacher G, Hobson KA, Schaefer HM. 2009 Contemporary evolution of reproductive isolation and phenotypic divergence in sympatry along a migratory divide. *Current Biology* **19**, 1–5.

4.Grant PR, Grant BR. 2009 The secondary contact phase of allopatric speciation in Darwin's finches. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 20141–20148.

5.Ober C. 1999 Studies of HLA, fertility and mate choice in a human isolate. *Human Reproduction Update* **5**, 103–107.

6.Vacquier VD. 1998 Evolution of gamete recognition proteins. *Science* **281**, 1995–1998.

7.Primakoff P. 2002 Penetration, adhesion, and fusion in mammalian sperm-egg interaction. *Science* **296**, 2183–2185.

8.Okabe M. 2013 The cell biology of mammalian fertilization. *Development* **140**, 4471–4479.

9.Galindo BE, Vacquier VD, Swanson WJ. 2003 Positive selection in the egg receptor for abalone sperm lysin. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 4639–4643.

10.Kresge N, Vacquier VD, Stout CD. 2001 Abalone lysin: the dissolving and evolving sperm protein. *BioEssays* **23**, 95–103.

11.Panhuis TM, Clark NL, Swanson WJ. 2006 Rapid evolution of reproductive proteins in abalone and *Drosophila*. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **361**, 261–268.

12.Swanson WJ, Vacquier VD. 1998 Concerted evolution in an egg receptor for a rapidly evolving abalone sperm protein. *Science* **281**, 710–712.

13.Metz EC, Gómez-Gutiérrez G, Vacquier VD. 1998 Mitochondrial DNA and bindin gene sequence evolution among allopatric species of the sea urchin genus Arbacia. *Molecular Biology and Evolution* **15**, 185–195.

14.Palumbi SR. 1999 All males are not created equal: Fertility differences depend on gamete recognition polymorphisms in sea urchins. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 12632–12637.

15.Zigler KS, McCartney MA, Levitan DR, Lessios HA. 2005 Sea urchin bindin divergence predicts gamete compatibility. *Evolution* **59**, 2399–2404.

16.Zigler KS, Lessios HA. 2004 Speciation on the coasts of the new world: phylogeography and the evolution of bindin in the sea urchin genus *Lytechinus*. *Evolution* **58**, 1225–1241.

17.Gavrilets S, Waxman D. 2002 Sympatric speciation by sexual conflict. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 10533–10538.

18.Metz EC, Kane RE, Yanagimachi H, Palumbi SR. 1994 Fertilization between closely related sea urchins is blocked by incompatibilities during sperm-egg attachment and early stages of fusion. *The Biological Bulletin* **187**, 23–34.

19.Palumbi SR. 2009 Speciation and the evolution of gamete recognition genes: pattern and process. *Heredity* **102**, 66–76.

20.Corbett-Detig RB, Zhou J, Clark AG, Hartl DL, Ayroles JF. 2014 Genetic incompatibilities are widespread within species. *Nature* **504**, 135–137.

21. Swanson WJ, Vacquier VD. 2002 The rapid evolution of reproductive proteins. *Nature Reviews Genetics* **3**, 137–144.

22. Turner LM, Hoekstra HE. 2008 Causes and consequences of the evolution of reproductive proteins. *The International Journal of Developmental Biology* **52**, 769–780.

23. Hart MW, Sunday JM, Popovic I, Learning KJ, Konrad CM. 2014 Incipient speciation of sea start populations by adaptive gamete recognition coevolution. *Evolution* **68**, 1294–1305.

24. Dorus S, Evans PD, Wyckoff GJ, Choi SS, Lahn BT. 2004 Rate of molecular evolution of the seminal protein gene SEMG2 correlates with levels of female promiscuity. *Nature Genetics* **36**, 1326–1329.

25. Hamm D, Mautz BS, Wolfner MF, Aquadro CF, Swanson WJ. 2007 Evidence of amino acid diversity-enhancing selection within humans and among primates at the candidate sperm-receptor gene PKDREJ. *The American Journal of Human Genetics* **81**, 44–52.

26. Wassarman PM, Litscher ES. 2008 Mammalian fertilization: the eggs multifunctional zona pellucida. *The International Journal of Developmental Biology* **52**, 665–676.

27. Cohen DJ, Busso D, Da Ros V, Ellerman DA, Maldera JA, Goldweic N, Cuasnicu PS. 2008 Participation of cysteine-rich secretory proteins (CRISP) in mammalian sperm-egg interaction. *The International Journal of Developmental Biology* **52**, 737–742.

28. Claw KG, George RD, Swanson WJ. 2014 Detecting coevolution in mammalian sperm-egg fusion proteins. *Molecular Reproduction and Development* **81**, 531–538.

29. Bianchi E, Doe B, Goulding D, Wright GJ. 2014 Juno is the egg Izumo receptor and is essential for mammalian fertilization. *Nature* **508**, 483–487.

30. Wassarman PM. 2014 Sperm protein finds its mate. *Nature* **508**, 466–467.

31. Hardy DM, editor. 2002 *Fertilization*. Academic Press.

32.Diaz FJ, Wigglesworth K, Eppig JJ. 2007 Oocytes determine cumulus cell lineage in mouse ovarian follicles. *Journal of Cell Science* **120**, 1330–1340.

33.Oren-Benaroya R, Orvieto R, Gakamsky A, Pinchasov M, Eisenbach M. 2008 The sperm chemoattractant secreted from human cumulus cells is progesterone. *Human Reproduction* **23**, 2339–2345.

34.Rodger JC, Bedford JM. 1982 Separation of sperm pairs and sperm egg interaction in the opossum, *Didelphis virginiana*. *Journal of Reproduction and Fertilization* **64**, 171–179.

35.Cardullo RA, Thaler CD. 2002 Function of the egg's extracellular matrix. In *Fertilization* (ed DM Hardy), pp. 119–152. New York, NY: Academic Press.

36.Wilburn DB, Swanson WJ. 2015 From molecules to mating: Rapid evolution and biochemical studies of reproductive proteins. *Journal of Proteomics* **135**, 12-25.

37.Wassarman PM, Jovine L, Litscher ES, Qi H, Williams Z. 2004 Egg-sperm interactions at fertilization in mammals. *European Journal of Obstetrics & Gynecology and Reproductive Biology* **115**, S57–S60.

38.Shu L, Suter MJF, Räsänen K. 2015 Evolution of egg coats: linking molecular biology and ecology. *Molecular Ecology* **24**, 4052–4073.

39.Pesch S, Bergmann M. 2006 Structure of mammalian spermatozoa in respect to viability, fertility and cryopreservation. *Micron* **37**, 597–612.

40.Gadella BM, Luna C. 2014 Cell biology and functional dynamics of the mammalian sperm surface. *Theriogenology* **81**, 74–84.

41.Kahn JA, Sunde A, Düring V von, Sørdal T, Remen A, Lippe B, Siegel J, Molne K. 1993 Formation of antisperm antibodies in women treated with fallopian tube sperm perfusion. *Human Reproduction* **8**, 1414–1419.

42.Mazumdar S, Levine AS. 1998 Antisperm antibodies: etiology, pathogenesis, diagnosis, and treatment. *Fertility and Sterility* **70**, 799–810.

43.Robertson SA. 2005 Seminal plasma and male factor signalling in the female reproductive tract. *Cell and Tissue Research* **322**, 43–52.

44.Phillips DM, Mahler S. 1977 Phagocytosis of spermatozoa by the rabbit vagina. *The Anatomical Record: Advances in Integrative Anatomy and Evolutionary Biology* **189**, 61–71.

45.Woelders H, Matthijs A. 2001 Phagocytosis of boar spermatozoa *in vitro* and *in vivo*. *Reproduction Supplement* **58**, 113–127.

46.Suarez SS. 2002 Gamete transport. In *Fertilization* (ed DM Hardy), pp. 3–28. San Diego, CA.

47.Dostál J, Veselský L, Marounek M, Zelezná B, Jonáková V. 1997 Inhibition of bacterial and boar epididymal sperm immunogenicity by boar seminal immunosuppressive component in mice. *Journal of Reproduction and Fertility* **111**, 135–141.

48.Quill TA, Garbers DL. 2002 Sperm motility activation and chemoattraction. In *Fertilization*, pp. 29–55. San Diego, CA: Academic Press.

49.Eisenbach M, Giojalas LC. 2006 Sperm guidance in mammals an unpaved road to the egg. *Nature Reviews Molecular Cell Biology* **7**, 276–285.

50.Bahat A, Tur-Kaspa I, Gakamsky A, Giojalas LC, Breitbart H, Eisenbach M. 2003 Thermotaxis of mammalian sperm cells: A potential navigation mechanism in the female genital tract. *Nature Medicine* **9**, 149–150.

51.Eisenbach M. 1999 Sperm chemotaxis. *Reviews of reproduction* **4**, 56–66.

52.Chang MC. 1951 Fertilizing capacity of spermatozoa deposited into the Fallopian tubes. *Nature* **168**, 697–698.

53.Austin CR. 1951 Observations on the penetration of the sperm in the mammalian egg. *Australian journal of scientific research. Ser. B: Biological sciences* **4**, 581–596.

54.Austin CR. 1952 The 'capacitation' of the mammalian sperm. *Nature* **170**, 326.

55.Jaiswal BS, Eisenbach M. 2002 Capacitation. In *Fertilization* (ed DM Hardy), pp. 57–115. San Diego, CA: Academic Press.

56.Overstreet JW, Gould JE, Katz DF, Hanson FW. 1980 In vitro capacitation of human spermatozoa after passage through a columns of cervical mucus. *Fertility and Sterility* **34**, 604–606.

57.Austin CR. 1985 Sperm maturation in the male and female genital tracts. In *Biology of Fertilization* (eds CB Metz, A Monroy), pp. 121–155. New York, NY: Academic Press.

58.Cohen-Dayag A, Tur-Kaspa I, Dor J, Mashiach S, Eisenbach M. 1995 Sperm capacitation in humans is transient and correlates with chemotactic responsiveness to follicular factors. *Proceedings of the National Academy of Sciences of the United States of America* **92**, 11039–11043.

59.Chang MC. 1957 A detrimental effect of seminal plasma on the fertilizing capacity of sperm. *Nature* **179**, 258–259.

60.Bedford JM. 1970 Sperm capacitation and fertilization in mammals. *Biology of Reproduction. Supplement* **2**, 128–158.

61.Demott RP, Suarez SS. 1992 Hyperactivated sperm progress in the mouse oviduct. *Biology of Reproduction* **46**, 779–785.

62.Stauss CR, Votta TJ, Suarez SS. 1995 Sperm motility hyperactivation facilitates penetration of the hamster zona pellucida. *Biology of Reproduction* **53**, 1280–1285.

63.Yanagimachi R. 1994 Mammalian fertilization. In *The Physiology of Reproduction* (eds E Knobil, J Neill), pp. 189–317. New York, NY: Raven Press.

64.Lamirande E de, Leclerc P, Gagnon C. 1997 Capacitation as a regulatory event that primes spermatozoa for the acrosome reaction and fertilization. *Molecular Human Reproduction* **3**, 175–194.

65.Kopf GS. 2002 Signal transduction mechanisms regulating sperm acrosomal exocytosis. In *Fertilization*, pp. 181–223. San Diego, CA: Academic Press.

66.Gerton GL. 2002 Function of the sperm acrosome. In *Fertilization*, pp. 265–302. San Diego, CA: Academic Press.

67.Wassarman PM, Jovine L, Qi H, Williams Z, Darie C, Litscher ES. 2005 Recent aspects of mammalian fertilization research. *Molecular and Cellular Endocrinology* **234**, 95–103.

68.Liu C, Litscher ES, Mortillo S, Sakai Y, Kinloch RA, Stewart CL, Wassarman PM. 1996 Targeted disruption of the mZP3 gene results in production of eggs lacking a zona pellucida and infertility in female mice. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 5431–5436.

69.Rankin T, Familari M, Lee E, Ginsberg A, Dwyer N, Blanchette-Mackie J, Drago J, Westphal H, Dean J. 1996 Mice homozygous for an insertional mutation in the Zp3 gene lack a zona pellucida and are infertile. *Development* **122**, 2903–2910.

70.Wassarman PM, Jovine L, Litscher ES. 2001 A profile of fertilization in mammals. *Nature Cell Biology* **3**, E59–64.

71.Jin M, Fujiwara E, Kakiuchi Y, Okabe M, Satouh Y, Baba SA, Chiba K, Hirohashi N. 2011 Most fertilizing mouse spermatozoa begin their acrosome reaction before contact with the zona pellucida during in vitro fertilization. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 4892–4896.

72.La Spina FA, Puga Molina LC, Romarowski A, Vitale AM, Falzone TL, Krapf D, Hirohashi N, Buffone MG. 2016 Mouse sperm begin to undergo acrosomal exocytosis in the upper isthmus of the oviduct. *Developmental Biology* **411**, 172–182.

73.Primakoff P, Myles DG. 2002 Gamete fusion in mammals. In *Fertilization*, pp. 303–318. San Diego, CA: Academic Press.

74.Inoue N, Hamada D, Kamikubo H, Hirata K, Kataoka M, Yamamoto M, Ikawa M, Okabe M, Hagihara Y. 2013 Molecular dissection of IZUMO1, a sperm protein essential for sperm-egg fusion. *Development* **140**, 3221–3229.

75.Klinovska K, Sebkova N, Dvorakova-Hortova K. 2014 Sperm-egg fusion: A molecular enigma of mammalian reproduction. *International Journal of Molecular Sciences* **15**, 10652–10668.

76.Inoue N, Ikawa M, Okabe M. 2010 The mechanism of sperm-egg interaction and the involvement of IZUMO1 in fusion. *Asian Journal of Andrology* **13**, 81–87.

77.Inoue N, Ikawa M, Isotani A, Okabe M. 2005 The immunoglobulin superfamily protein Izumo is required for sperm to fuse with eggs. *Nature* **434**, 234–238.

78.Aydin H, Sultana A, Li S, Thavalingam A, Lee JE. 2016 Molecular architecture of the human sperm IZUMO1 and egg JUNO fertilization complex. *Nature* **534**, 562–565.

79.Ohto U, Ishida H, Krayukhina E, Uchiyama S, Inoue N, Shimizu T. 2016 Structure of IZUMO1-JUNO reveals sperm-oocyte recognition during mammalian fertilization. *Nature* **534**, 566–569.

80.Cohen DJ, Maldera JA, Vasen G, Ernesto JI, Munoz MW, Battistone MA, Cuasnicu PS. 2011 Epididymal protein CRISP1 plays different roles during the fertilization process. *Journal of Andrology* **32**, 672–678.

81.Busso D, Cohen DJ, Maldera JA, Dematteis A, Cuasnicu PS. 2007 A Novel Function for CRISP1 in rodent fertilization: Involvement in sperm-zona pellucida interaction. *Biology of Reproduction* **77**, 848–854.

82.Milne TJ, Abbenante G, Tyndall JDA, Halliday J, Lewis RJ. 2003 Isolation and characterization of a cone snail protease with homology to CRISP proteins of the pathogenesis-related protein superfamily. *Journal of Biological Chemistry* **278**, 31105–31110.

83.Yamazaki Y, Morita T. 2004 Structure and function of snake venom cysteine-rich secretory proteins. *Toxicon* **44**, 227–231.

84.Da Ros VG, Maldera JA, Willis WD, Cohen DJ, Goulding EH, Gelman DM, Rubinstein M, Eddy EM, Cuasnicu PS. 2008 Impaired sperm fertilizing ability in mice lacking Cysteine-Rich Secretory Protein 1 (CRISP1). *Developmental Biology* **320**, 12–18.

85.Ellerman DA, Cohen DJ, Da Ros VG, Morgenfeld MM, Busso D, Cuasnicu PS. 2006 Sperm protein 'DE' mediates gamete fusion through an evolutionarily conserved site of the CRISP family. *Developmental Biology* **297**, 228–237.

86.Busso D, Goldweic NM, Hayashi M, Kasahara M, Cuasnicu PS. 2007 Evidence for the involvement of testicular protein CRISP2 in mouse sperm-egg fusion. *Biology of Reproduction* **76**, 701–708.

87.Jamsai D, Reilly A, Smith SJ, Gibbs GM, Baker HWG, McLachlan RI, Kretser DM de, O'Bryan MK. 2008 Polymorphisms in the human cysteine-rich secretory protein 2 (CRISP2) gene in Australian men. *Human Reproduction* **23**, 2151–2159.

88.Hughes J, Ward CJ, Aspinwall R, Butler R, Harris PC. 1999 Identification of a human homologue of the sea urchin receptor for egg jelly: a polycystic kidney disease-like protein. *Human Molecular Genetics* **8**, 543–549.

89.Gallagher AR, Hidaka S, Gretz N, Witzgall R. 2002 Molecular basis of autosomal-dominant polycystic kidney disease. *Cellular and Molecular Life Sciences CMLS* **59**, 682–693.

90.Moy GW, Mendoza LM, Schulz JR, Swanson WJ, Glabe CG, Vacquier VD. 1996 The sea urchin sperm receptor for egg jelly is a modular protein with extensive homology to the human polycystic kidney disease protein, PKD1. *The Journal of Cell Biology* **133**, 809–817.

91.Butscheid Y, Chubanov V, Steger K, Meyer D, Dietrich A, Gudermann T. 2006 Polycystic kidney disease and receptor for egg jelly is a plasma membrane protein of mouse sperm head. *Molecular Reproduction and Development* **73**, 350–360.

92.Sutton KA, Jungnickel MK, Florman HM. 2008 A polycystin-1 controls postcopulatory reproductive selection in mice. *Proceedings of the National Academy of Sciences* **105**, 8661–8666.

93.Primakoff P, Lathrop W, Woolman L, Cowan A, Myles D. 1988 Fully effective contraception in male and female guinea pigs immunized with the sperm protein PH-20. *Nature* **335**, 543–546.

94.Sabeur K, Cherr GN, Yudin AI, Primakoff P, Li MW, Overstreet JW. 1997 The PH-20 protein in human spermatozoa. *Journal of Andrology* **18**, 151–158.

95.Tung KS, Primakoff P, Woolman-Gamer L, Myles DG. 1997 Mechanism of infertility in male guinea pigs immunized with sperm PH-20. *Biology of Reproduction* **56**, 1133–1141.

96.Baba D, Kashiwabara S i, Honda A, Yamagata K, Wu Q, Ikawa M, Okabe M, Baba T. 2002 Mouse sperm lacking cell surface hyaluronidase PH-20 can pass through the layer of cumulus cells and fertilize the egg. *Journal of Biological Chemistry* **277**, 30310–30314.

97.Hunnicutt GR, Primakoff P, Myles DG. 1996 Sperm surface protein PH-20 is bifunctional: one activity is a hyaluronidase and a second, distinct activity is required in secondary sperm-zona binding. *Biology of Reproduction* **55**, 80–86.

98.Cherr GN, Yudin AI, Overstreet JW. 2001 The dual functions of GPI-anchored PH-20: hyaluronidase and intracellular signaling. *Matrix Biology* **20**, 515–525.

99.Vines CA, Li MW, Deng X, Yudin AI, Cherr GN, Overstreet JW. 2001 Identification of a hyaluronic acid (HA) binding domain in the PH-20 protein that may function in cell signaling. *Molecular Reproduction and Development* **60**, 542–552.

100.Gasper J, Swanson WJ. 2006 Molecular population genetics of the gene encoding the human fertilization protein zonadhesin reveals rapid adaptive evolution. *The American Journal of Human Genetics* **79**, 820–830.

101.Hardy DM, Garbers DL. 1995 A sperm membrane protein that binds in a species-specific manner to the egg extracellular matrix is homologous to von Willebrand factor. *Journal of Biological Chemistry* **270**, 26025–26028.

102.Tardif S, Wilson MD, Wagner R, Hunt P. 2010 Zonadhesin is essential for species specificity of sperm adhesion to the egg zona pellucida. *Journal of Biological Chemistry* **285**, 14863–14870.

103.Tardif S, Cormier N. 2011 Role of zonadhesin during sperm-egg interaction: a species-specific acrosomal molecule with multiple functions. *Molecular Human Reproduction* **17**, 661–668.

104.Herlyn H, Zischler H. 2005 Identification of a positively evolving putative binding region with increased variability in posttranslational motifs in zonadhesin MAM domain 2. *Molecular Phylogenetics and Evolution* **37**, 62–72.

105.Meczekalski B. 2009 Oocyte-specific genes: role in fertility and infertility. *Journal of Endocrinological Investigation* **32**, 474–481.

106.Rankin T, Talbot P, Lee E, Dean J. 1999 Abnormal zonae pellucidae in mice lacking ZP1 result in early embryonic loss. *Development* **126**, 3847–3855.

107.Avella MA, Baibakov B, Dean J. 2014 A single domain of the ZP2 zona pellucida protein mediates gamete recognition in mice and humans. *The Journal of Cell Biology* **205**, 801–809.

108.Baibakov B, Boggs NA, Yauger B, Baibakov G, Dean J. 2012 Human sperm bind to the N-terminal domain of ZP2 in humanized zonae pellucidae in transgenic mice. *The Journal of Cell Biology* **197**, 897–905.

109.Rankin TL, O'Brien M, Lee E, Wigglesworth K, Eppig J, Dean J. 2001 Defective zonae pellucidae in Zp2-null mice disrupt folliculogenesis, fertility and development. *Development* **128**, 1119–1126.

110.Florman HM, Wassarman PM. 1985 O-linked oligosaccharides of mouse egg ZP3 account for its sperm receptor activity. *Cell* **41**, 313–324.

111.Chen J, Litscher ES, Wassarman PM. 1998 Inactivation of the mouse sperm receptor, mZP3, by site-directed mutagenesis of individual serine residues located at the combining site for sperm. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 6193–6197.

112.Liu C, Litscher ES, Wassarman PM. 1995 Transgenic mice with reduced numbers of functional sperm receptors on their eggs reproduce normally. *Molecular Biology of the Cell* **6**, 577–585.

113.Williams SA, Xia L, Cummings RD, McEver RP, Stanley P. 2007 Fertilization in mouse does not require terminal galactose or N-acetylglucosamine on the zona pellucida glycans. *Journal of Cell Science* **120**, 1341–1349.

114.Swanson WJ, Yang Z, Wolfner MF, Aquadro CF. 2001 Positive Darwinian selection drives the evolution of several female reproductive proteins in mammals. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 2509–2514.

115.Zhu G-Z, Miller BJ, Boucheix C, Rubinstein E, Liu CC, Hynes RO, Myles DG, Primakoff P. 2002 Residues SFQ (173-175) in the large extracellular loop of CD9 are required for gamete fusion. *Development* **129**, 1995–2002.

116.Diekman AB. 2003 Glycoconjugates in sperm function and gamete interactions: how much sugar does it take to sweet-talk the egg? *Cellular and Molecular Life Sciences* **60**, 298–308.

117.Mengerink KJ, Vacquier VD. 2001 Glycobiology of sperm-egg interactions in deuterostomes. *Glycobiology* **11**, 37R–43R.

118.Tulsiani D, Yoshida-Komiya H, Araki Y. 1997 Mammalian fertilization: a carbohydrate-mediated event. *Biology of Reproduction* **57**, 487–494.

119.Olsson M, Shine R, Madsen T, Gullberg A, Tegelström H. 1996 Sperm selection by females. *Nature* **383**, 585.

120.Tregenza T, Wedell N. 2000 Genetic compatibility, mate choice and patterns of parentage: Invited Review. *Molecular Ecology* **9**, 1013–1027.

121.Birkhead TR, Pizzari T. 2002 Postcopulatory sexual selection. *Nature Reviews Genetics* **3**, 262–273.

122.Birkhead T, Chaline N, Biggins J, Burke T. 2004 Nontransitivity of paternity in a bird. *Evolution* **58**, 416–420.

123.Mays H, Hill G. 2004 Choosing mates: good genes versus genes that are a good fit. *Trends in Ecology & Evolution* **19**, 554–559.

124.Evans J, Marshall D. 2005 Male-by-female interactions influence fertilization success and mediate the benefits of polyandry in the sea urchin *Heliocidaris erythrogramma*. *Evolution* **59**, 106–112.

125.Dziminski MA, Roberts JD, Simmons LW. 2008 Fitness consequences of parental compatibility in the frog *Crinia georgiana*. *Evolution* **62**, 879–886.

126.Kempenaers B, Adriaensen F, Van Noordwijk A, Dhondt A. 1996 Genetic similarity, inbreeding and hatching failure in blue tits: are unhatched eggs infertile? *Proceedings: Biological Sciences* **263**, 179–185.

127.Firman RC, Simmons LW. 2008 Polyandry facilitates postcopulatory inbreeding avoidance in house mice. *Evolution* **62**, 603–611.

128.Bensch S, Andrén H, Hansson B, Pedersen HC, Sand H, Sejberg D, Wabakken P, Åkesson M, Liberg O. 2006 Selection for heterozygosity gives hope to a wild population of inbred wolves. *PLoS ONE* **1**, e72–7.

129.Frasier TR, Gillett RM, Hamilton PK, Brown MW, Kraus SD, White BN. 2013 Postcopulatory selection for dissimilar gametes maintains heterozygosity in the endangered North Atlantic right whale. *Ecology and Evolution* **3**, 3483–3494.

130.Olsson M, Shine R. 1997 Advantages of multiple matings to females: a test of the infertility hypothesis using lizards. *Evolution* **51**, 1684–1688.

131.Tregenza T, Wedell N. 2002 Polyandrous females avoid costs of inbreeding. *Nature* **415**, 71–73.

132.Stockley P. 2003 Female multiple mating behaviour, early reproductive failure and litter size variation in mammals. *Proceedings of the Royal Society of London Series B* **270**, 271–278.

133.Fisher DO, Double MC, Blomberg SP, Jennions MD, Cockburn A. 2006 Post-mating sexual selection increases lifetime fitness of polyandrous females in the wild. *Nature* **444**, 89–92.

134.Mayr E. 1942 *Systematics and the Origin of Species from the Viewpoint of a Zoologist*. New York, NY: Columbia University Press.

135.De Queiroz K. 2007 Species concepts and species delimitation. *Systematic Biology* **56**, 879–886.

136.Rice WR. 1998 Intergenomic conflict, interlocus antagonistic coevolution, and the evolution of reproductive isolation. In *Endless forms: Species and speciation* (eds DJ Howard, SH Berlocher), pp. 261–270. New York, NY.

137.Howard DJ, Reece M, Gregory PG, Chu J, Cain ML. 1998 The evolution of barriers to fertilization between closely related organisms. In *Endless forms: Species and speciation* (eds DJ Howard, SH Berlocher), pp. 279–288. New York, NY: Oxford University Press.

138.Palumbi SR. 1998 Species formation and the evolution of gamete recognition loci. In *Endless forms: Species and speciation* (eds DJ Howard, SH Berlocher), pp. 271–278. New York, NY: Oxford University Press.

139.Hohenlohe PA, Bassham S, Etter PD, Stiffler N, Johnson EA, Cresko WA. 2010 Population Genomics of Parallel Adaptation in Threespine Stickleback using Sequenced RAD Tags. *PLoS Genetics* **6**, e1000862–23.

140.Cammen KM, Schultz TF, Rosel PE, Wells RS, Read AJ. 2015 Genomewide investigation of adaptation to harmful algal blooms in common bottlenose dolphins (*Tursiops truncatus*). *Molecular Ecology* **24**, 4697–4710.

141.Timmann C *et al.* 2017 Genome-wide association study indicates two novel resistance loci for severe malaria. *Nature* **489**, 443–446.

142.Amos W, Driscoll E, Hoffman JI. 2011 Candidate genes versus genome-wide associations: which are better for detecting genetic susceptibility to infectious disease? *Proceedings of the Royal Society of London Series B* **278**, 1183–1188.

Figure 1. Overview of egg structure. The ovulated egg is surrounded by a hyaluronic acid matrix, which contains cumulus cells. The zona pellucida (ZP) separates the cumulus cells from the egg. The perivitelline space is the space between the ZP and the membrane of the egg. Drawn from references 8 & 32.

Figure 2. Overview of major structures, sites, and transport/guiding processes involved in fertilization. Although there is variation across mammals, this is meant to represent generic features of mammalian reproduction. Figure drawn from reference 49 & 51.

Figure 3. Sperm structure and capacitation. Diagram of the sperm head containing the nucleus with the haploid genome, and the acrosome which is a secretory organelle. The acrosome has two membranes, an inner and outer. Capacitation causes multiple physiological changes in the head, acrosome, and tail of the sperm, which is necessary for fertilization. Figure drawn from reference 8.

Figure 4. Changes to spermatozoa during the acrosome reaction. The inner acrosomal membrane is exposed allowing the spermatozoa to bind to and penetrate the zona pellucida, and to bind to the egg plasma membrane. Drawn from references 8 & 37.

Figure 5. Major steps in fertilization. (1) Spermatozoa undergo the acrosome reaction likely prior to reaching the cumulus mass [71,72]; (2) Spermatozoa penetrate the

cumulus cells; (3) spermatozoa binds to the zona pellucida; (4) sperm moves through the zona pellucida into the perivitelline space; (5) sperm binds to the egg plasma membrane; (6) sperm fuses with the egg plasma membrane. Note that binding (step 5) and fusion (step 6) are distinct processes, and studies have shown that sperm can bind to the plasma membrane without fusing with it. Drawn from reference 7.

Figure 1.

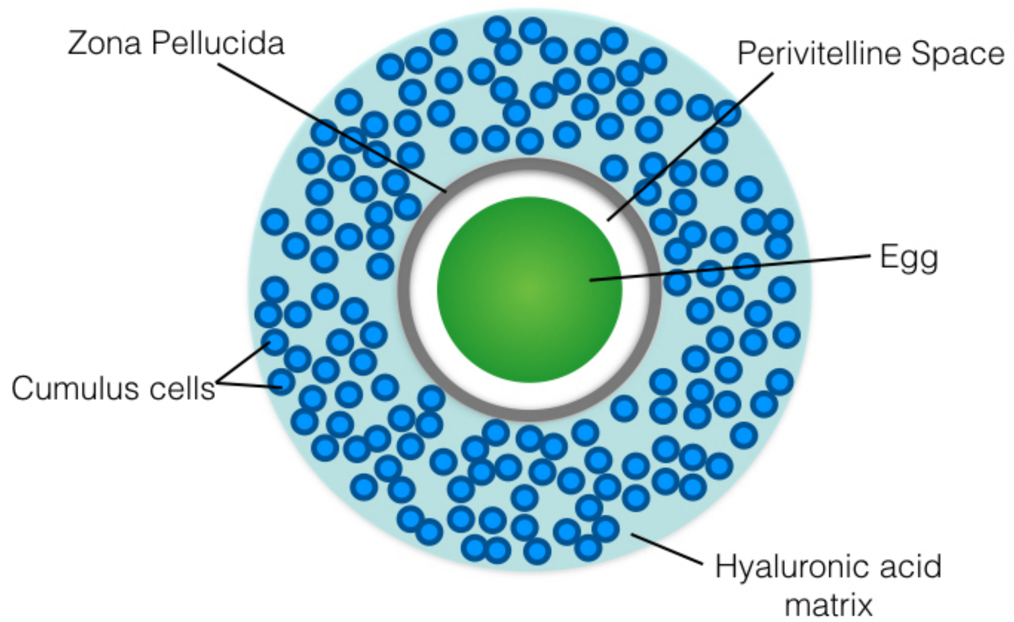


Figure 2.

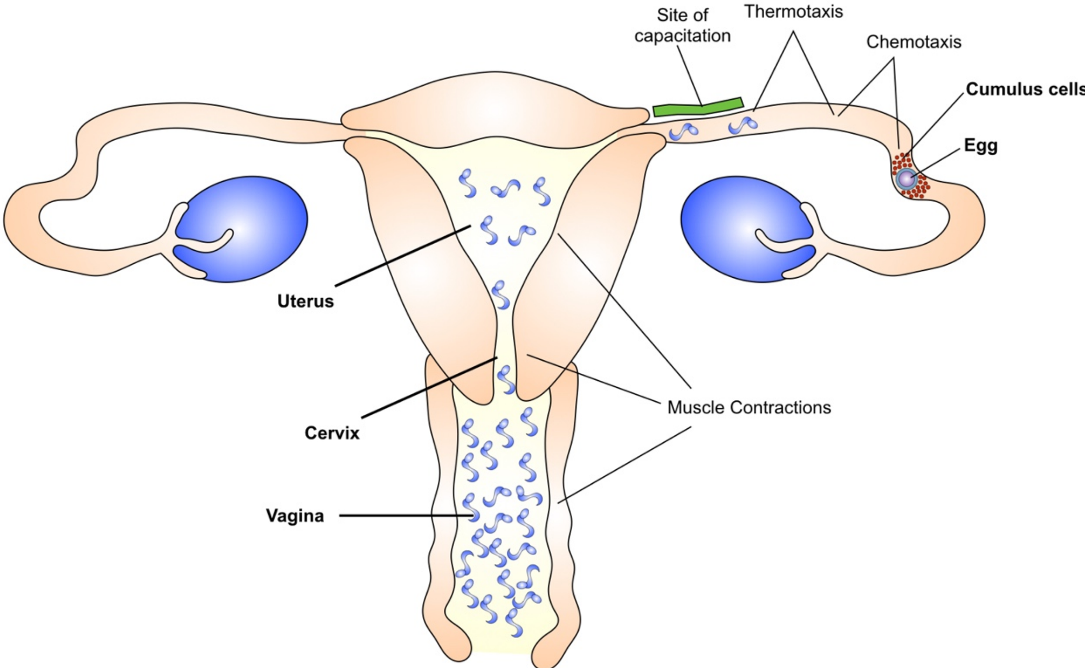


Figure 3.

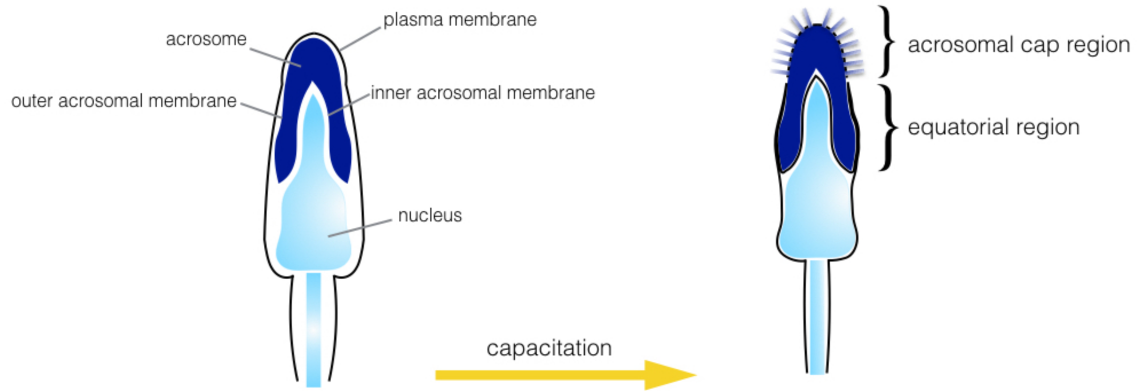


Figure 4.

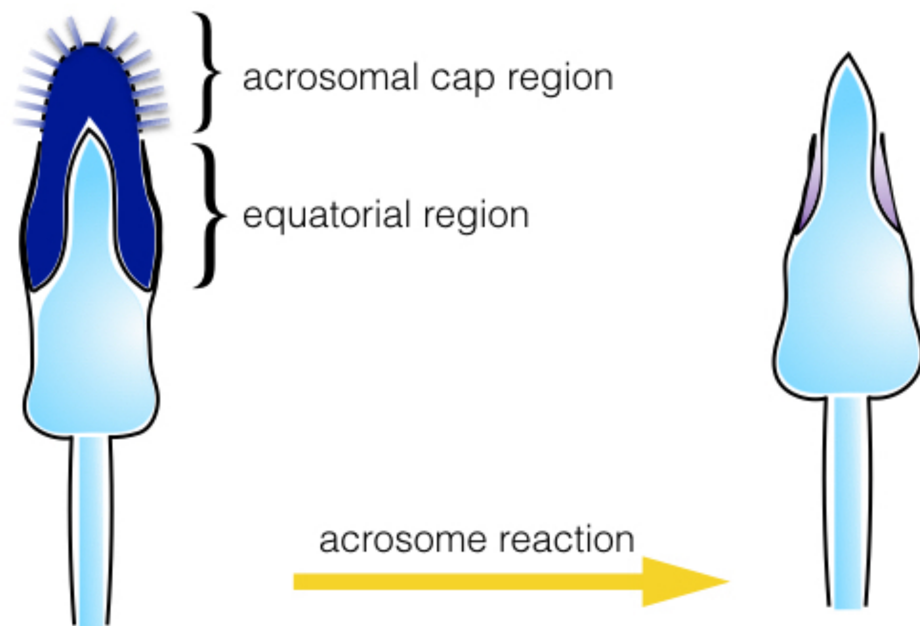
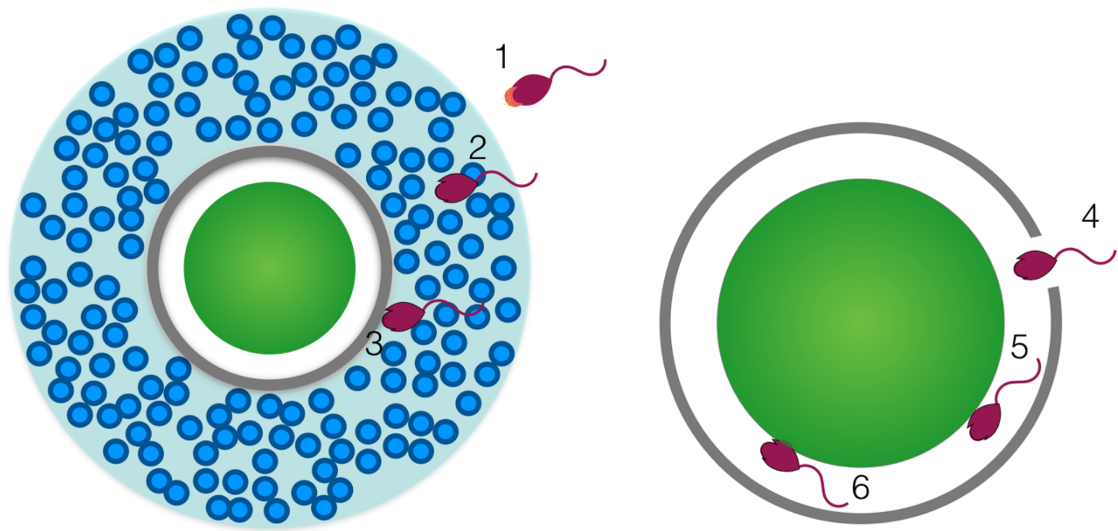


Figure 5.



Chapter 4:

Primer development for gamete compatibility genes in cetaceans

(Formatted for Marine Mammal Science)

ABSTRACT:

It is well known that inbreeding can have negative effects on fitness, and on reproductive performance in particular. However, such information comes primarily from studies correlating some measure of inbreeding with reproductive success, and the specific mechanisms involved are unknown in most species. The North Atlantic right whale (*Eubalaena glacialis*) represents an endangered species where inbreeding appears to be limiting reproductive success and ultimately the recovery potential of the species. Specifically, previous work has detected biased fertilization patterns, suggesting that most matings are unsuccessful, whereas successful fertilizations are biased towards genetically dissimilar gametes, resulting in offspring more heterozygous than expected from this gene pool. To understand the mechanism driving this pattern, and to aid the study of the molecular aspects of fertilization in general, primers were developed to amplify the putative functional sites of nine candidate genes thought to be involved in gamete interactions and fertilization. To maximize their functionality, these primers were designed using sequences from across the diversity of cetaceans (whales, dolphins, and porpoises). Analyses showed that they successfully amplify the desired regions across all cetacean species tested, and will therefore be of value for the study of fertilization across a range of species.

1. Introduction

As ancestry - and therefore genetic similarity - between mates increases, the fitness of offspring tends to decrease, resulting in inbreeding depression. The resulting decrease in heterozygosity can lead to reduced reproductive success and survival of affected individuals (Amos et al. 2001; Frankham 2003, 2005). The overwhelming majority of studies testing for inbreeding depression are based on detecting correlations between heterozygosity of a few microsatellite loci and fitness traits ("Heterozygosity Fitness Correlations" or "HFCs") (e.g., Chapman et al. 2009; Forstmeier et al. 2012). However, more recent data indicate that such an approach has severe limitations, primarily due to the fact that a few microsatellite loci are unlikely to: (a) be representative of genome-wide patterns of heterozygosity, or (b) be in linkage disequilibrium with the specific loci influencing the studied fitness traits (Reed and Frankham 2003; Hoffman et al. 2014).

As a result, there is now a shift towards either genome wide association studies (GWAs), or candidate gene studies (Patnala et al. 2013; Zhu and Zhao 2007). Genome wide association studies correlate allele frequencies from tens to hundreds of thousands of single nucleotide polymorphisms (SNPs) spaced throughout the genome with variation in the desired trait. It is based on the idea that if enough variable markers are spaced throughout the genome, then some of them should be in linkage disequilibrium with the gene(s) influencing the fitness trait, and therefore a correlation will be detected. It is unbiased in regard to the genomic structure and previous knowledge of the trait etiology (Stranger et al. 2011). In contrast, candidate gene studies are focused on analysis of genes hypothesized

to be influential to the desired trait, and therefore require some prior knowledge about gene function (Patnala et al. 2013).

This shift in research approaches can help to directly identify what genes could play a role in the negative effects of inbreeding. Although GWAs are a widely used, they are still a type of HFC test looking for correlations between genes and the trait, and require subsequent candidate gene analyses to identify specific relationships between genotypes and phenotypes. Therefore, for some research questions where suitable candidate genes exist, it may be more efficient to begin with a candidate gene approach.

A recent increase in information regarding which genes influence fertilization, and how they do so, is resulting in new opportunities to understand fertilization at a molecular level (e.g., Bianchi et al. 2014). For example, analyses of these genes may provide information on exactly how genetic variation influences patterns of fertilization and individual reproductive success. Such genes are responsible for maintaining species-specific barriers during fertilization, can bias fertilization towards specific combinations of sperm-egg genotypes, and can scale up during the formation of new species. However, there are only a few examples where the specific proteins and molecular mechanisms that underlie these processes are understood (Turner and Hoekstra 2008; Findlay and Swanson 2015).

Reproductive proteins have been studied extensively in marine invertebrates, such as abalone, sea urchin, and sea stars, due to the large availability and ease of manipulation of gametes from externally fertilizing taxa (Palumbi 1999, 2008; Kamei and Glabe 2003; Harper and Hart 2005; Zigler et al. 2005; Aagaard et al. 2009, 2013; Pujolar and Pogson

2011; Vacquier and Swanson 2011; Hart 2013; Hart et al. 2014). Sea urchins have a fertilization protein called bindin that is present on both the sperm and the egg, which was the first fertilization protein to be identified (Wilburn and Swanson 2015). The main role of bindin is sperm-egg recognition, and the female bindin genotype is a good indicator of the sperm genotype chosen for fertilization. Since bindin is expressed on the surface of each gamete, it is through the interaction of these proteins that determines which combinations will result in successful fertilization, with fertilizations being biased towards more similar alleles. Therefore, there is no "best" bindin genotype, instead, a bindin allele only has an advantage in the presence of a certain female bindin genotype (Palumbi 1999).

At the molecular level, fertilization is a complex act composed of interactions between the sperm and the egg, key steps of which require complimentary interactions between proteins expressed on the surface of each gamete (Primakoff and Myles 2002; Findlay and Swanson 2015; Wilburn and Swanson 2015). In mammals, there are two main steps of fertilization that require the interaction of the surface proteins: when the sperm binds to the zona pellucida (the outer protective layer of the egg), and when the sperm binds to the plasma membrane of the egg (Wassarman et al. 2004). In mammals, while there are multiple proteins that are involved in fertilization, no specific ligand-receptor pair had been identified until recently, when a surface protein on the sperm, “Izumo1”, was identified to interact with the “Juno” receptor on the egg (Bianchi et al. 2014). The interaction of these proteins is necessary for successful fertilization (Bianchi et al. 2014). Several other genes have been implicated in playing a role in fertilization across different species, such as ZP3 and Cd9 on the egg, and CRISP1, CRISP2, PKDREJ, PH-20 and Zonadhesin on the sperm

(Primakoff et al. 1988; Zhu et al. 2002; Wassarman et al. 2004; Busso et al. 2007a; Hamm et al. 2007; Tardif et al. 2010; Nimlamool et al. 2013).

The North Atlantic right whale, *Eubalaena glacialis*, is the world's most endangered large whale, and suffers from a reduced reproductive rate that is three times lower than their known potential (Kraus and Rolland 2007). This reduced reproductive performance is potentially due, at least in part, to a lack of genetically dissimilar mates (Frasier et al. 2013). Right whales have extremely low levels of genetic variation (Frasier et al. 2007), and they show biased fertilization patterns, with offspring being more heterozygous than expected if fertilizations were random based on the genotypes of the parents. This is due to offspring inheriting paternal alleles that differ from maternal alleles more often than expected based on Mendelain inheritance (Frasier et al. 2013). It is thought that females select for sperm from genetically compatible males through cryptic female choice (Kraus and Rolland 2007; Frasier et al. 2013). Similar biased fertilization patterns are found throughout nature (Kempenaers et al. 1996; Olsson et al. 1996; Tregenza and Wedell 2000; Foerster et al. 2003; Bensch et al. 2006; Dziminski et al. 2008; Firman and Simmons 2008).

In mammals, the mechanism(s) that drives these biased fertilization patterns remains unknown, so identifying the genes and mechanisms is considered a high priority (Tregenza and Wedell 2000; Birkhead and Pizzari 2002; Mays Jr and Hill 2004). With the identification of some genes involved in fertilization comes an opportunity to understand the details and mechanisms of gamete compatibility in mammals, and to identify how genetic variation at these genes influence fertilization patterns and fitness. This can also

apply to conservation genetics where reproductive performance is often associated with low genetic diversity (Frankham 2005).

To provide a means to analyze candidate gamete compatibility genes in right whales, and test hypotheses regarding their role in influencing fertilization patterns, it was first necessary to: (a) identify suitable candidate genes; (b) design primers to amplify the putative functional sites of these genes; and (c) sequence right whale mother-father-calf triads to test hypotheses regarding the role of these genes in biasing fertilization patterns. Objectives (a) and (c) are addressed in Chapters 3 and 5, respectively. Here, the work is described that was conducted to design primers for amplification of each candidate gene, and to optimize the laboratory protocols for the analysis of each locus.

2. Methods

Candidate genes were selected based on an extensive literature review for genes involved in gamete interactions (Chapter 3). The interactive sites of the candidate genes were identified either through published genetic sequences that defined a putative functional site, or through the structure of the proteins, focusing on the extracellular section (Cherr et al. 2001; Vines et al. 2001; Zhu et al. 2002; Herlyn and Zischler 2005; Inoue et al. 2005; Ellerman et al. 2006; Busso et al. 2007a; Busso et al. 2007b; Hamm et al. 2007; Jamsai et al. 2008; Han et al. 2010; Chen et al. 2014).

2.1 Primer Design

Primers were designed to amplify the putative regions of each of these candidate genes involved in gamete interaction. To do this, the appropriate sequences were aligned across a range of cetacean (whale, dolphin, and porpoise) species. Primers were then designed in conserved areas adjacent to the regions of interest. Specifically, the species used were the sperm whale - *Physeter macrocephalus*, the orca - *Orcinus orca*, and the minke whale - *Balaenoptera acutorostrata*. For primers that were designed in introns, which are more variable than exons, sites were used that were also conserved in the common bottlenose dolphin - *Tursiops truncatus*, and the Yangtze river dolphin - *Lipotes vexillifer*, to ensure that the priming sites were not variable in cetaceans. All sequences were taken from published data on GenBank (Table 1). Generally, the research for these genes had been done on mice or humans. Therefore, the desired region in human or mice mRNA was BLASTed against a whale genome to confirm the interactive site in the whale genome. To ensure that the correct site was being sequenced in the whales, a published amino acid sequence was compared to the nucleotide sequences on GenBank. The primers were designed following the suggested guidelines: a length of 18-22 BP, a GC content of 40-60%, a GC clamp (but with no more than 3 Gs or Cs in the last 5 bases), avoiding repeats or long runs, and preventing primer dimers (Dieffenbach et al. 1993).

Table 1: Accession numbers from GenBank for mRNA and genome sequences

Gene	mRNA	Cetacean Species and GenBank Accession #
Cd9	*Homo sapiens* (M38690.1)	*Physeter macrocephalus* (NW_006713091.1) *Balaenoptera acutorostrata* (NW_006725621.1) *Orcinus orca* (NW_004438521.1) *Tursiops truncatus* (NW_017842545.1) *Lipotes vexillifer* (NW_006797390.1)
CRISP1	*Homo sapiens* (NM_001131.2) *Sus scrofa* (NM_001128434.2)	*Orcinus orca* (NW_004438437.1) *Balaenoptera acutorostrata* (NW_006732678.1) *Tursiops truncatus* (NW_004210126.1) *Lipotes vexillifer* (NW_006778796.1) *Physeter macrocephalus* (NW_006712912.1)
CRISP2	*Homo sapiens* (NM_001142407.2)	*Orcinus orca* (NW_004438437.1) *Physeter macrocephalus* (NW_006712912.1) *Balaenoptera acutorostrata* (NW_006732678.1)
Izumo1	*Mus musculus* (NM_001018013) *Bos taurus* (XM_015458228)	*Orcinus orca* (NW_004438455.1) *Balaenoptera acutorostrata* (NW_006725499.1) *Physeter macrocephalus* (NW_006714298.1)
Juno	*Homo sapiens* (NM_001199206)	*Orcinus orca* (NW_004438427.1) *Balaenoptera acutorostrata* (NW_006730123.1) *Physeter macrocephalus* (NW_006714503.1) *Tursiops truncatus* (NW_017844693.1) *Lipotes vexillifer* (NW_006775074.1)
PH-20	*Homo sapiens* (S67798)	*Orcinus orca* (NW_004438427.1) *Physeter macrocephalus* (NW_006712825.1) *Balaenoptera acutorostrata* (NW_006726354.1)
PKDREJ	*Homo sapiens* (NM_006071.1)	*Orcinus orca* (NW_004438525.1) *Physeter macrocephalus* (NW_006724313.1) *Balaenoptera acutorostrata* (NW_006729790.1)

Zonadhesin	*Homo sapiens* (NM_003386)	*Orcinus orca* (NW_004438442.1) *Physeter macrocephalus* (NW_006713905.1) *Balaenoptera acutorostrata* (NW_006728570.1) *Lipotes vexillifer* (NW_006796716.1) *Tursiops truncatus* (NW_004200502.1)
ZP3	*Mus musculus* (M20026.1) *Sus scrofa* (NM_213893)	*Balaenoptera acutorostrata* (NW_006728570.1) *Physeter macrocephalus* (NW_006713114.1) *Tursiops truncatus* (NW_004202001.1) *Lipotes vexillifer* (NW_006796716.1)

2.2 Optimizing PCR conditions

To optimize the amplification conditions, primers were initially tested using a range of annealing temperatures in fin whales (*Balaenoptera physalus*). Initially, annealing temperatures of 50°C, 55°C, and 60°C were tested, of 3 individual samples at each temperature, for a total of 9 reactions. Table 2 shows the amount of cycles and annealing temperature for each primer. For some loci, such as Izumo1 and ZP3, due to difficulties with amplification, multiple primer sets were ordered, and different combinations of forward and reverse primers were tested. Each 20 µl PCR cocktail contained 2 µl of DNA (at a concentration of 5 ng/µl for total of 10 ng), as well as 1x PCR buffer (Promega), 0.2 mM each dNTP (Invitrogen), 0.4 mg/mL bovine serum albumin (Ambion), 1.5 mM $MgCl_2$ (Promega), 0.3 µM of each the forward and reverse primer, and 0.05 u/µl *Taq* polymerase (Promega). The PCR conditions for each locus are described in Table 2.

In all cases, PCR products were size-separated and visualized by combining 5µl of PCR product with 3µl of Orange G dye, and loading this into 1.5% agarose gels, stained with GelRed (Biotium). The conditions that yielded the brightest and cleanest bands were

chosen as the optimal conditions. If necessary, based on the results of the gel, either an optimal annealing temperature would be chosen, or the reaction would be further manipulated through changing the annealing temperature, and/or increasing the number of cycles.

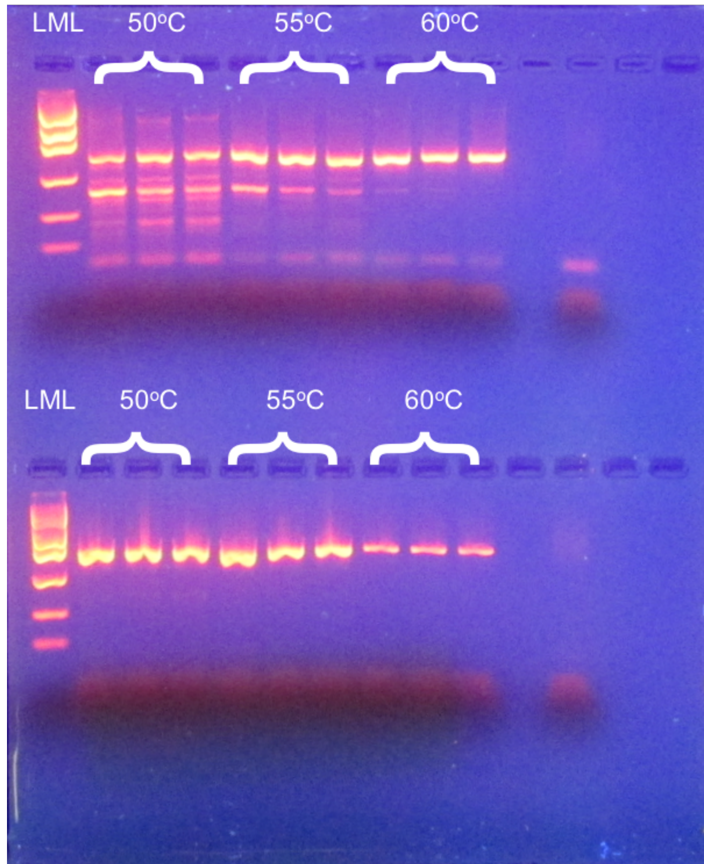


Figure 1: Testing annealing temperatures in fin whales (*Balaenoptera physalus*) to determine optimal PCR conditions. For locus PKDREJ-S1a (top row) an annealing temperature of 60°C was chosen, and for locus PKDREJ-S1b (bottom row) an annealing temperature of 55°C was chosen.

Using the optimized PCR conditions developed from testing the fin whale samples, these same conditions were used for testing the additional cetacean species (Table 3). To ensure that the primers would amplify across a range of cetaceans, all primers were tested across 6 different cetacean species, with 3 individuals of each species: fin whale - *Balaenoptera physalus*, North Atlantic right whale - *Eubalaena glacialis*, beluga - *Delphinapterus leucas*, pilot whale - *Globicephala melas*, sperm whale - *Physeter macrocephalus*, and gray whale - *Eschrichtius robustus*. As before, PCR products were size-separated and visualized by combining 5μl of PCR product with 3μl of Orange G dye, and loading this into 1.5% agarose gels stained with GelRed. Low mass ladder (Invitrogen) was also used as a size standard to ensure the correct region was amplified by confirming the length was correct. The DNA was then visualized with a UV light.

Table 2: Primer sets and PCR conditions

Gene	Primer Name	Forward Primer	Reverse Primer	PCR Cycling Conditions	Locus Length (bp)
Cd9	Cet-Cd9a	GTT CTT TGG CTT CCT CTT GGT G	CAA ATC CCA CGG GGC TTC TG	5 min @ 94°C 30 sec @ 94°C 1 min @ 60°C 30 cycles 1 min @ 72°C 45 min @ 60°C	630
	Cet-Cd9b	GCA GAA ACT CAT GTA CAG AGA GG	TGT CCT TTT TGG GGC AGA TG	5 min @ 94°C 30 sec @ 94°C 1 min @ 55°C 30 cycles 1 min @ 72°C 45 min @ 60°C	700
	Cet-Cd9c	CAT CTG CCC CAA AAA GGA C	AGT GCC TGG CCA GAG CAA G	5 min @ 94°C 30 sec @ 94°C 1 min @ 55°C 30 cycles 1 min @ 72°C 45 min @ 60°C	800

CRISP1	Cet-CRISP1a	GGG CCA TTC TTA CCT TAT TGG	CTT GGA ATT GTT GCT CAC ACC	5 min @ 94°C 30 sec @ 94°C 1 min @ 55°C 30 cycles 1 min @ 72°C 45 min @ 60°C	600
	Cet-CRISP1b	CTA AGG CTA GAA ATC TTA CAT GTC	CTT CAC AGT TAT TTG GAC AGT C	5 min @ 94°C 30 sec @ 94°C 1 min @ 50°C 30 cycles 1 min @ 72°C 45 min @ 60°C	260
CRISP2	Cet-CRISP2	CTC CTT AGT AAC TGT GAT TCC	TCT TGG AGT AGG AAT GAC AAT G	5 min @ 94°C 30 sec @ 94°C 1 min @ 50°C 40 cycles 1 min @ 72°C 45 min @ 60°C	530
Izumo1	Izumo1-e/f	GCC AAG GAA ACC TCA GGG	TTG TGA TGC TTG GCC TCC AGG	5 min @ 94°C 30 sec @ 94°C 1 min @ 55°C 30 cycles 1 min @ 72°C 45 min @ 60°C	600
Juno	Cet-Juno1	TTC CTA GGG GAC TCA CAG TCT G	TGC CAG TTG GAT TTG CAG GTG	5 min @ 94°C 30 sec @ 94°C 1 min @ 55°C 30 cycles 1 min @ 72°C 45 min @ 60°C	900
	Cet-Juno2	TAC ACC TGC AAA TCC AAC TGG C	CAG GCT TGA ACC ACT TCT GCA G	5 min @ 94°C 30 sec @ 94°C 1 min @ 55°C 30 cycles 1 min @ 72°C 45 min @ 60°C	380
PH-20	PH-20-c/d	AAC GTG GGC TGG CTC TCA TTG	AAG AGG GCA GCT TGT GGA G	5 min @ 94°C 30 sec @ 94°C 1 min @ 55°C 30 cycles 1 min @ 72°C 45 min @ 60°C	420
PKDREJ	Cet-PKDREJ-S1a	TCG TCT TAG AGG CTG TCA ACC	CGT AGT TTC TTG GGT CTG TGG	5 min @ 94°C 30 sec @ 94°C 1 min @ 60°C 30 cycles 1 min @ 72°C 45 min @ 60°C	600

	Cet-PKDREJ-S1b	GGA AAT ATG TCC TCT GAT CCA G	GGA TCA ATT TTG CAT TCT GTG G	5 min @ 94°C 30 sec @ 94°C 1 min @ 55°C 30 cycles 1 min @ 72°C 45 min @ 60°C	660
	Cet-PKDREJ-S2a	CAC GGT GAG GAT TTG GCA AGC	TGA ACA ACA CGG CCA CTT CCG	5 min @ 94°C 30 sec @ 94°C 1 min @ 55°C 30 cycles 1 min @ 72°C 45 min @ 60°C	740
	Cet-PKDREJ-S2b	GGA TTC CAA TGA CAG GAA CCG	AGA TCC ACG GGG TTA GGG AC	5 min @ 94°C 30 sec @ 94°C 1 min @ 55°C 30 cycles 1 min @ 72°C 45 min @ 60°C	740
Zonadhesin	Cet-Zonad	TGA GAG CAG CTC TGA GAA TGG G	AGG CCT TGG GCT CAA TGA AGC	5 min @ 94°C 30 sec @ 94°C 1 min @ 55°C 30 cycles 1 min @ 72°C 45 min @ 60°C	540
ZP3	ZP3-c/d	TGC CAC CTG AAG GTC ACT CC	ACT GTG ACA TCT GCT TCT TCT G	5 min @ 94°C 30 sec @ 94°C 1 min @ 65°C 35 cycles 1 min @ 72°C 45 min @ 60°C	680

3. Results and Discussion

Each species successfully amplified at the correct region for each primer pair (Figure 2). Since these primers amplified across all of the tested species, it is likely that they will work for most, or potentially all, cetacean species.

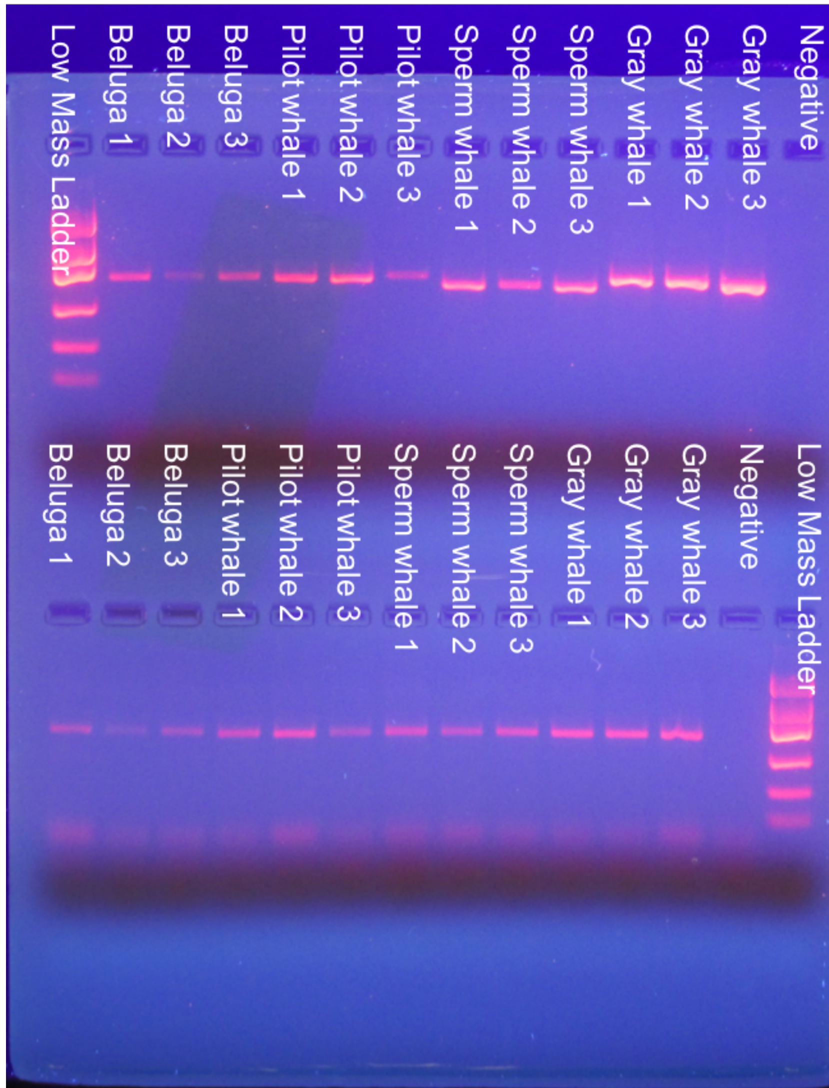


Figure 2: Gel Image of Loci Cd9a (row 1) and PKDREJ-S1a (row 2)

In addition to being helpful for understanding how genetic factors influence reproductive success in right whales, these primers may also prove useful for many other species. For example, due to intensive whaling practices, many cetacean species are endangered, and also suffer from low genetic variability (Clapham *et al*. 1999; Perry *et al*. 1999; Roman *et al*. 2013). Some species are similar to the North Atlantic right whale, in that they are showing little signs of recovery. This includes Bowhead Whales (*Balaena*

mysticetus) of the Okhotsk Sea and various eastern Arctic populations, western Pacific Gray Whales (*Eschrichtius robustus*), and Blue Whale (*Balaenoptera musculus*) populations. Comparing sequences from these genes involved in fertilization in such species, and comparing them to species that are recovering well (such as Humpback whales (*Megaptera novaeangliae*) and the Southern right whale (*Eubalaena australis*)), may help improve our understanding of exactly how genetic characteristics can influence reproductive performance and recovery potential in endangered cetacean species.

Additionally, genes involved in gamete compatibility also likely play a key role in the evolution of reproductive barriers during the speciation process (Gavrilets & Waxman 2002; Palumbi 2008). It is assumed that time-based acquisition of genetic compatibilities drives reproductive isolation, and that the raw material needed to cause this reproductive isolation (gamete compatibility alleles with different affinities) is already segregating within species. Reproductive isolation can be attained though divergence in frequencies of these compatibility alleles (Corbett-Detig *et al.* 2013; Hedges *et al.* 2015). Since gamete compatibility genes are responsible for maintaining species-specific barriers during fertilization, these interactions result in non-random fertilization patterns, and can have large impacts on patterns of individual reproductive success within populations, and on the development of reproductive barriers during speciation (Turner & Hoekstra 2008). Cetaceans represent ideal case studies for testing such hypotheses because there are many examples of recent speciation events and periods of adaptive radiation (such as the finless porpoise (*Neophocaena*) in the strait of Taiwan (Wang *et al*. 2008), and the evolution of killer whale ecotypes (Morin et al. 2015). Therefore, the analyses of these genes could

prove useful in understanding the molecular aspects of the evolution of reproductive barriers and speciation within whales.

Due to their potential ability to improve our understanding of the links between genetic characteristics and reproductive performance, and the subsequent implications on conservation biology and speciation, the analysis of these genes should be an important new tool for researchers interested in understanding the molecular aspects of reproduction.

References

AAGAARD, J. E., S. A. SPRINGER, S. D. SOELBERG, AND W. J. SWANSON. 2013. Duplicate Abalone Egg Coat Proteins Bind Sperm Lysin Similarly, but Evolve Oppositely, Consistent with Molecular Mimicry at Fertilization. PLoS Genetics 9:e1003287–11.

AAGAARD, J. E., V. D. VACQUIER, M. J. MACCOSS, AND W. J. SWANSON. 2009. ZP Domain Proteins in the Abalone Egg Coat Include a Paralog of VERL under Positive Selection That Binds Lysin and 18-kDa Sperm Proteins. Molecular Biology and Evolution 27:193–203.

AMOS, W. ET AL. 2001. The influence of parental relatedness on reproductive success. Proceedings of the Royal Society B: Biological Sciences 268:2021–2027.

BENSCH, S. ET AL. 2006. Selection for heterozygosity gives hope to a wild population of inbred wolves. PLOS ONE 1:e72.

BIANCHI, E., B. DOE, D. GOULDING, AND G. J. WRIGHT. 2014. Juno is the egg Izumo receptor and is essential for mammalian fertilization. Nature 508:483–487.

BIRKHEAD, T. R., AND T. PIZZARI. 2002. Postcopulatory Sexual Selection. Nature Reviews Genetics 3:262–273.

BUSSO, D., D. J. COHEN, J. A. MALDERA, A. DEMATTEIS, AND P. S. CUASNICU. 2007a. A Novel Function for CRISP1 in Rodent Fertilization: Involvement in Sperm-Zona Pellucida Interaction. Biology of Reproduction 77:848–854.

BUSSO, D., N. M. GOLDWEIC, M. HAYASHI, M. KASAHARA, AND P. S. CUASNICU. 2007b. Evidence for the Involvement of Testicular Protein CRISP2 in Mouse Sperm-Egg Fusion. Biology of Reproduction 76:701–708.

CHAPMAN, J. R., S. NAKAGAWA, D. W. COLTMAN, J. SLATE, AND B. C. SHELDON. 2009. A quantitative review of heterozygosity-fitness correlations in animal populations. Molecular Ecology 18:2746–2765.

CHEN, C. ET AL. 2014. Structural basis for molecular recognition of folic acid by folate receptors. Nature 500:486–489.

CHERR, G. N., A. I. YUDIN, AND J. W. OVERSTREET. 2001. The dual functions of GPI-anchored PH-20: hyaluronidase and intracellular signaling. Matrix Biology 20:515–525.

CLAPHAM, P. J., S. B. YOUNG, AND L. R. BROWNELL.1999. Baleen whales: conservation issues and the status of the most endangered populations. Mammal Review 29: 37–62.

CORBETT-DETIG, R.B., J. ZHOU, A. G. CLARK, D. L. HARTL, AND J. F. AYROLES. 2013. Genetic incompatibilities are widespread within species. Nature 504:135–137.

DIEFFENBACH, C. W, T. M. LOWE, T. AND G. S. DVEKSLER. 1993. General concepts for PCR primer design. PCR Methods and Applications 3:S30-37.

DZIMINSKI, M. A., J. D. ROBERTS, AND L. W. SIMMONS. 2008. Fitness consequences of parental compatibility in the frog Crinia georgiana. Evolution 62:879–886.

ELLERMAN, D. A., D. J. COHEN, V. G. DA ROS, M. M. MORGENFELD, D. BUSSO, AND P. S. CUASNICU. 2006. Sperm protein “DE” mediates gamete fusion through an evolutionarily conserved site of the CRISP family. Developmental Biology 297:228–237.

FINDLAY, G. D., AND W. J. SWANSON. 2015. Proteomics enhances evolutionary and functional analysis of reproductive proteins. BioEssays 32:26–36.

FIRMAN, R. C., AND L. W. SIMMONS. 2008. Polyandry facilitates postcopulatory inbreeding avoidance in house mice. Evolution 62:603–611.

FOERSTER, K., K. DELHEY, A. JOHNSEN, AND J. T. LIFJELD. 2003. Females increase offspring heterozygosity and fitness through extra-pair matings. Nature 425:714–717.

FORSTMEIER, W., H. SCHIELZETH, J. C. MUELLER, H. ELLEGREN, AND B. KEMPENAERS. 2012. Heterozygosity-fitness correlations in zebra finches: microsatellite markers can be better than their reputation. Molecular Ecology 21:3237–3249.

FRANKHAM, R. 2003. Genetics and conservation biology. Comptes Rendus Biologies 326:22–29.

FRANKHAM, R. 2005. Genetics and extinction. Biological Conservation 126:131–140.

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. ET AL. 2007. Patterns of male reproductive success in a highly promiscuous whale species: the endangered North Atlantic right whale. Molecular Ecology 16:5277–5293.

GAVRILETS, S., & D. WAXMAN, 2002. Sympatric speciation by sexual conflict. Proceedings of the National Academy of Sciences 99:10533–10538.

HAMM, D., B. S. MAUTZ, M. F. WOLFNER, C. F. AQUADRO, AND W. J. SWANSON. 2007. Evidence of Amino Acid Diversity Enhancing Selection within Humans and among Primates at the Candidate Sperm-Receptor Gene PKDREJ. The American Journal of Human Genetics 81:44–52.

HAN, L. ET AL. 2010. Insights into Egg Coat Assembly and Egg-Sperm Interaction from the X-Ray Structure of Full-Length ZP3. Cell 143:404–415.

HARPER, F. M., AND M. W. HART. 2005. Gamete compatibility and sperm competition affect paternity and hybridization between sympatric Asterias sea stars. The Biological bulletin 209:113–126.

HART, M. W. 2013. Structure and evolution of the sea star egg receptor for sperm bindin. Molecular Ecology 22:2143–2156.

HART, M. W., J. M. SUNDAY, I. POPOVIC, K. J. LEARNING, AND C. M. KONRAD. 2014. Incipient speciation of sea star populations by adaptive gamete recognition coevolution. Evolution 68:1294–1305.

HEDGES, S. B., J. MARIN, M. SULESKI, M. PAYMER, AND S. KUMAR. 2015. Tree of life reveals clock-like speciation and diversification. Molecular Biology and Evolution 32:835–845.

HERLYN, H., AND H. ZISCHLER. 2005. Identification of a positively evolving putative binding region with increased variability in posttranslational motifs in zonadhesin MAM domain 2. Molecular Phylogenetics and Evolution 37:62–72.

HOFFMAN, J. I. ET AL. 2014. High-throughput sequencing reveals inbreeding depression in a natural population. Proceedings of the National Academy of Sciences 111:3775–3780.

INOUE, N., M. IKAWA, A. ISOTANI, AND M. OKABE. 2005. The immunoglobulin superfamily protein Izumo is required for sperm to fuse with eggs. Nature 434:234–238.

JAMSAI, D. ET AL. 2008. Polymorphisms in the human cysteine-rich secretory protein 2 (CRISP2) gene in Australian men. Human Reproduction 23:2151–2159.

KAMEI, N., AND C. G. GLABE. 2003. The species-specific egg receptor for sea urchin sperm adhesion is EBR1, a novel ADAMTS protein. Genes & development 17:2502–2507.

KEMPENAERS, B., F. ADRIAENSEN, A. J. VAN NOORDWIJK, AND A. A. DHONDT. 1996. Genetic Similarity, Inbreeding and Hatching Failure in Blue Tits: Are Unhatched Eggs Infertile? Proceedings of the Royal Society B: Biological Sciences 263:179–185.

KRAUS, S. D., AND R. M. ROLLAND (EDS.). 2007. The Urban Whale. Harvard University Press.

MAYS JR, H. L., AND G. E. HILL. 2004. Choosing mates: good genes versus genes that are a good fit. Trends in Ecology & Evolution 19:554–559.

MORIN, P.A., K. M. PARSONS, F. I. ARHCER, M. V. ÁVILA-ARCOS, L. G. BARRETT-LENNARD, L. D. ROSA, S. DUCHÊNE, J. W. DURBAN, G. M. ELLIS, S. H. FERGUSON, J. K. FORD, M. J. FORD, C. GARILAO, M. T. P. GILBERT, K. KASCHNER, C. O. MATKIN, S. D. PETERSEN, K. M. ROBERTSON, I. N. VISSER, P. R. WADE, S Y. W. HO, AND A. D. FOOTE. 2015. Geographic and temporal dynamics of a global radiation and diversification in the killer whale. Molecular Ecology 24:3964-3979.

NIMLAMOOL, W., B. S. BEAN, AND L. J. LOWE-KRENTZ. 2013. Human sperm CRISP2 is released from the acrosome during the acrosome reaction and re-associates at the equatorial segment. Molecular Reproduction and Development 80:488–502.

OLSSON, M., R. SHINE, T. MADSEN, A. GULLBERG, AND H. TEGELSTROM. 1996. Sperm selection by females. Nature 383:585.

PALUMBI, S. R. 1999. All males are not created equal: fertility differences depend on gamete recognition polymorphisms in sea urchins. Proceedings of the National Academy of Sciences 96:12632–12637.

PALUMBI, S. R. 2008. Speciation and the evolution of gamete recognition genes: pattern and process. Heredity 102:66–76.

PATNALA, R., J. CLEMENTS, AND J. BATRA. 2013. Candidate gene association studies: a comprehensive guide to useful in silico tools. BMC genetics 14:39.

PERRY, S. L., D. P. DEMASTER, & G. K. SILBER. 1999. The Great Whales: History and Status of Six Species Listed as Endangered Under the U.S. Endangered Species Act of 1973. Marine Fisheries Review 61:1–74.

PRIMAKOFF, P., AND D. G. MYLES. 2002. Penetration, adhesion, and fusion in mammalian sperm-egg interaction. Science 296:2183–2185.

PRIMAKOFF, P., W. LATHROP, L. WOOLMAN, A. COWAN, AND D. MYLES. 1988. Fully effective contraception in male and female guinea pigs immunized with the sperm protein PH-20. Nature 335:543–546.

PUJOLAR, J. M., AND G. H. POGSON. 2011. Positive Darwinian selection in gamete recognition proteins of Strongylocentrotus sea urchins. Molecular Ecology 20:4968–4982.

REED, D. H., AND R. FRANKHAM. 2003. Correlation between fitness and genetic diversity. Conservation biology 17:230–237.

ROMAN, J., I. ALTMAN, M. M. DUNPHY-DALY, C. CAMPBELL, M. JASNY, AND A. J. READ. 2013. The Marine Mammal Protection Act at 40: status, recovery, and future of U.S. marine mammals. Annals of the New York Academy of Sciences 1286:29–49.

STRANGER, B. E., E. A. STAHL, AND T. RAJ. 2011. Progress and promise of genome-wide association studies for human complex trait genetics. Genetics 187:367–383.

TARDIF, S., M. D. WILSON, R. WAGNER, AND P. HUNT. 2010. Zonadhesin is essential for species specificity of sperm adhesion to the egg zona pellucida. Journal of Biological Chemistry 285:14863–14870.

TREGENZA, T., AND N. WEDELL. 2000. Genetic compatibility, mate choice and patterns of parentage: invited review. Molecular Ecology 9:1013–1027.

TURNER, L. M., AND H. E. HOEKSTRA. 2008. Causes and consequences of the evolution of reproductive proteins. The International Journal of Developmental Biology 52:769–780.

VACQUIER, V. D., AND W. J. SWANSON. 2011. Selection in the Rapid Evolution of Gamete Recognition Proteins in Marine Invertebrates. Cold Spring Harbor Perspectives in Biology 3:a002931–a002931.

VINES, C. A., M. W. LI, X. DENG, A. I. YUDIN, G. N. CHERR, AND J. W. OVERSTREET. 2001. Identification of a hyaluronic acid (HA) binding domain in the PH-20 protein that may function in cell signaling. Molecular Reproduction and Development 60:542–552.

WANG, J. Y., T. R. FRASIER, S. C. YANG, AND B. N. WHITE. 2008. Detecting recent speciation events: the case of the finless porpoise (genus *Neophocaena*). Heredity 101:145–155.

WASSARMAN, P. M., L. JOVINE, E. S. LITSCHER, H. QI, AND Z. WILLIAMS. 2004. Egg sperm interactions at fertilization in mammals. European Journal of Obstetrics & Gynecology and Reproductive Biology 115:S57–S60.

Wilburn, D. B., and W. J. Swanson. 2015. From molecules to mating: Rapid evolution and biochemical studies of reproductive proteins. Journal of proteomics.

Zhu, G.-Z. et al. 2002. Residues SFQ (173-175) in the large extracellular loop of CD9 are required for gamete fusion. Development 129:1995–2002.

Zhu, M., and S. Zhao. 2007. Candidate gene identification approach: progress and challenges. International journal of biological sciences 3:420–427.

Zigler, K. S., M. A. McCartney, D. R. Levitan, and H. A. Lessios. 2005. Sea urchin bindin divergence predicts gamete compatibility. Evolution 59:2399–2404.

Chapter 5:

Gamete compatibility and reproductive success in the North Atlantic right whale

ABSTRACT:

The North Atlantic right whale is considered the world's most endangered baleen whale. Despite international protection since 1935, the species has shown little sign of recovery. This is partly due to a reduced reproductive rate that appears to be due, at least in part, to low genetic diversity. Specifically, previous studies found that fertilizations were biased towards genetically dissimilar gametes, which results in offspring that are born more heterozygous than expected from this gene pool. However, these data were based on non-coding microsatellite loci, and therefore the underlying mechanism remains unknown. Here, 9 genes were analyzed that are putatively involved in gamete compatibility, in 69 mother-father-calf triads, to test for biased fertilization patterns and therefore to aid in the identification of how genetic characteristics are influencing reproductive performance. Only two loci had variable sites, or single nucleotide polymorphisms (SNPs), one at Cd9a, and two at PKDREJ-S1b. Some genotypes are over/under-represented among mating pairs, suggesting potential mate-choice effects based on these loci. When testing for a bias in inheritance patterns, only locus PKDREJ-S1b showed a strong signal for non-Mendelian patterns, however this pattern was in the opposite direction than expected, with calves being more homozygous than expected. Thus, these loci may be influencing patterns of reproductive success, but in ways different than those detected based on the previous microsatellite data.

1. Introduction

The North Atlantic right whale (*Eubalaena glacialis*) is considered the world's most endangered baleen whale (Perry *et al.* 1999). Entanglement in fishing gear and ship strikes are the leading causes of death (Kraus 1990; Clapham *et al.* 1999; Perry *et al.* 1999; Kraus *et al.* 2005). Despite international protection since 1935, the species has shown little signs of recovery. They have been increasing at an average annual rate of ~2%; however, there are wide fluctuations in mortality and birth rates between years, leading to periods of presumed decline (Fujiwara & Caswell 2001; Kraus *et al.* 2005). This is in stark contrast to populations of South Atlantic right whales (*Eubalaena australis*), which were also extensively hunted, but are now increasing at rates of ~7% (Best *et al.* 2001). The current population size of the North Atlantic right whale is estimated at approximately 500 individuals (Kraus *et al.* 2016).

The compromised growth rate of the North Atlantic right whale is not solely attributed to anthropogenic mortalities (Kraus *et al.* 2005). Instead, they also suffer from a reduced reproductive rate that is three times lower than their known potential (Frasier *et al.* 2007b). Based on the number of females in the current population, approximately 30-35 calves should be born per year, but the birth rate is on average around 13 calves per year (Frasier *et al.* 2007a). However, yearly counts can vary from as low as 1 calf up to 31 (with just 4 calves being born this past calving season).

The right whale mating system is one of the most extreme examples of sperm competition seen in mammals (Frasier *et al.* 2007a). In species with sperm competition, males tend to have larger than expected testes, given their body size (Ginsberg & Huck

1989; Preston *et al.* 2003), and right whales have the highest testes-size to body-size ratio in mammals, and one of the highest penis-size to body-size ratio in baleen whales (Frasier *et al.* 2007a). Their mating groups consist of one female and multiple males (Mate *et al.* 2005). These groups are called surface active groups, or SAGs. The female initiates these by making a specific call that attracts males (Parks 2003). When males arrive, she rolls onto her back, making copulation difficult, while males try to move into the closest spot next to the female. When the female rolls back over to breathe, the closest male is able to mate with her. On average, one female will have mated 60 times in a single SAG (Frasier *et al.* 2007a). The presence of sperm from multiple males in the female reproductive tract creates an environment that promotes sperm competition. This is an ideal situation for the female to select for advantageous sperm, which is known as cryptic female choice. Post-copulatory mechanisms of mate choice are especially difficult to study in species with internal fertilization. Due to cryptic female choice being subtle, the effects are often masked by male-driven processes and more difficult to detect (Birkhead & Pizzari 2002).

Right whales have some of the lowest levels of genetic variability found in any wild population. Previous studies have found that this low variation appears to be influencing reproductive performance. Specifically, Frasier *et al*. (2013) found that the majority of matings are unsuccessful, and that successful fertilization only occurs between gametes that are particularly genetically dissimilar (based on genotypes at 35 microsatellite loci). Successful fertilizations therefore result in calves that are more heterozygous than expected from this gene pool (Frasier *et al.* 2013). Right whales do not appear to make a pre-copulatory choice for mates, so the "choice" must be post-copulatory, most likely through biased fertilization patterns (Frasier *et al.* 2013). Similar biased fertilization patterns are

seen throughout nature (Olsson *et al.* 1996; Kempenaers *et al.* 1996; Tregenza & Wedell 2000; Foerster *et al.* 2003; Firman & Simmons 2008; Dziminski *et al.* 2008). However, the mechanism that drives these patterns remains unknown, and identifying the responsible genes is considered high priority (Tregenza & Wedell 2000; Birkhead & Pizzari 2002; Mays Jr & Hill 2004).

1.1 Fertilization and Candidate Genes

Fertilization is mediated by a complex series of interactions between the sperm and the egg, key steps of which require complimentary interactions between proteins expressed on the surface of each gamete (Primakoff & Myles 2002). These interactions result in non-random fertilization patterns, and can scale-up to have large impacts on patterns of individual reproductive success within populations, and on the development of reproductive barriers during speciation (Turner & Hoekstra 2008). Reproductive proteins have been studied extensively in sea urchins and abalones, mainly because of the large availability of gametes from species that have external fertilization (Turner & Hoekstra 2008; Aagaard *et al.* 2009, 2013; Pujolar & Pogson 2011; Vacquier & Swanson 2011; Kosman & Levitan 2014; Wilburn & Swanson 2016). In sea urchins, the genotype of the receptor on the egg determines which sperm genotype can result in successful fertilization. Therefore, there is not a single best, or favoured, genotype, but rather which sperm genotype is "best" depends on the genotype of the female (Palumbi 1999). It is likely that a similar mechanism to the one in marine invertebrates may exist in mammals, except evidence suggests that many more genes are involved, and therefore the gamete interactions are likely much more complex (Swanson & Vacquier 2002).

Until 2014, no interacting ligand-receptor genes had been identified in mammals. However, in 2014 two interacting genes were identified: Izumo1 expressed on the sperm, and Juno the receptor on the egg. The interaction of Juno and Izumo1 is a necessary event for adhesion between the cell membranes of the two gametes (Bianchi *et al.* 2014). In addition to these two complementary genes, several genes that influence fertilization have been identified, but their exact roles and complimentary nature remain to be identified. These include CRISP1, CRISP2, PKDREJ, PH-20, and Zonadhesin on the sperm, and Cd9a and ZP3 on the egg (Chapters 3 & 4).

1.2 Objectives

Although there are multiple reasons why the North Atlantic right whale remains endangered, the reduced reproductive rate is a large factor. The low genetic diversity of the population may be a contributing factor to this low reproductive rate (Frasier *et al.* 2013), but the underlying mechanism remains unknown. To improve our understanding of the mechanism underlying this process in right whales, and to increase our understanding of the molecular aspects of fertilization in mammals, in general, sequence analyses were conducted of putative genes involved in gamete compatibility in identified mother-father-calf triads to test for biased inheritance patterns, as were found in the previous study.

The candidate genes reviewed in earlier chapters have known structures and sequences. By sequencing the interactive sites of these genes in mother-father-calf triads, mating pattern bias could be tested for, with some combinations being over or under represented to determine if there are pre- or post-copulatory mate choices present in this species based on the characteristics of these genes. Biased inheritance patterns were also

tested for in the offspring to test for cryptic female choice with an egg selecting a sperm with a certain genotype over another. Variability was also assessed within the North Atlantic right whale population, and how it is influencing an individual's overall reproductive performance.

These analyses can potentially identify what genes play a role in gamete compatibility, not only in right whales, but in other species that show these biased fertilization patterns. Also, they can provide some insight as to how gamete compatibility can affect reproductive success in populations that have been reduced to a small size. These analyses have the potential to address still unknown processes involved in reproductive success, speciation, and the maintenance of genetic variation in small populations.

2. Methods

2.1 Samples and Paternity Data

To obtain small skin samples from free-swimming right whales, a crossbow is used with a modified bolt (arrow) (Brown *et al.* 1991). This method is widespread in the study of cetaceans (whales, dolphins, and porpoises), and, if done properly, does not cause any negative short- or long-term effects on the individuals, other than an initial startle response (Best *et al.* 2005; Noren & Mocklin 2011). Samples have been collected from North Atlantic right whales since 1988, and are collected in conjunction with photo identification of the individual whales, which can be identified on an individual level using callosity and scaring patterns (Payne *et al.* 1983; Kraus *et al.* 1986). DNA is extracted using standard

phenol:chloroform protocols (this was step was completed at Trent University by another lab) (e.g., Wang *et al.* 2008).

Since right whale calves are not weaned for approximately 12 months (Hamilton *et al.* 1995), mother-calf relationships can be identified using this close association between mothers and calves throughout the year. Fathers are assigned on an annual basis using genetic data. Briefly, once DNA is extracted from each sample it is profiled at a range of molecular markers, including: (1) a sex-specific marker, (2) sequence analysis at a portion of the mitochondrial control region (Malik *et al.* 2000), and (3) genotype analysis at 35 microsatellite loci (Frasier *et al.* 2006). These genetic data are used to confirm maternities, as well as to assign paternities to the identified calves using the methods described in Frasier et al. (2007b). These paternity assignments are stored in the genetic database.

2.2 Sequencing

The gamete compatibility genes that were chosen to analyze were CRISP1, CRISP2, Izumo1, PH-20, PKDREJ, and Zonadhesin on the sperm and Cd9, Juno, and ZP3 on the egg. Justification for these genes can be found in Chapters 2 & 3. A total of 16 primer sets, for 9 candidate genes, were used for sequencing. Amplification of each sample at each locus took place in 40 µl reactions containing 2 µl of DNA (at a concentration of 5 ng/µl for total of 10 ng), as well as 1x PCR buffer (Promega), 0.2 mM each dNTP (Invitrogen), 0.4 mg/mL bovine serum albumin (Ambion), 1.5 mM $MgCl_2$ (Promega), 0.3 µM of each the forward and reverse primer, and 0.05 U/µl *Taq* polymerase (Promega). The PCR amplification conditions differed for some loci (Chapter 4, Table 2). After PCR, excess unincorporated nucleoside triphosphates and primers were removed through an enzymatic

process using a standardized amount of DNA (usually 5 μl), 0.129 μl of Antarctic phosphatase buffer/μl of DNA (New England BioLabs), 0.02 μl of Antarctic phosphatase/μl of DNA (New England BioLabs), and 0.00614 μl of Exonuclease I/μl DNA (New England BioLabs). The samples were heated to 37°C for 15 minutes, and then 80°C for 15 minutes. The sequencing reaction was based on the BigDye v3.1 sequence kit (Applied Biosystems) and consisted of a cocktail of 15 μl with 5.78 μl of product from the previous reaction (if standardized to 5 μl), 0.25x reaction mix, 1x sequencing buffer, and 1μl of primer at a concentration of 10 μM. The sequencing PCR conditions were 96°C for 2 min; and then 30 cycles of 96°C for 20 seconds, 50°C for 20 seconds, and 60°C for 4 minutes. After the sequencing reaction, the salts, unincorporated dNTPs and primers need to be removed from the samples for proper injection and capillary electrophoresis. This was done using the protocol described in Irwin *et al.* (2003). The samples were then resuspended in 10 μl of HiDi formamide (Applied Biosystems).

Sequencing products were size-separated and visualized on an ABI 3500xl Genetic Analyzer (Applied Biosystems). Initially, individuals from 3 different mother-father-calf triads (for a total of nine individuals) were sequenced at each locus to look for variable sites, both in the forward and reverse direction to confirm potential variable sites. Sequences were manually trimmed using 4peaks (Mekentosj), and aligned using Clustalx (Jeanmougin *et al.* 1998; Thompson *et al.* 2002). This initial screening only identified two loci with variable sites: one at locus PKDREJ-S1b, and one at locus Cd9a. The remaining right whale samples were sequenced at these two loci, in the forward direction, and approximately half of the variable individuals were sequenced in the reverse direction to confirm the variable sites. An additional variable site was subsequently found in locus

PKDREJ-S1b after sequencing more individuals. It is possible that there are more variable sites in other loci that were not seen in the original nine individuals. Because of this, additional individuals were sequenced at both the Juno and the Izumo1 locus, since these could be the most telling because they are already known to interact. However, after sequencing 13 individuals at locus Juno1, 18 individuals at locus Juno2, and 33 individuals at locus Izumo1, no variable sites were seen.

3. Statistical Analyses

3.1 Mating Pairs

A total of 152 individuals, that made up 69 mother-father-calf triads, were sequenced at loci Cd9a and PKDREJ-S1b. Due to some of the issues with frequentist statistics and p-values (e.g., Berger & Berry 1988; Halsey *et al.* 2015; Kruschke 2014), a Bayesian approach was used. To test if the frequency of mating pairs were biased towards individuals with certain allelic combinations (within and between loci), a Bayesian version of a contingency test was used, in which the observed versus expected frequency of mating pairs are compared for each allelic combination. A detailed description of the Bayesian model used to perform these analyses is provided in **Appendix 1**. This test provides a means to test if some allelic combinations are over- or under-represented, and therefore to test if mate choice patterns appear to be influenced by the genetic characteristics of potential mates at these genes. The mating pairs were combined regardless of the sex of the parent (i.e. if the mother or father was heterozygous it would be counted with AA/AG).

3.2 Variation and Sex

Because some of the loci are expressed on the sperm and others are expressed on the egg, patterns of allelic combinations were tested with respect to sex. For this the same version of a Bayesian contingency table was used, where the observed counts of individuals of each sex with each genotype were used as the observed data, and compared to those expected if genotypes were distributed evenly across the sexes.

3.3 Inheritance Patterns

Lastly, biased allele inheritance patterns were tested for among mother-father-calf triads, as was found with the microsatellite data. Specifically, a signal of biased fertilization patterns was tested for by testing if certain allelic combinations were more common in offspring than expected given straight Mendelian inheritance. Because my sample size was relatively small, instead of testing for specific allelic combinations, the data were combined and tested if offspring tended to inherit the same, or different, alleles from each parent more often than expected. The previous microsatellite data found a bias where offspring inherited different alleles from each parent more often than expected, but the opposite could be true for these loci as well, because for some gamete compatibility genes, in some species, fertilizations are biased towards similar alleles. The specific model used to conduct this analysis is described in **Appendix 2**.

4. Results

A total of three SNPs (single nucleotide polymorphisms) were found across two different loci. Cd9a was variable at one site, with an A/G substitution at position 448, in a total of 13 individuals (Table 1). This substitution is non-synonymous; GAT codes for aspartic acid, whereas the allele with GGT codes for glycine.

Table 1: SNPs found in the North Atlantic right whale (line represents the base pairs that make up the amino acid)

Locus	Allele 1	Allele 2
Cd9a	AGAGATGCATGG	AGAGGTGCATGG
PKDREJ-S1b (Site 185)	AAGAACGACAGA	AAGAATGACAGA
PKDREJ-S1b (Site 464)	GTTTCGTTGAA	GTTTTGTTGAA

Table 2: Variation counts for locus Cd9a

Genotype	**Triad**			**Total**
	Mother	**Father**	**Calf**	
AA	44	42	53	139
AG	6	3	4	13
GG	0	0	0	0
Total	50	45	57	152

Genotype	**Sex**		**Total**
	Male	**Female**	
AA	68	71	139
AG	5	8	13
GG	0	0	0
Total	73	79	152

PKDREJ-S1b had two variable sites, the first one at site 185, containing a C/T substitution. Thirteen individuals were heterozygous at this site, with one individual homozygous for the T allele (Table 3). The second PKDREJ-S1b variable site was at

position 464, with 31 heterozygous individuals, also a C/T substitution (Table 4). Both SNP's in PKDREJ-S1b are synonymous substitutions, meaning that they code for the same amino acids. At site 185, AAC and AAT both code for Asparagine, and at site 464 TTC and TTT code for Phenylalanine.

Table 3: Variation counts for locus PKDREJ-S1b: Site 185

Genotype	**Triad**			**Total**
	Mother	**Father**	**Calf**	
CC	45	41	52	138
CT	5	3	5	13
TT	1	0	0	1
Total	51	44	57	152

Genotype	**Sex**		**Total**
	Male	**Female**	
CC	65	73	138
CT	7	6	13
TT	0	1	1
Total	72	80	152

Table 4: Variation counts for locus PKDREJ-S1b: Site 464

Genotype	**Triad**			**Total**
	Mother	**Father**	**Calf**	
CC	43	30	48	121
CT	8	14	9	31
TT	0	0	0	0
Total	51	44	57	152

Genotype	**Sex**		**Total**
	Male	**Female**	
CC	53	68	121
CT	19	12	31
TT	0	0	0
Total	72	80	152

Only three individuals were heterozygous at multiple variable sites. One calf (female) was heterozygous at Cd9a and PKDREJ-S1b site 464. Two individuals, one calf (male), and one father were heterozygous at Cd9a and PKDREJ-S1b site 185.

Bayesian statistics do not involve testing to see if an effect is “significant” or not. Instead, the results are presented as posterior probability distributions that are associated

with each hypothesized value (Kruschke 2014). Therefore, instead of *P*-values, my results will be plots of the observed and expected values, and the statistics will be plots of the resulting posterior probability distributions. The dashed red bars will indicate the value expected for a "zero" effect, which would be the case if there is no interaction between the two sets of predictors. A plot is shown for each possible combination.

4.1 Variation and Mating Pairs

For locus Cd9a, there were no differences between the number of observed and expected mating pairs of each genotyping combination (Figure 1). As a result, there was not any statistical evidence for some combinations being over- or under-represented (data not shown).

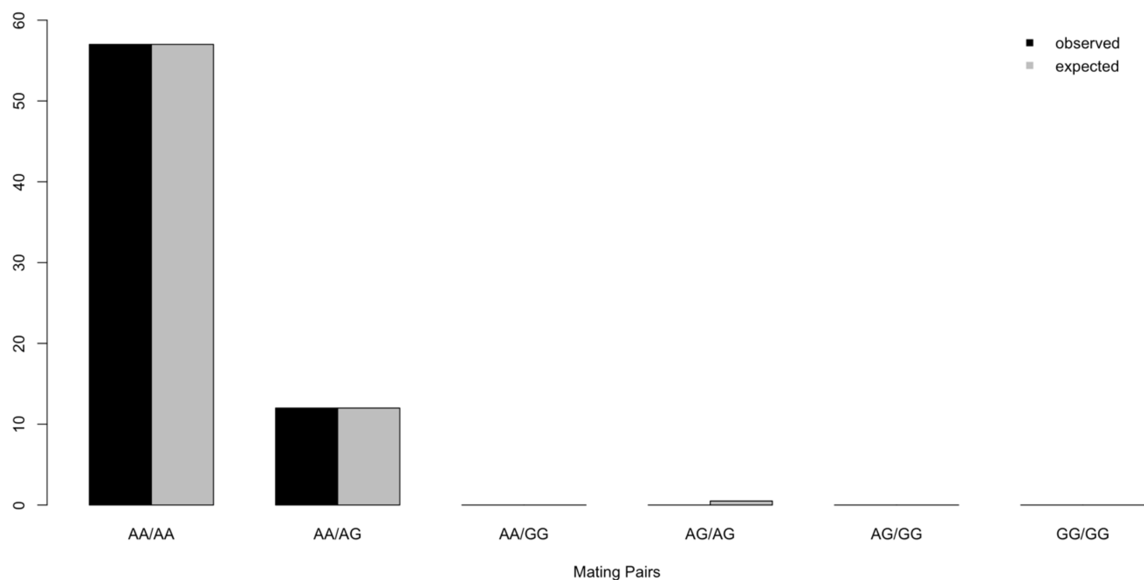


Figure 1: Observed vs. Expected Counts of Mating Pairs for Variable Site 448 at locus Cd9a

There was a similar lack of substantial bias in mating pairs for locus PKDREJ-S1b site 185. The observed and expected number of mating pairs are similar (Figure 2). There are only low interactions effects seen between parents, but there are some interesting patterns. For example, mating pairs both homozygous for the same alleles (CC with CC) are substantially over-represented in the data set, whereas those homozygous for different alleles (CC with TT) are substantially under-represented (Figure 3). However, these differences are too slight to really be seen in the expected mating pairs.

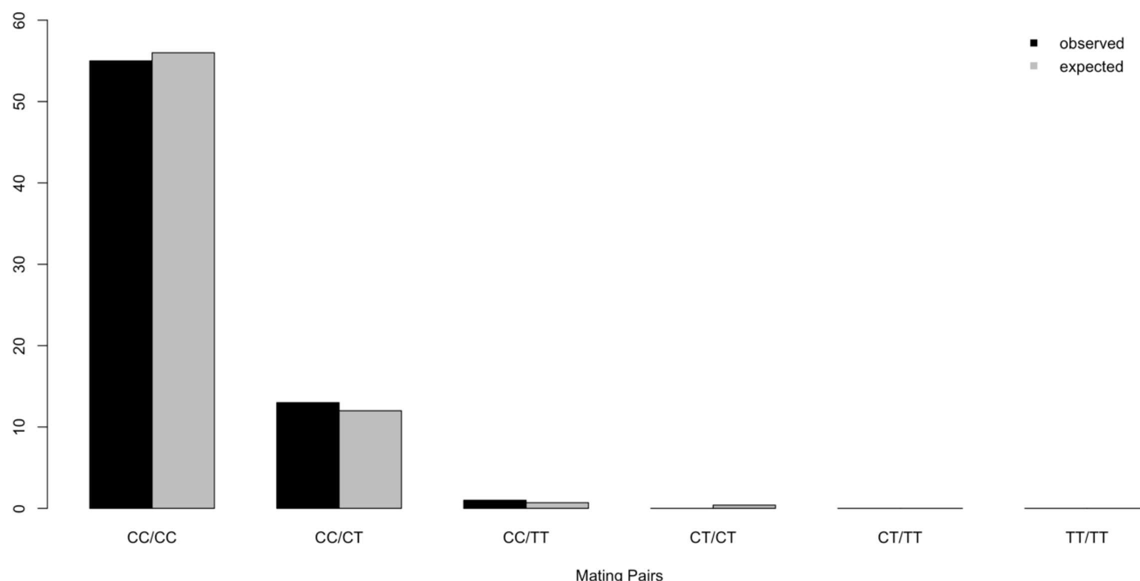


Figure 2: Observed vs. Expected Counts of Mating Pairs for Variable Site 185 at locus PKDREJ-S1b

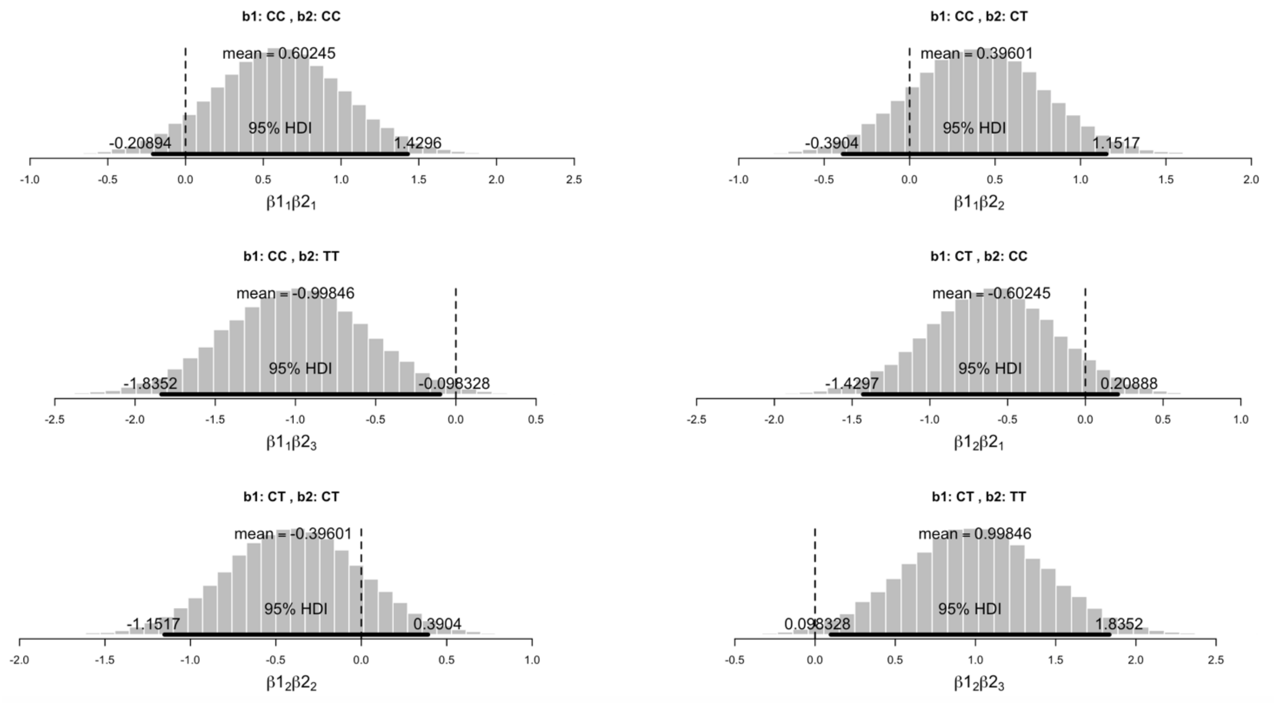


Figure 3: Interaction effects for different combinations of mating pairs at locus PKDREJ-S1b site 185. A positive value represents a combination that is occurring more than expected, and a negative value represents combinations that are occurring less than expected. The dashed line at zero represents the value of a "zero" effect.

With locus PKDREJ-S1b at site 464 there was a similar trend as locus Cd9a. The observed and expected number of mating pairs are the same. As a result, there was not any statistical evidence for some combinations being over- or under-represented (data not shown).

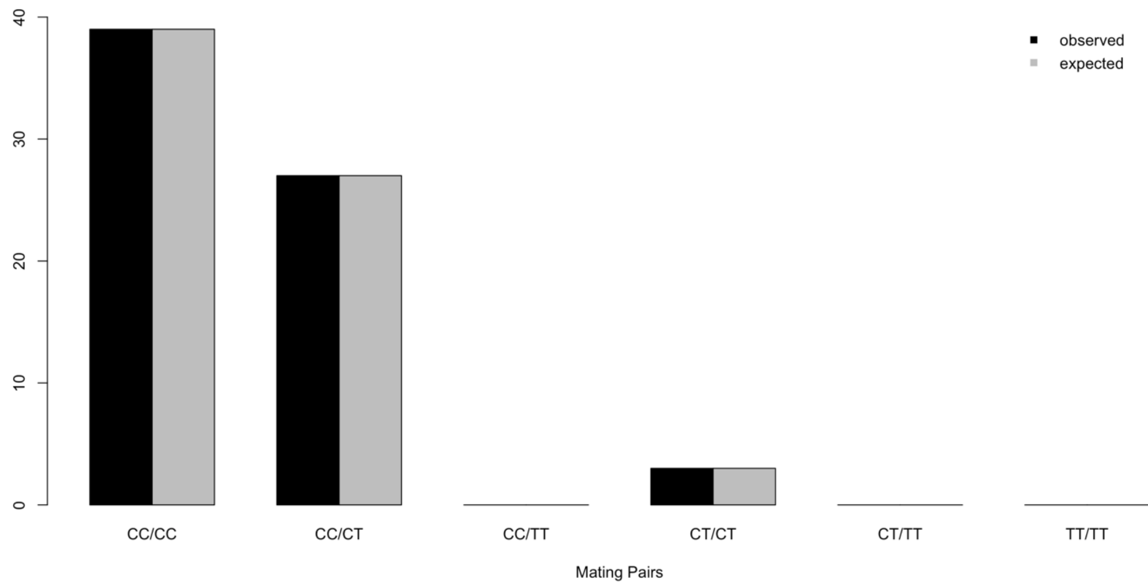
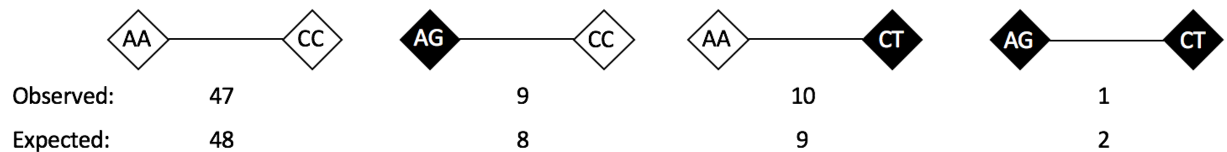


Figure 4: Observed vs. Expected Counts of Mating Pairs for Variable Site 464 at locus PKDREJ-S1b

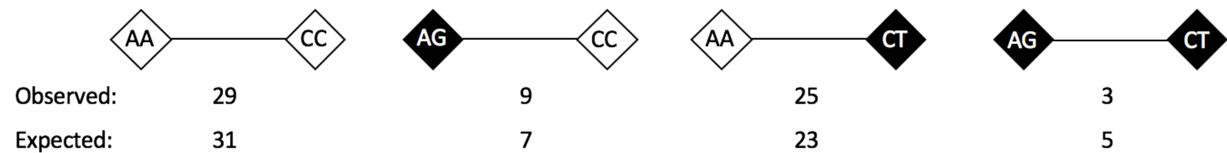
However, when testing for interactions between loci, there are notable differences in the number of observed and expected mating pairs (Figure 5). For interactions between locus Cd9a and PKDREJ-S1b site 185, the differences between observed and expected mating pairs are small, and no signs of interaction are seen, with the mean values centering around zero (Figure 6). However, for interactions between Cd9a and PKDREJ-S1b site 464, and between PKDREJ-S1b site 185 and PKDREJ-S1b site 464, there are larger differences between observed and expected, and in similar patterns. For both sets, there are more sets of mating pairs when one parent is homozygous at one locus, and heterozygous at the other. However, if both parents are heterozygous at the locus, there are fewer mating pairs than expected. This same pattern of fewer mating pairs than expected is also seen if both parents are homozygous at each locus. This is supported by the interaction effects,

with homozygous mating pairs being underrepresented in the population as shown by a negative interaction value, and mating pairs where one parent is homozygous and the other is heterozygous has a positive value, suggesting they are overrepresented (Figure 7 & 8). Although these apparent interaction effects are interesting, and worth pursuing for future studies, my current sample size seems too small to draw any conclusions at this time.

Cd9a vs. PKDREJ-S1b site 185



Cd9a vs. PKDREJ-S1b site 464



PKDREJ-S1b site 185 vs. PKDREJ-S1b site 464

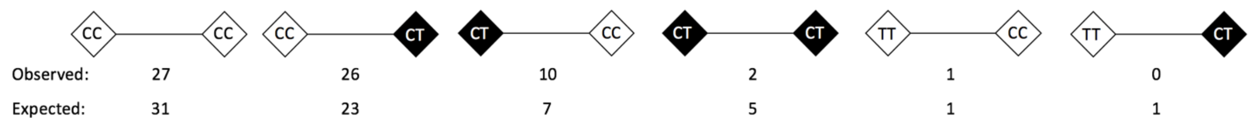


Figure 5: Observed versus expected counts of mating pairs when comparing between loci (black diamonds represent heterozygous individuals).

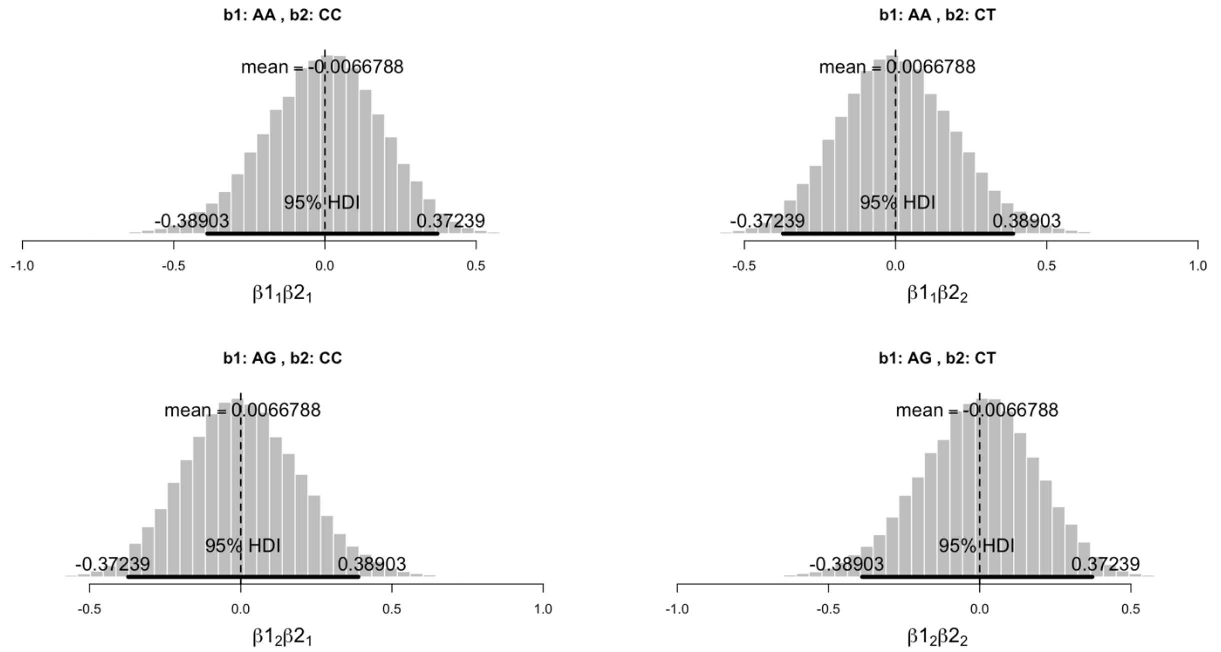


Figure 6: Interaction effects of Cd9a vs. PKDREJ-S1b site 185. A positive value represents a combination that is occurring more than expected, and a negative value represents combinations that are occurring less than expected. The dashed line at zero represents the value of a “zero” effect.

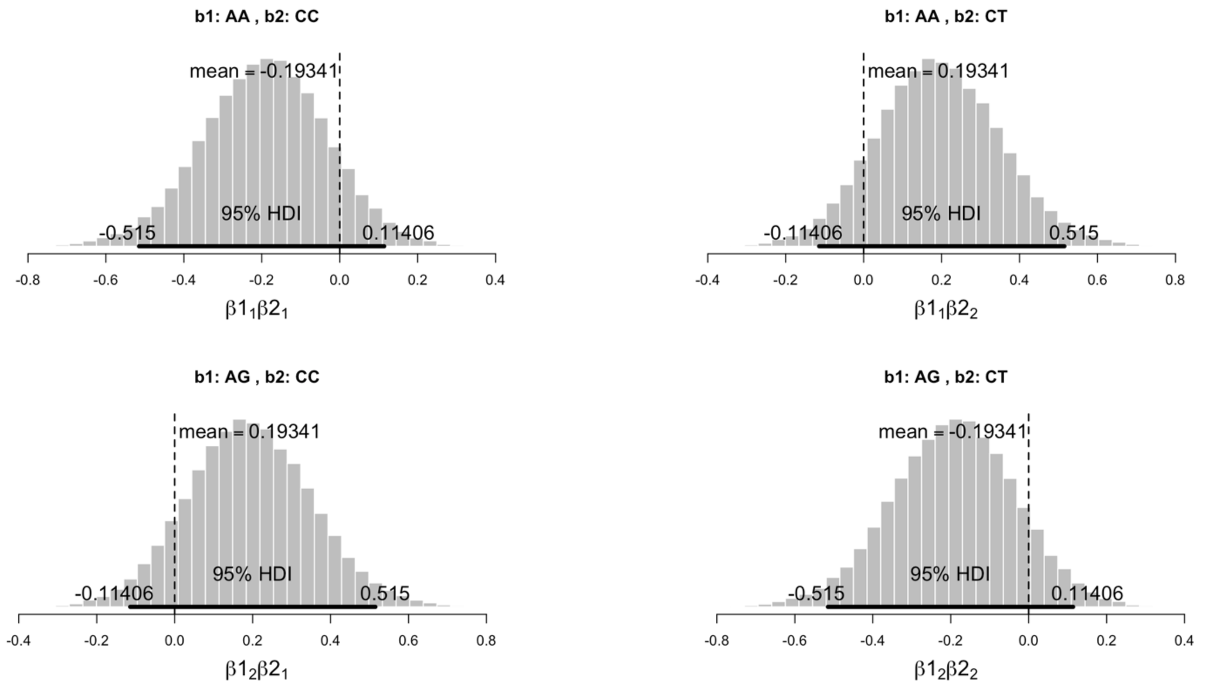


Figure 7: Interaction effects of Cd9a vs. PKDREJ-S1b site 464. A positive value represents a combination that is occurring more than expected, and a negative value represents combinations that are occurring less than expected. The dashed line at zero represents the value of a “zero” effect.

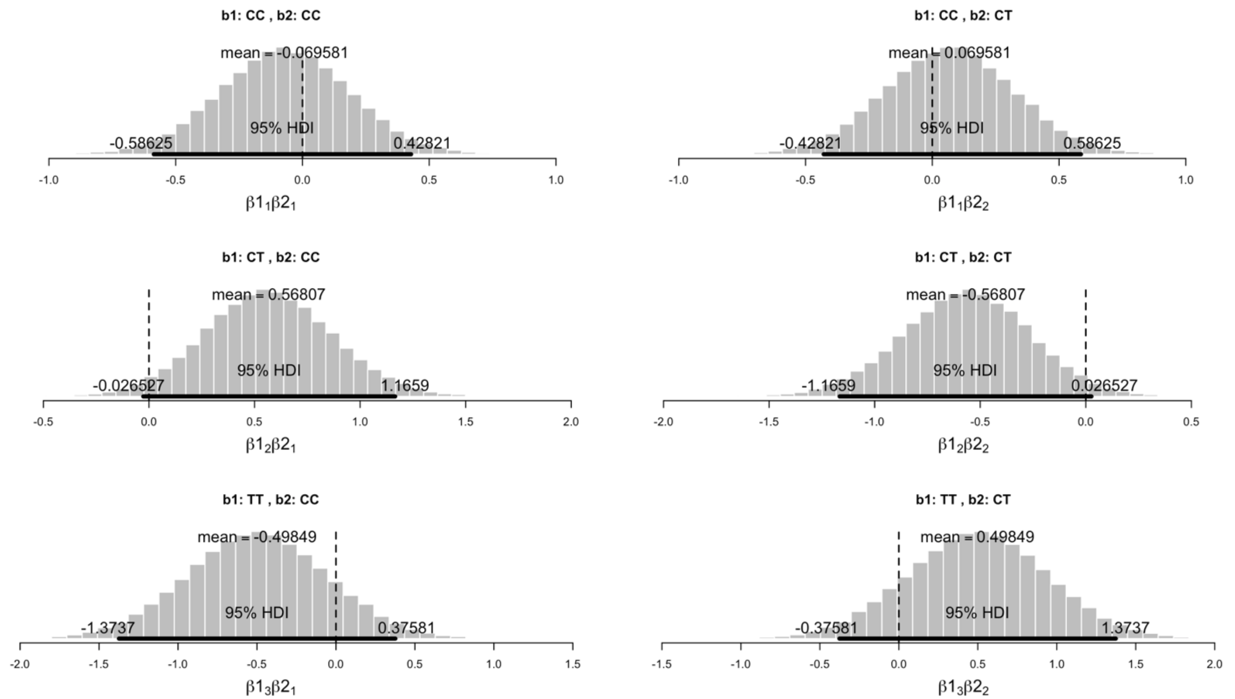


Figure 8: Interaction effects of PKDREJ-S1b site 185 vs. PKDREJ-S1b site 464. A positive value represents a combination that is occurring more than expected, and a negative value represents combinations that are occurring less than expected. The dashed line at zero represents the value of a "zero" effect.

4.2 Variation and Sex

PKDREJ is only expressed on the sperm, and even though the bias is not large, there is a slightly higher association between being male and being heterozygous at either site (Figure 10 & 12). The posterior probabilities represent the interaction effect between sex and genotype. A higher number shows that these two groups are more likely to be found together. However, there are no large differences between the number of expected and observed individuals of either sex for the first variable site (185) (Figure 10). The second site at base pair 464 does show a slight pattern where males are more likely to be

heterozygous than expected, and females are more likely to be homozygous. The interaction values reflect this, but the values are too small to represent a large effect of sex on this variable site.

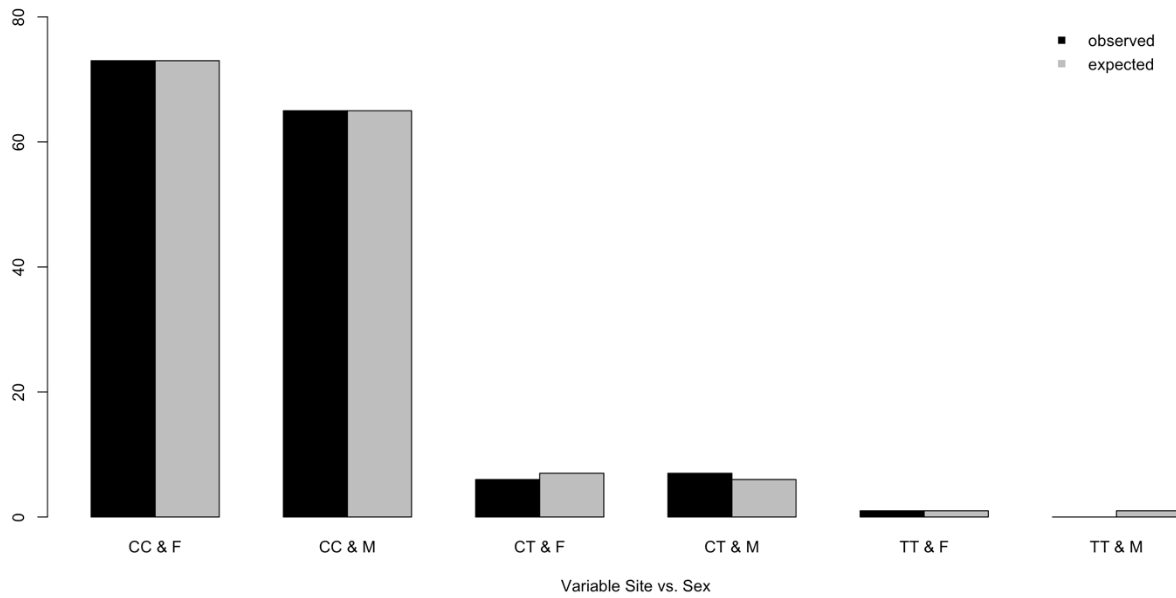


Figure 9: Observed vs. Expected values for PKDREJ-S1b Site 185

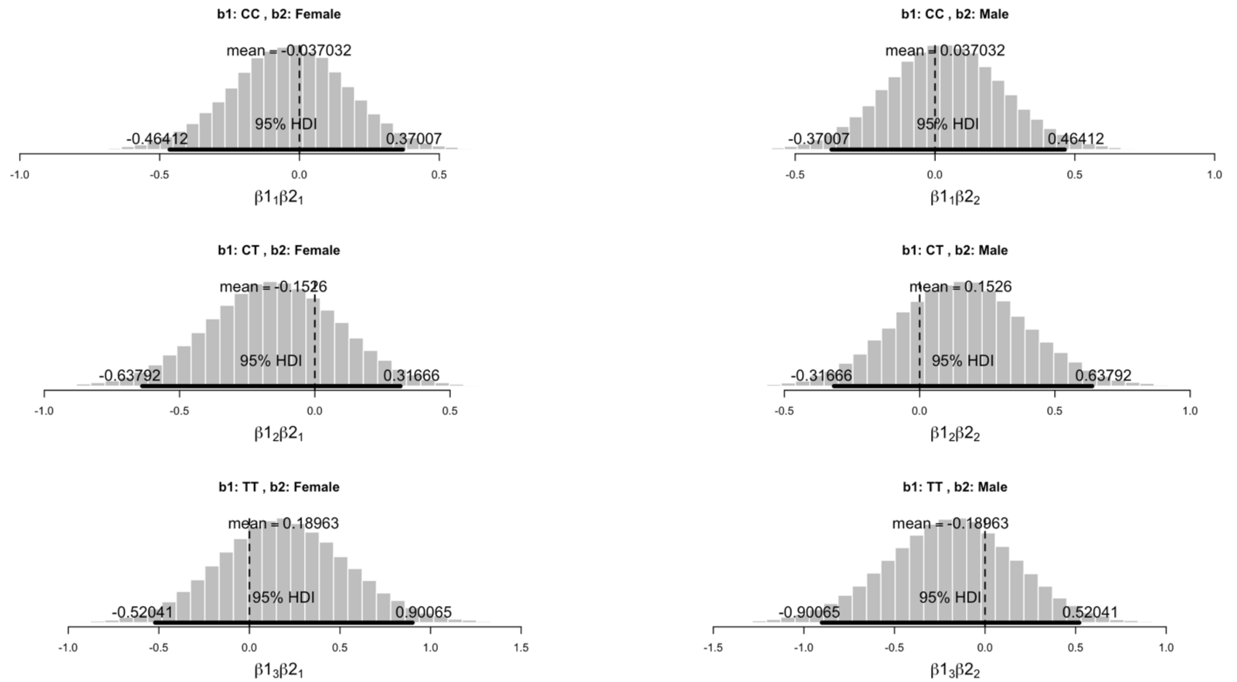


Figure 10: Interaction Values for PKDREJ-S1b Site 185 and Sex. A positive value represents a combination that is occurring more than expected, and a negative value represents combinations that are occurring less than expected. The dashed line at zero represents the value of a "zero" effect.

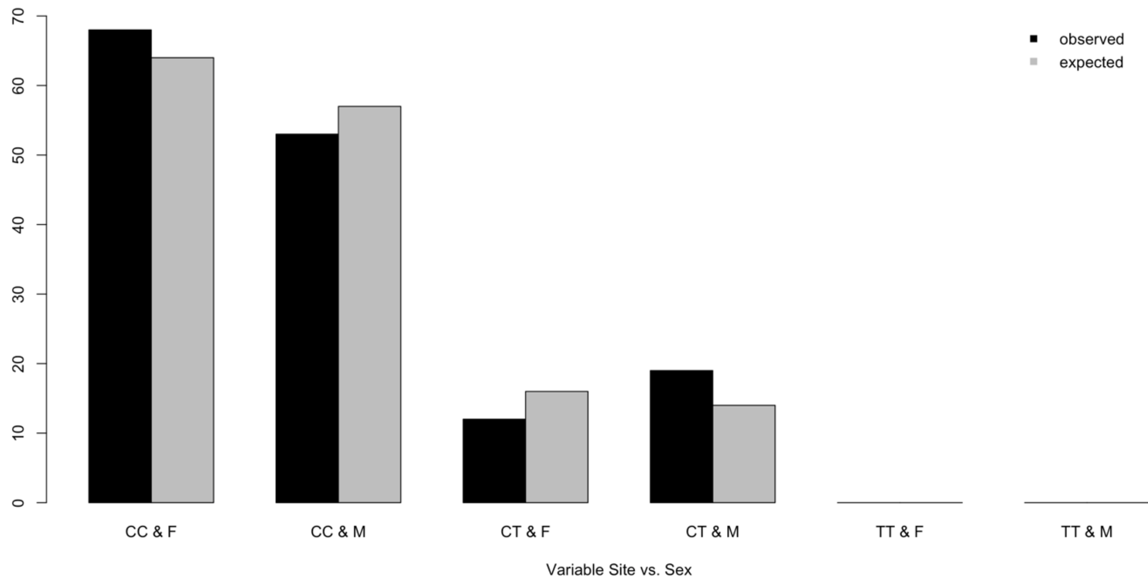


Figure 11: Observed vs. Expected values for PKDREJ-S1b Site 464

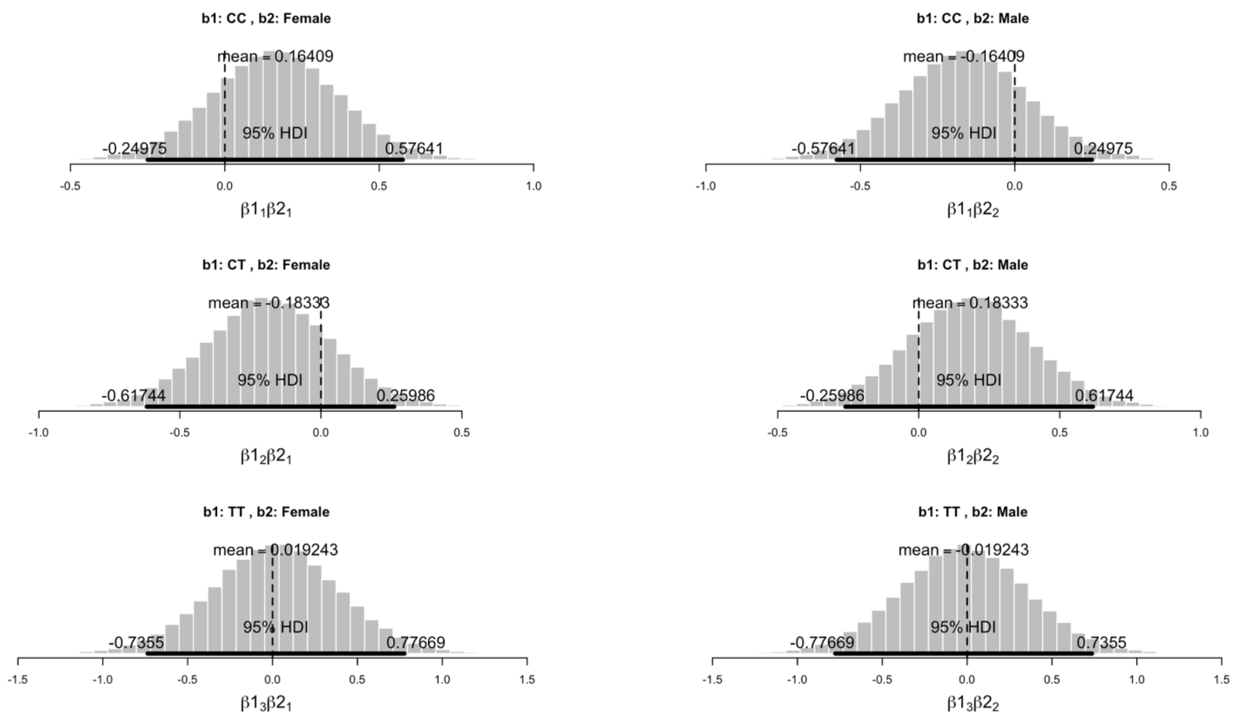


Figure 12: Interaction Values for PKDREJ-S1b Site 464 and Sex. A positive value represents a combination that is occurring more than expected, and a negative value represents combinations that are occurring less than expected. The dashed line at zero represents the value of a “zero” effect.

Cd9 originally was thought to be expressed only on the egg, but recently was found also on the sperm (Antalíková *et al.* 2015). There are only slight differences seen in the observed and expected numbers, where slightly more females are heterozygous than expected. This is supported by the interaction values (Figure 15), which show that females are more likely to be heterozygous.

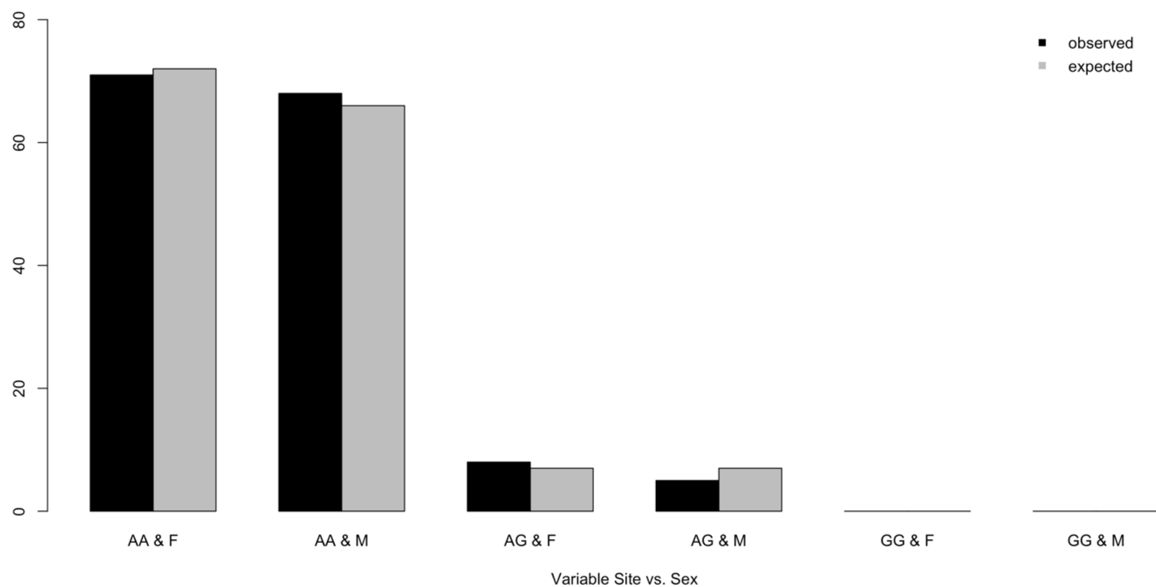


Figure 13: Observed vs. Expected values for Cd9

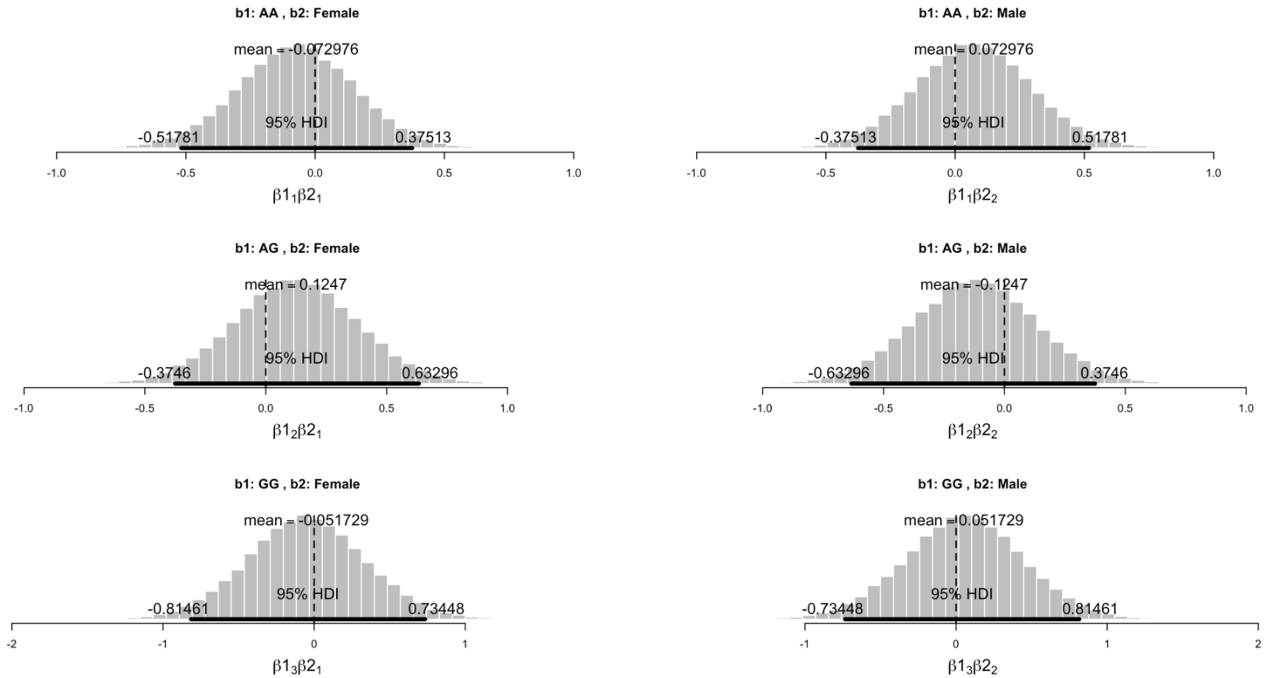


Figure 14: Interaction Values for Cd9 and Sex. A positive value represents a combination that is occurring more than expected, and a negative value represents combinations that are occurring less than expected. The dashed line at zero represents the value of a "zero" effect.

4.3 Inheritance Patterns

Two of the three mother-father-calf triads combinations showed notable differences from Mendelian inheritance patterns. For instance, at locus PKDREJ-S1b site 464, there was a total of 23 triads with one homozygous parent. Offspring were more likely to inherit a maternal allele that was the same as the paternal, resulting in more homozygous offspring than expected based on Mendelain inheritance at this locus. A similar, albeit smaller pattern is seen at locus Cd9a, with the offspring more likely to be homozygous than expected. Locus PKDREJ-S1b site 185 follows Mendelian inheritance patterns.

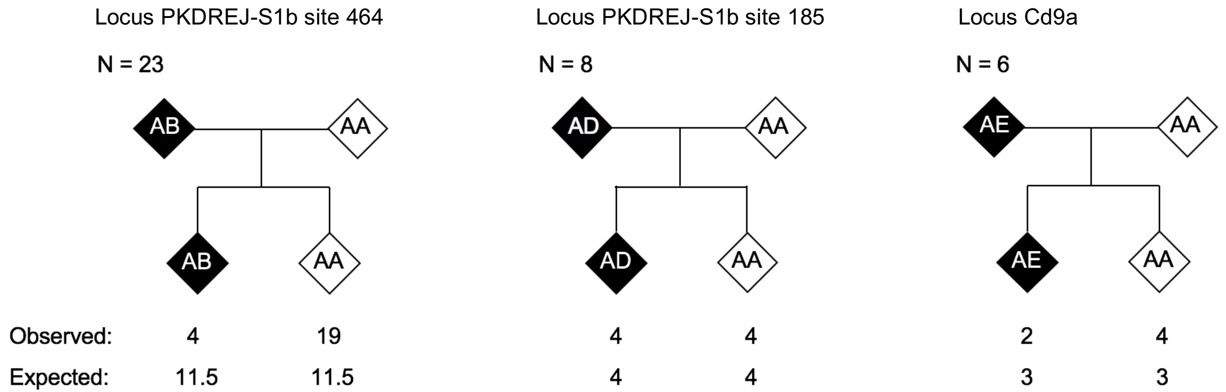


Figure 15: Pedigrees for the North Atlantic right whale based on genotype. Uninformative genotypes (ones with low sample numbers, or instances where offspring would have to inherit differing alleles) were not analyzed. Pedigrees were also designed without separating based on which parent had the mutation. Haplotype Definition: A: A Cd9a, C 185 PKDREJ-S1b, C 464 PKDREJ-S1b; B: A Cd9a, C 185 PKDREJ-S1b, T 464 PKDREJ-S1b; D: A Cd9a, T 185 PKDREJ-S1b, C 464 PKDREJ-S1b; E: G Cd9a, C 185 PKDREJ-S1b, C 464 PKDREJ-S1b;

Haplotypes not shown or seen: C: A Cd9a, T 185 PKDREJ-S1b, T 464 PKDREJ-S1b; F: G Cd9a, T 185 PKDREJ-S1b, C 464 PKDREJ-S1b; G: G Cd9a, T 185 PKDREJ-S1b, T 464 PKDREJ-S1b; H: G Cd9a, C 185 PKDREJ-S1b, T 464 PKDREJ-S1b

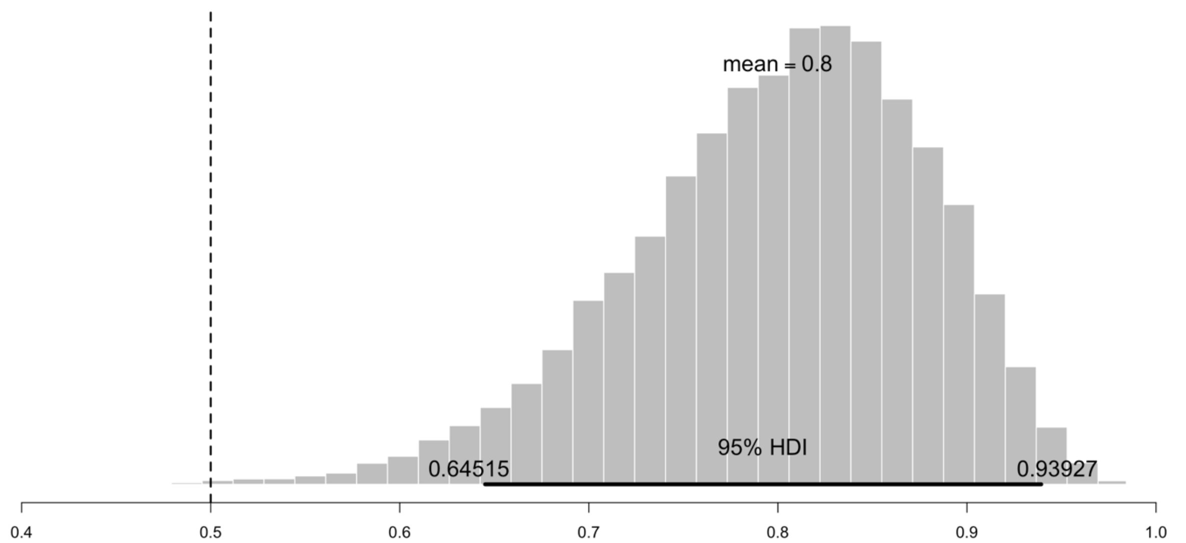


Figure 16: Probability of an offspring inheriting an allele that is the same as the other parent - being homozygous at locus PKDREJ-S1b site 464. The dashed line at 0.5 represents the value the mean should be centered around if inheritance patterns were following Mendelian inheritance patterns.

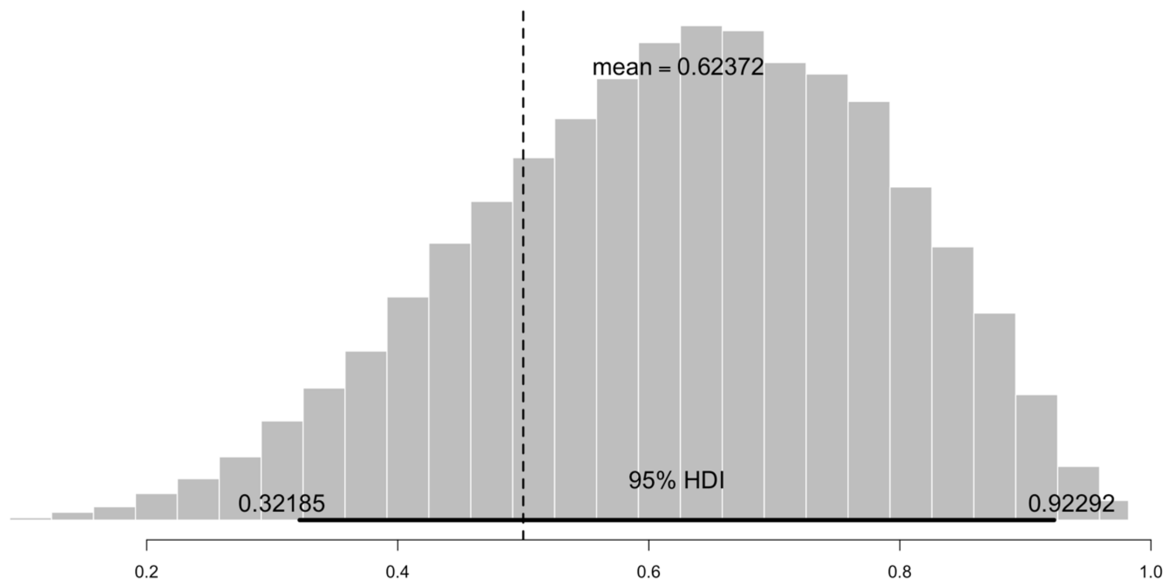


Figure 17: Probability of an offspring inheriting an allele that is the same as the other parent - being homozygous at locus Cd9a. The dashed line at 0.5 represents the value the mean should be centered around if inheritance patterns were following Mendelian inheritance patterns.

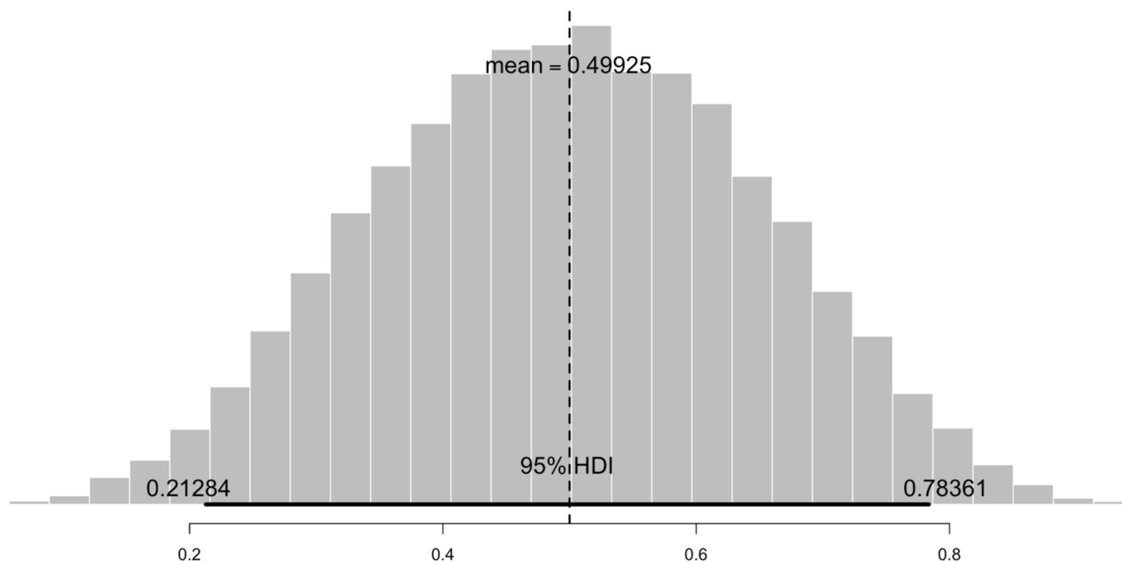


Figure 18: Probability of an offspring inheriting an allele that is the same as the other parent - being homozygous at locus PKDREJ-S1b site 185. The dashed line at 0.5 represents the value the mean should be centered around if inheritance patterns were following Mendelian inheritance patterns.

For PKDREJ-S1b, site 464, based on Mendelian inheritance, there should be a 50% chance that the offspring of AA and AB parents would be AA. However, the Bernoulli distribution shows that offspring are more likely to be homozygous, as the probability is higher for being homozygous than the normal 50%. For Cd9, there is also a bias, although smaller, for the offspring to be homozygous. The probability for being homozygous at this locus is also higher than 50%. At locus PKDREJ-S1b, site 185, the inheritance patterns following Mendelian inheritance, with the observed and expected number of offspring being equal.

5. Discussion

The North Atlantic right whale has among the lowest levels of genetic variability of all studied wildlife populations (Kraus & Rolland 2007). Thus, it was not unexpected that low levels of variation were found in the candidate genes, with only three variable sites. Additionally, understanding of the genes influencing fertilization is in its infancy, and likely there are genes important to gamete compatibility, but which are still unidentified, and therefore were not analyzed.

5.1 Variation and Mating Pairs

By comparing the genotypes of mating pairs, it can be tested to see if there are certain patterns of mate choice in the North Atlantic right whale with respect to alleles at these genes. When considering data from each locus independently, there is not a large bias seen for any loci: the observed and expected counts are nearly the same for each combination of mating pair. There is a slight bias within loci in which homozygous mating pairs are more likely to occur, but due to my low sample size it is difficult to say if this is indicative of a strong effect in mate choice.

When comparing between pairs of loci, instead of within, an interesting pattern emerges. The fact that there are more observed than expected mating pairs when one parent is heterozygous at one locus, and the other parent is homozygous at the other locus. This could indicate that there is mate selection for mates of differing genotypes at these fertilization genes, with the interaction occurring across (rather than within) loci. This is supported by some combinations showing relatively strong interaction effects (Figures 7 &

8). It is possible that locus PKDREJ-S1b site 464 interacts with both locus Cd9a and locus PKDREJ-S1b site 185. Despite this, the small sample size is limiting, and these differences are slight, as are the interaction values, which suggest that this may not be a strong indicator of mate choice within this population, but is worth further investigation.

A similar pattern of mate selection for a mate that had heterozygous genotypes was seen in wolves (*Canis lupus*), which are highly inbred. In this population, heterozygous males are preferentially selected for as breeders, and their offspring were more heterozygous than offspring from random matings. This selection allows for heterozygosity to be maintained within the population, despite inbreeding (Bensch *et al.* 2006). It is possible that there is a similar effect in right whales, in which mating with individuals that have differing genotypes can increase the fitness of the offspring. There is an unknown mechanism that is increasing heterozygosity in right whale offspring (Frasier *et* al. 2013). However, further research would be needed before being able to conclusively say any effect this may have.

5.2 Variation and Sex

If one allele is beneficial to one sex, and detrimental in the other, it would be expected to see a bias in which sex is heterozygous (Haig *et* al. 2014). Cd9 was thought to only be expressed on the egg, but more recently was shown to also be expressed on the sperm plasma membrane (Antalíková *et al.* 2015), and may interact with itself during sperm-egg interaction. There is not a large difference between which sex is heterozygous, with the G substitution being present in 8 females, and 5 males. The interaction effects are slightly higher for being female and being heterozygous, but the most likely effect size is

not far from zero (Figure 14), suggesting there is little to no effect on what sex is heterozygous.

Research suggests that PKDREJ is only expressed on the sperm (Hughes *et al*. 1999), and is not known to have an effect in females. The first variable site only has one more male than female with a T substitution, showing no clear bias, suggesting that this allele does not have an effect (either negative or positive) in males or females. However, being heterozygous at the second variable at site 464, is seen in 19 males and only 12 females. The interaction between being male and having either allele for PKDREJ-S1b is only slightly higher than the interaction between females and the alleles (Figure 12). This suggests that sex does not have an effect on heterozygosity at these genes.

5.3 Inheritance Patterns

To try to increase the power of the analyses, the pedigrees were combined so that instead of testing for specific allelic combinations, triads were combined indicating the same scenarios (i.e., one parent being homozygous and the second parent being heterozygous with one shared allele, regardless of what those alleles are). These pedigrees were then used to test if the offspring inherited alleles that were different from the other parent's alleles more often than expected. This pattern has been seen previously in the North Atlantic right whale with different molecular markers (Frasier *et al.* 2013). However, a surprising pattern emerged in which offspring were more likely to inherit the same allele from both parents, resulting in offspring that were more homozygous than expected. This bias was most drastic when looking at the PKDREJ-S1b locus at site 464. Out of a total of

23 triads, only 4 offspring inherited an allele that differed from that inherited from the other parent (when the expected number from these triads was ~12).

This is not entirely unexpected due to the complicated nature of fertilization. In sea urchins, the egg selects for sperm that has the same bindin genotype (Palumbi 1999). Since PKDREJ is predicted to interact with ZP3 (Hamm *et al.* 2007), and the ZP maintains species specificity (Wassarman *et al.* 2004), it may be beneficial to be homozygous for this allele. A mutation could result in the sperm not being “recognized” by the ZP, and therefore would not result in a successful fertilization. Since this allele is passed on to some offspring, it may just cause a reduction in eggs that are fertilized by sperm with this haplotype. Also, a previous study that inserted a mutation into the PKDREJ allele showed that mice with the mutation had no fertilization success when in sperm competition with a normal mouse (Sutton *et al*. 2008). Since right whales compete through sperm competition, it is possible that this allele is less successful in this environment, and sperm with the “normal” haplotype are more likely to be successful.

Even though at PKDREJ-S1b both alleles represent silent substitutions, this may still affect the gene. Selection for synonymous substitutions can results from codon usage bias, and is seen in multiple organisms (Du *et al.* 2014). However, these selection pressures would not apply to a gene's role in gamete compatibility. It is more likely that this mutation at site 464 is linked to a different, undiscovered, functional mutation on this gene. A previous study has shown that in mammals, the frequency of nonsynonymous substitutions is correlated with synonymous substitutions in the same gene (Mouchiroud *et al*. 1995).

Thus, important selection patterns may still be detected within this gene based on variation at linked sites, rather than having identified the influential site itself.

5.4 Conclusion

This research represents one of the first attempts at studying gamete compatibility genes in mammals. Although there were no strong patterns of mate choice seen in this population, a bias in inheritance patterns was found. Even though it is an opposite pattern from those seen in previous studies, it is possible that these genes play a role in cryptic female choice in this species, and potentially in other cetacean species. Additional research in cetacean species with similar issues may help to understand the role these genes have in individual's reproductive success, maintaining heterozygosity in small populations, and even on the development of reproductive barriers for speciation.

Appendix 1: Bayesian model used for analysis of mating pairs

A contingency test is based on count data and attempts to detect if there are interaction effects among considered variables by comparing observed counts in each combined set of categories to expected values if there are not any interaction effects. For a model using count data, a Poisson distribution is used since the data are positive integers. The model equation is: $y_i \sim Poisson(\lambda)$ and lambda is estimated from $\lambda = \exp(\beta_0 + \beta_1[x[1]] + \beta_2[x[2]] + \beta_{1,2}[x[1],x[2]])$. This represents a log linear model, where the effects are estimated on a log scale (so that they are additive), and then brought back to the initial scale by raising them to the exponent. Thus, the estimated values (and their priors) are on the log scale. β_0 represents the average value across all categories of each predictor variable. β_1 represents the deflection away from the baseline due to being in each category of the first nominal predictor variable, which in this model is the genotype of the first parent. β_2 represents the deflection away from the baseline due to being in each category of the second nominal predictor variable, which in this model is the genotype of the second parent. $\beta_{1,2}$ represents the interaction effects of these two predictors.

The prior probabilities for each predictor variable (β_1, β_2, and $\beta_{1,2}$) are all assumed to follow a normal distribution centered around zero (i.e., no effect), but with wide tails (a precision value of 1, where precision is $1/sd^2$). Thus, I am *a priori* assuming no effect of each predictor variable, but am not placing much weight in this assumption. The prior for β_0 is a normal distribution with a mean of the log of the mean from the y data. The precision around this estimate is also 1, which again, on the log scale, represents wide tails, and substantial uncertainty in the prior value.

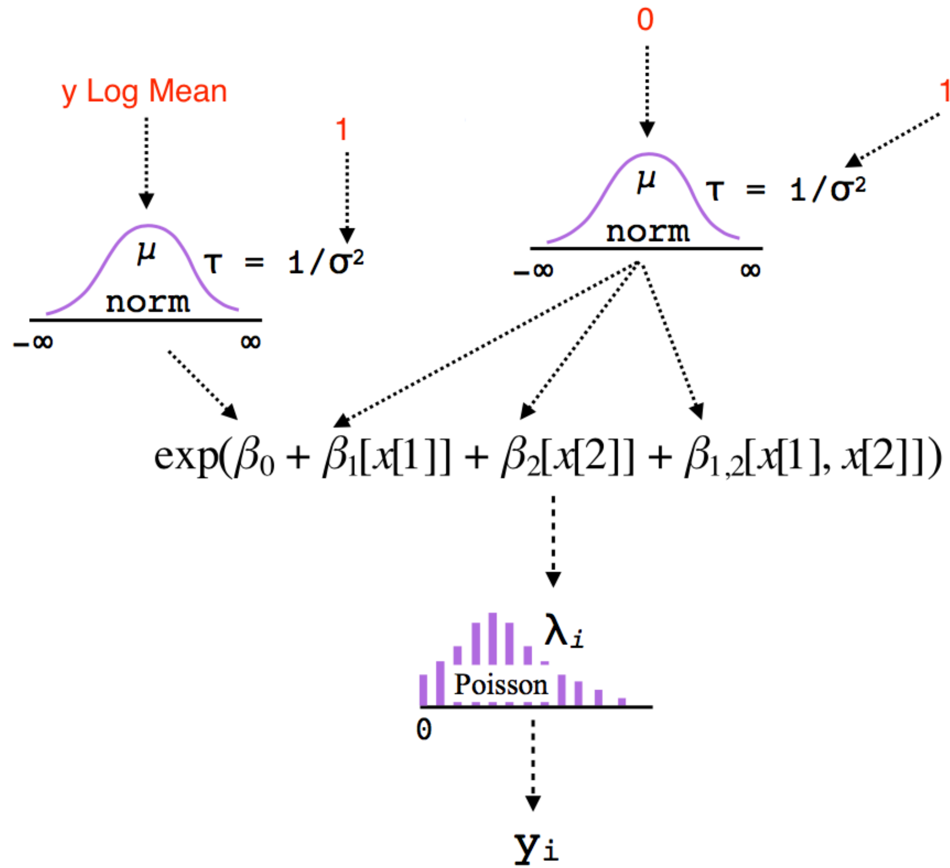


Figure 1: Diagram representation of model used for Bayesian alternative to a contingency test.

Appendix 2: Bayesian model for inheritance test

For analyzing inheritance patterns, the data was converted into a binary format, where the data were coded as 1 if the offspring was homozygous and 0 if the offspring was heterozygous. The Bayesian model then had a Bernoulli predictor variable, which is appropriate when outcomes can be either 0 or 1. Only one parameter, θ, is needed for estimation of the Bernoulli distribution, which represents the probability of a "success" (a value of 1). Given the way the pedigrees were organized, and strict Mendelian inheritance, the estimate of θ should be 0.5 (i.e., offspring have a 50% of being homozygous or heterozygous given informative parental genotypes where one parent is homozygous and the other parent is heterozygous, with one allele shared with the other parent). The prior distribution that was used for θ was a beta distribution with the two shape parameters both being 1. This represents an "uninformative" prior because it results in a flat distribution, with all values between 0 and 1 being equally likely.

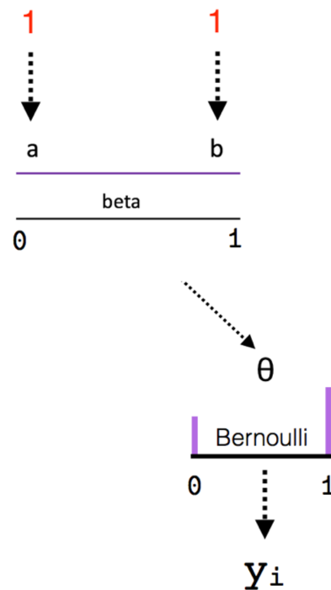


Figure 1: Diagram representation of the Bayesian model used to test for biased inheritance patterns.

References

Aagaard JE, Springer SA, Soelberg SD, Swanson WJ (2013) Duplicate Abalone Egg Coat Proteins Bind Sperm Lysin Similarly, but Evolve Oppositely, Consistent with Molecular Mimicry at Fertilization. *PLoS Genetics*, **9**, e1003287–11.

Aagaard JE, Vacquier VD, MacCoss MJ, Swanson WJ (2009) ZP Domain Proteins in the Abalone Egg Coat Include a Paralog of VERL under Positive Selection That Binds Lysin and 18-kDa Sperm Proteins. *Molecular Biology and Evolution*, **27**, 193–203.

Amaral AR, Moller LM, Beheregaray LB, Coelho MM (2011) Evolution of 2 Reproductive Proteins, ZP3 and PKDREJ, in Cetaceans. *Journal of Heredity*, **102**, 275–282.

Antalíková J, Jankovičová J, Simon M *et al.* (2015) Localization of CD9 Molecule on Bull Spermatozoa: Its Involvement in the Sperm-Egg Interaction. *Reproduction in Domestic Animals*, **50**, 423–430.

Bensch S, Andrén H, Hansson B *et al.* (2006) Selection for heterozygosity gives hope to a wild population of inbred wolves. *PLOS ONE*, **1**, e72.

Berger JO, Berry DA (1988) Statistical analysis and the illusion of objectivity. *American Scientist*, **76**, 159–165.

Best PB, Brandao A, Butterworth D (2001) Demographic parameters of southern right whales off South Africa. *Journal Cetacean Resource Management (Special Issue)*, 161–169.

Best PB, Reeb D, Rew MB *et al.* (2005) Biopsying southern right whales: their reactions and effects on reproduction. *Journal of Wildlife Management*, **69**, 1171–1180.

Bianchi E, Doe B, Goulding D, Wright GJ (2014) Juno is the egg Izumo receptor and is essential for mammalian fertilization. *Nature*, **508**, 483–487.

Birkhead TR, Pizzari T (2002) Postcopulatory Sexual Selection. *Nature Reviews Genetics*, **3**, 262–273.

Brown MW, Kraus SD, Gaskin DE (1991) Reaction of North Atlantic right whales (Eubalaena glacialis) to skin biopsy sampling for genetic and pollutant analysis. *Rep. Int. Whal. Commn. (Special Issue)*, **13**.

Clapham PJ, Young SB, Brownell RL (1999) Baleen whales: conservation issues and the status of the most endangered populations. *Mammal review*, **29**, 37–62.

Du J, Dungan SZ, Sabouhanian A, Chang BSW (2014) Selection on synonymous codons in mammalian rhodopsins: a possible role in optimizing translational processes. *BMC evolutionary biology*, **14**, 96.

Dziminski MA, Roberts JD, Simmons LW (2008) Fitness consequences of parental compatibility in the frog Crinia georgiana. *Evolution*, **62**, 879–886.

Firman RC, Simmons LW (2008) Polyandry facilitates postcopulatory inbreeding avoidance in house mice. *Evolution*, **62**, 603–611.

Foerster K, Delhey K, Johnsen A, Lifjeld JT (2003) Females increase offspring heterozygosity and fitness through extra-pair matings. *Nature*, **425**, 714–717.

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, Hamilton PK, Brown MW *et al.* (2007a) Patterns of male reproductive success in a highly promiscuous whale species: the endangered North Atlantic right whale. *Molecular Ecology*, **16**, 5277–5293.

Frasier TR, McLeod BA, Gillett RM, Brown MW, White BN (2007b) Right whale past and present as revealed by their genes. In: *The urban whale north atlantic right whales at the crossroads* (eds Kraus SD, Rolland RM)

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). *Molecular Ecology Notes*, **6**, 1025–1029.

Fujiwara M, Caswell H (2001) Demography of the endangered North Atlantic right whale. *Nature*, **414**, 537–541.

Ginsberg JR, Huck UW (1989) Sperm competition in mammals. *Trends in Ecology & Evolution*, **4**, 74–79.

Haig D, Ubeda F, Patten MM (2014) Specialists and Generalists: The Sexual Ecology of the Genome. *Cold Spring Harbor Perspectives in Biology*, **6**, a017525–a017525.

Halsey LG, Curran-Everett D, Vowler SL, Drummond GB (2015) The fickle P value generates irreproducible results. *Nature methods*, **12**, 179–185.

Hamilton PK, Marx MK, Kraus SD (1995) Weaning in North Atlantic right whales. *Marine Mammal Science*, **11**, 386–390.

Hamm D, Mautz BS, Wolfner MF, Aquadro CF, Swanson WJ (2007) Evidence of Amino Acid Diversity Enhancing Selection within Humans and among Primates at the Candidate Sperm-Receptor Gene PKDREJ. *The American Journal of Human Genetics*, **81**, 44–52.

Hughes J, Ward CJ, Aspinwall R, Butler R, Harris PC (1999) Identification of a human homologue of the sea urchin receptor for egg jelly: a polycystic kidney disease-like protein. *Human Molecular Genetics*, **8**, 543–549.

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.

Jeanmougin F, Thompson JD, Gouy M, Higgins DG, Gibson TJ (1998) Multiple sequence alignment with Clustal X. *Trends in Biochem Science*, **10**, 403–405.

Kempenaers B, Adriaensen F, Van Noordwijk AJ, Dhondt AA (1996) Genetic Similarity, Inbreeding and Hatching Failure in Blue Tits: Are Unhatched Eggs Infertile? *Proceedings of the Royal Society B: Biological Sciences*, **263**, 179–185.

Kosman ET, Levitan DR (2014) Sperm competition and the evolution of gametic compatibility in externally fertilizing taxa. *Molecular human reproduction*, **20**, 1190–1197.

Kraus SD (1990) Rates and Potential Causes of Mortality in North Atlantic Right Whales (*Eubalaena Glacialis*). *Marine Mammal Science*, **6**, 278–291.

Kraus SD, Rolland RM (Eds.) (2007) *The Urban Whale*. Harvard University Press.

Kraus SD, Brown MW, Caswell H *et al.* (2005) North Atlantic Right Whales in Crisis. *Science*, **309**, 561–562.

Kraus SD, Kenney RD, Mayo CA *et al.* (2016) Recent Scientific Publications Cast Doubt on North Atlantic Right Whale Future. *Frontiers in Marine Science*, **3**, 1401–3.

Kraus SD, Moore KE, Price CA *et al.* (1986) The use of photographs to identify individual North Atlantic right whales (*Eubalaena glacialis*)., **Special Issue 10**, 145–151.

Kruschke J (2014) *Doing Bayesian Data Analysis, Second Edition: A Tutorial with R, JAGS, and Stan*. Academic Press.

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.

Mate B, Duley P, Lagerquist B (2005) Observations of a female North Atlantic right whale (Eubalaena glacialis) in simultaneous copulation with two males: supporting evidence for sperm competition. *Aquatic Mammals*.

Mays Jr HL, Hill GE (2004) Choosing mates: good genes versus genes that are a good fit. *Trends in Ecology & Evolution*, **19**, 554–559.

Noren DP, Mocklin JA (2011) Review of cetacean biopsy techniques: Factors contributing to successful sample collection and physiological and behavioral impacts. *Marine Mammal Science*, **28**, 154–199.

Mouchiroud D, Gautier C, Bernardi G (1995) Frequencies of synonymous substitutions in mammals are gene-specific and correlated with frequencies of nonsynonymous substitutions. *Journal of Molecular Evolution*, **40**, 107-113.

Olsson M, Shine R, Madsen T, Gullberg A, Tegelstrom H (1996) Sperm selection by females. *Nature*, **383**, 585.

Palumbi SR (1999) All males are not created equal: fertility differences depend on gamete recognition polymorphisms in sea urchins. *Proceedings of the National Academy of Sciences*, **96**, 12632–12637.

Parks S (2003) Response of North Atlantic right whales (*Eubalaena glacialis*) to playback of calls recorded from surface active groups in both the North and South Atlantic. *Marine Mammal Science*, **19**, 563-580.

Payne R, Brazier O, Dorsey EM *et al.* (1983) External features in southern right whales (*Eubalaena australis*) and their use in identifying individuals. In: *Communication and behavior of whales* (ed Payne R), pp. 371–445. Boulder, CO.

Perry SL, DeMaster DP, Silber GK (1999) The Great Whales: History and Status of Six Species Listed as Endangered Under the U.S. Endangered Species Act of 1973. *Marine Fisheries Review*, **61**, 1–74.

Preston BT, Stevenson IR, Pemberton JM, Coltman DW, Wilson K (2003) Overt and covert competition in a promiscuous mammal: the importance of weaponry and testes size to male reproductive success. *Proceedings of the Royal Society B: Biological Sciences*, **270**, 633–640.

Primakoff P, Myles DG (2002) Penetration, adhesion, and fusion in mammalian sperm-egg interaction. *Science*, **296**, 2183–2185.

Pujolar JM, Pogson GH (2011) Positive Darwinian selection in gamete recognition proteins of Strongylocentrotus sea urchins. *Molecular Ecology*, **20**, 4968–4982.

Sutton KA, Jungnickel MK, Florman HM (2008) A polycystin-1 controls postcopulatory reproductive selection in mice. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 8661–8666.

Swanson WJ, Vacquier VD (2002) The Rapid Evolution of Reproductive Proteins. *Nature Reviews Genetics*, **3**, 137–144.

Thompson J, Gibson T, Higgins D (2002) Multiple sequence alignment using ClustalW and ClustalX. *Current Protocols in Bioinformatics*, **2**.

Tregenza T, Wedell N (2000) Genetic compatibility, mate choice and patterns of parentage: invited review. *Molecular Ecology*, **9**, 1013–1027.

Turner LM, Hoekstra HE (2008) Causes and consequences of the evolution of reproductive proteins. *The International Journal of Developmental Biology*, **52**, 769–780.

Vacquier VD, Swanson WJ (2011) Selection in the Rapid Evolution of Gamete Recognition Proteins in Marine Invertebrates. *Cold Spring Harbor Perspectives in Biology*, **3**, a002931–a002931.

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.

Wassarman PM, Jovine L, Litscher ES, Qi H, Williams Z (2004) Egg sperm interactions at fertilization in mammals. *European Journal of Obstetrics & Gynecology and Reproductive Biology*, **115**, S57–S60.

Wilburn DB, Swanson WJ (2016) From molecules to mating: Rapid evolution and biochemical studies of reproductive proteins. *Journal of proteomics*, **135**, 12-25.

Chapter 6:

Conclusion and Future Work

1. Main Results

To examine reproductive success and inheritance patterns in the North Atlantic right whale, a total of nine candidate genes were selected to analyze based on the role they play in fertilization, and are expressed on the surface of either the sperm or egg. These candidate genes were sequenced at their putative functional sites. For these nine candidate genes, sixteen primer sets were developed, and only 2 loci had variable sites, a single nucleotide polymorphism (SNP) - one at Cd9a and two at PKDREJ-S1b. Cd9a is expressed on the egg, and PKDREJ-S1b on the sperm. For Cd9a, the variable site had a G or an A, which is a non-synonymous substitution resulting in a different amino acid. In PKDREJ-S1b, both variable sites had either a T or a C, and were both synonymous substitutions resulting in the same amino acid. Statistical analyses for biases in mate choice showed a slight bias within loci for mates with the same genotype. However, when looking between loci, more mating pairs occurred than were expected when one parent was homozygous at one locus and the other was heterozygous at another. However, these differences were small, suggesting that the role of these genes in mate choice, if any, is likely small.

Inheritance pattern analyses to test for bias from cryptic female choice showed an interesting pattern. Even though a previous study using different markers found that offspring were more likely to inherit a paternal allele that differed from the maternal (Frasier *et al.* 2013), offspring were more likely to inherit an allele that was the same as the maternal allele more often than expected based on Mendelian inheritance, for the second site of locus PKDREJ-S1b. The different pattern observed here than in the previous studies could be due to the different stages of gamete compatibility. Specifically, there are two key

stages at which the sperm interacts with the egg: the first is when the capacitated sperm interacts with the ZP of the egg, and the second is when the acrosome-reacted sperm interact with and bind to the plasma membrane of the egg. The first interaction is that which is most often involved in species recognition, reducing the chances of cross-species fertilization. PKDREJ-S1b is thought to be expressed on capacitated sperm, and would therefore interact with the ZP of the egg. Thus, it makes sense that there could be a bias for similar alleles, because at this stage similarity is "good" in terms of species-specificity. At the second interaction stage, however, it may be beneficial for fertilization to occur between dissimilar alleles at interacting genes, to reduce the impacts of inbreeding. Therefore, although the patterns observed is different than that expected, and than that found with the microsatellite loci, it seems likely that there is a functional reason that makes intuitive sense when placed in the proper context of which genes are interacting when.

These genes do seem to have an impact on cryptic female choice in this species, but further research is needed to say whether similar effects will be seen in other cetacean species. They also may play a role in mate choice, but again with the limited sample size, more research is needed to be able to say anything conclusively. The effect of these genes on individual's reproductive success is still unknown. There are multiple factors that affect reproduction in the North Atlantic right whale such as pollutants, limited food supply and carrying capacity, disease, marine biotoxins, and habitat loss (Kraus & Rolland 2007). Thus, basing an individual's reproductive success on only their genetics would be oversimplifying reproduction in this population.

2. Going Forward

With only 152 individuals, and 69 mother-father-calf triads, the sample size was limiting, and also budget. Since only 9 individuals were originally sequenced to look for variable sites, there are potentially missed variable sites that were not detected. Samples are also continuously being collected, so the sample size is increasing with each year. When more up-to-date parentage information is available, it will allow for more triads to be sequenced to increase sample size. Due to the complicated nature of fertilization, it is also possible that there are still unidentified genes that play a role in this process that were not analyzed, so future research can add to this. In addition, since both mutations at PKDREJ-S1b were synonymous, but the second site still showed biased distribution patterns in what sex had the variable site, and also in inheritance patterns, sequencing around this site may reveal the actual functional site from which the effects are being seen.

It may be beneficial to use a genome wide association study (GWAS) approach in this population since using a candidate gene approach was fairly unsuccessful in this population. GWAS correlate allele frequencies at each of several hundred thousand markers throughout the genome with the desired trait. If enough variable markers are spread throughout the genome, then some should be in linkage disequilibrium with genes influencing the desired trait, and will show a correlation. Unlike candidate gene studies, GWAS do not require previous knowledge of the trait, and could potentially identify still unknown genes that may be linked to reproductive traits in this population (Stranger *et* al. 2011). Also, with the development of the new lab protocols for studying these genes in cetaceans, it may be worth analyzing these genes in similar cetacean species that are also

struggling to recover despite protection, such as Bowhead Whales (*Balaena mysticetus*) of the Okhotsk Sea and various eastern Arctic populations, western Gray Whales (*Eschrichtius robustus*), and Blue Whale (*Balaenoptera musculus*) populations (Clapham *et al.* 1999).

Since the mechanism(s) that drives the biased fertilization patterns seen in multiple species throughout nature is not known, identifying this mechanism and the genes that cause these patterns remains a priority (Olsson *et al.* 1996; Kempenaers *et al.* 1996; Tregenza & Wedell 2000; Foerster *et al.* 2003; Bensch *et al.* 2006; Firman & Simmons 2008; Dziminski *et al.* 2008). Further research into genes that influence fertilization can add to the list of candidate genes for study, and also could identify receptor-ligand pairs that are currently unknown.

Identifying the mechanism(s) behind biased fertilization patterns can help increase the understanding of how gamete compatibility can influence reproductive success in small populations, and also on still unidentified processes involved in reproductive success, speciation, and the maintenance of genetic variation in small populations. This study only represents one small step in the understanding of gamete compatibility in mammals. This field is in its infancy, and therefore much further research is required to help clarify the understanding of this mechanism in mammals.

References

Amaral AR, Moller LM, Beheregaray LB, Coelho MM (2011) Evolution of 2 Reproductive Proteins, ZP3 and PKDREJ, in Cetaceans. *Journal of Heredity*, **102**, 275–282.

Bensch S, Andrén H, Hansson B *et al.* (2006) Selection for heterozygosity gives hope to a wild population of inbred wolves. *PLOS ONE*, **1**, e72.

Clapham PJ, Young SB, Brownell RL (1999) Baleen whales: conservation issues and the status of the most endangered populations. *Mammal review*, **29**, 37–62.

Dziminski MA, Roberts JD, Simmons LW (2008) Fitness consequences of parental compatibility in the frog Crinia georgiana. *Evolution*, **62**, 879–886.

Firman RC, Simmons LW (2008) Polyandry facilitates postcopulatory inbreeding avoidance in house mice. *Evolution*, **62**, 603–611.

Foerster K, Delhey K, Johnsen A, Lifjeld JT (2003) Females increase offspring heterozygosity and fitness through extra-pair matings. *Nature*, **425**, 714–717.

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.

Kempenaers B, Adriaensen F, Van Noordwijk AJ, Dhondt AA (1996) Genetic Similarity, Inbreeding and Hatching Failure in Blue Tits: Are Unhatched Eggs Infertile? *Proceedings of the Royal Society B: Biological Sciences*, **263**, 179–185.

Olsson M, Shine R, Madsen T, Gullberg A, Tegelstrom H (1996) Sperm selection by females. *Nature*, **383**, 585.

Stranger BE, Stahl EA, Raj T (2011) Progress and promise of genome-wide association studies for human complex trait genetics. *Genetics*, **187**, 367–383.

Tregenza T, Wedell N (2000) Genetic compatibility, mate choice and patterns of parentage: invited review. *Molecular Ecology*, **9**, 1013–1027.