Proteomic identification and evolutionary analysis of primate reproductive proteins

Katrina G. Claw

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington
2013

Reading Committee:
Willie J. Swanson, Chair
Michael J. MacCoss, Chair
Deborah A. Nickerson

Program Authorized to Offer Degree:
Genome Sciences

University of Washington

Abstract

Proteomic identification and evolutionary analysis of primate reproductive proteins

Katrina G. Claw

Chair of the Supervisory Committee:
Associate Professor Willie J. Swanson and Associate Professor Michael J. MacCoss
Genome Sciences

Sex and reproduction have long been recognized as drivers of distinct evolutionary phenotypes. Studying the evolution and molecular variation of reproductive proteins can provide insights into how primates have evolved and adapted due to sexual pressures. In this dissertation, I explore the long-term evolution of reproductive proteins in human and non-human primates. I first describe the evolutionary diversification of sperm and eggs, and what drives them to diverge. I then describe rapidly evolving proteins in the egg and in sperm-egg interactions. I then present the use of a unique combination of genomic and proteomic technologies to study the evolution of seminal fluid proteins. With proteomics, I identify and quantify the abundance of a large proportion of uncharacterized seminal fluid proteins from 8 primate species with diverse mating systems. Using evolutionary analyses, I find rapidly evolving seminal fluid proteins and candidates genes with evolutionary rates and protein abundances that are correlated with mating system variation. I then explore the phylogentic relationships between putatively coevolving sperm-egg fusion genes. I find evidence that two pairs of sperm-egg fusion genes have correlated evolutionary rates along primate phylogenetic branches, which may indicate that they are genetically interacting. I conclude by discussing how to disentangle the selective pressures causing reproductive protein divergence, the value of using proteomics in comparative evolutionary studies, and future directions for identifying interacting fertilization proteins.

Table of Contents

List of Figures

List of Tables

Acknowledgements

So many people have contributed to completion of this dissertation.

First and foremost, I would like to thank my parents, George and Rose Claw. Their unwavering support and encouragement has helped me so much over the years.

Thank you to my partner in crime, Lyle Smith, for *everything*.

I would like to offer heartfelt thanks to my entire Claw and Benally family: my sister Gena, brothers Geoff and Doog, aunts, uncles, cousins, and to my late grandmothers, Alice Benally and Mary Claw. Thank you also to the Smith family who has always been there for me.

I would like to thank my mentors, past and present. I would especially like to thank my adviser, Willie. His support, encouragement, and quiet confidence in me for all my endeavors, whether research, travel, or outreach, has made me a better scientist. Thank you to the members of my committee (Michael MacCoss, Debbie Nickerson, Stan Fields, Evan Eichler, and Wylie Burke) for advice and guidance over the years.

Thank you to everyone below!

To former and current Swanson lab members, my fellow grad students, and colleagues: Melody Palmer, Jan Aagaard, Jennifer McCreight, Renee George, Joe Gasper, Steve Springer, and Geoff Findlay.

To the SACNAS Chapter at UW: You're such an awesome organization with awesome people! It was an honor to be part of a group of scientists who believe in outreach, mentoring, and serving the community.

To my Seattle, Arizona, and World friends: Thank you for your laughs and company, with a special shout out to A-team and the Saturday lunch crew.

To the Seattle Clear Sky Native Youth Council and the Seattle Native community: Thank you for sharing your stories and community with me.

To the Many Farms Chapter and the many Native American organizations who have sponsored my education over the years.

Dedication

For my parents, George and Rose Claw.
Ahéhee' shimá dóó shizhé'é. Nina'nitin baa ahééh nisin.

Chapter 1: Rapid Evolution in Eggs and Sperm

Introduction

The experience of learning about the birds and the bees is almost universally awkward, bringing to mind textbook cartoons of the male and female reproductive tracts viewed with amazement, confusion or embarrassment. Although most people do not contemplate reproductive systems every day, speaking of sex and reproduction only in secretive tones, like it or not, we are surrounded by sperm and eggs. On a typical day, you might enjoy a delectable breakfast of fried eggs (a chicken egg and the accompanying yolk) and take a walk outside that triggers a sneeze caused by pollen (plant sperm). You might decide to take a dip in the ocean, where external fertilizers such as sea urchins and abalone are spawning their eggs and sperm throughout the water.

Sexual reproduction is ubiquitous across plants and animals—more than 90 percent of vertebrates reproduce sexually. Sex is largely responsible for the biological diversity that first fascinated and continues to occupy biologists. Differences in shape, size and coloration, among other characteristics, are often attributed to the evolutionary influences of sex. But why did sex evolve in the first place? Observing sperm and egg diversity across species brings more questions to mind: Why are there more sperm than eggs? How does sexual promiscuity affect the evolution of the sperm and egg? One might think that the same mechanisms, sexual traits and reproductive genes are involved in sperm–egg composition and interactions. But is this the case?

Diversity of Sperm and Egg

The sperm and the egg vary in looks and structure. Each sperm or egg cell contains exactly half of the genetic information needed to make an individual. In humans, that amounts to

23 chromosomes, one of which is the sex chromosome (X chromosome in eggs and either X or Y chromosome in sperm). The other 22 chromosomes that do not contribute to sex determination are called autosomes. In the common fruit fly, Drosophila melanogaster, sperm and egg cells have three autosomes and an X or Y sex chromosome pair. Unlike in humans, the ratio of X chromosomes to autosomes determines sex in the fruit fly. Figure 1 illustrates the astounding morphological diversity of sperm and egg from various organisms spanning plants, mammals and invertebrates.

An important characteristic that may influence evolutionary dynamics of such diversity is whether an egg is fertilized inside or outside the female's body. Internal fertilization occurs inside the body and requires some type of insemination by the male, as in many mammals including humans. External fertilization, or free spawning, refers to species in which the sperm and eggs are released into the open environment, as in many marine invertebrates and some plants. Because external fertilization exposes the sperm and egg to a potentially hazardous environment without parental protection, these gametes often develop unique features for their protection and dispersal. For example, because plant pollen is immobile, some plant species have evolved pollen spines for better dispersal and for sticking to fertile flower parts (Figure 1, middle row on far left). Dispersal, or pollination, in plants can occur by animals, wind or water, and the wide variety of pollen morphology reflects the varying pressures to adapt to the surrounding environment.

Sperm and egg also differ dramatically in size and number. Typically, the female invests more into making an egg and will only release a few hundred in her lifetime compared to the ease and abundance of male sperm. Not only does the egg have multiple barriers to avoid polyspermy (multiple sperm entering the egg), it also contains nutritious resources for the new

zygote once fertilization has occurred. Because of this, eggs are often larger than sperm. The diameter of a human egg is about the width of a human hair, and a sperm is one-twentieth that size. But sheer numbers make up for what sperm lack in size. A human male will ejaculate about 200 to 500 million sperm. A man ejaculating 20 to 50 million sperm is considered sterile. Human sperm count is modest in the animal kingdom. By contrast, the common farm pig will expel 60 billion sperm in a single emission!

The size and shape of sperm vary between species. One of the biggest sperm is found in the puny common fruit fly. Researchers in Scott Pitnick's lab at Syracuse University showed that the common fruit fly's sperm could be measured with a ruler at a whopping two millimeters. But the winner in this sperm size contest goes to a related species of fruit fly, *Drosophila bifurca*, with a sperm length of 60 millimeters (Figure 2). The *D. bifurca* sperm is 1,000 times longer than a human sperm (1). How can such disparate forms of sperm and egg have evolved to perform a well-established function?

Various proteins, encoded by genes, make up sperm and eggs. Changes in these proteins over time may result in a single population becoming two separate, distinct species. In 2002, Willie Swanson and Victor Vacquier, respectively of University of Washington and the Scripps Institute of Oceanography, reviewed the many studies that have established that reproductive genes are some of the most rapidly evolving genes in the human genome, along with genes involved in immune response. Proteins are made up of amino acids, and changes in the nucleotide bases that make up genes can change the recipe of amino acids for a protein. Rapidly evolving genes have extremely high rates of amino acid change. As genome sequencing has become more affordable, this general trend is seen in a wide variety of other sequenced species, such as primates, rodents, fruit flies and butterflies. Swanson's lab, which I joined in 2009, has

shown that genes encoding for proteins found on the surfaces of the sperm and egg (potentially involved in sperm– egg interactions) are rapidly evolving in the marine invertebrate abalone (of the genus *Haliotis*). Our lab also found an important protein involved in forming the egg coat in primates to be rapidly evolving, and other studies have found a variety of other rapidly evolving sperm–egg proteins in different species.

In particular, researchers in the Vacquier and Swanson labs are using the abalone system to study species specificity of sperm and egg binding. Historically, abalone species lived in overlapping boundaries across the coast of the western United States. They provide an ideal system for studying reproduction and species specificity, because even though species of abalone have overlapping ranges and spawning times, there is very little hybridization between species. This fact indicates that there must be some type of species specificity happening at the sperm–egg level. Researchers can use a conservative test for rapid evolution to study changes in sperm and egg genes (Figure 3). Indeed, we see that rapid changes have led to the extreme diversification of reproductive genes, even within closely related species of abalone (Figure 4). Why are reproductive genes so diverse, in some cases evolving much faster than immunity and defense genes that are under constant pressure because of microbial attacks?

Every sexually reproducing organism has its own unique strategy for attracting mates of the opposite sex. Mating strategies may involve alluring chemicals emitted by one sex, extravagant coloration, extreme size, sexy vocalizations and many other oddities. The male bowerbird works hard to build a colorful nest of rocks, flowers and sticks to entice females. And just like its namesake, the male peacock spider of Australia raises his elaborately colored flaps and dances to attract female onlookers. The female porcupine, however, prefers a shower of urine from male porcupine suitors before choosing a mate. These bizarre strategies have evolved

over time in a complex balance that includes surviving to adulthood (time to mate!) and being attractive to potential mate(s) to pass on genes to the next generation.

A trait beneficial for male attractiveness may not necessarily be beneficial for male survival. For example, the male peacock has a colorful display of feathers irresistible to female peacocks, but which may attract undue attention from predators. The unwieldy tail might cause its demise. To boot, males and females may not have the same interests. Even if it is in the female's interest to pick males that look like they have the best sperm, it is not necessarily in the males' interest to look the part. Nevertheless, the process of sexual reproduction passes both sexes' genes to their offspring indiscriminately. This could create a situation in which there is conflict between male and female reproductive genes. Sexual conflict happens when males and females are in an evolutionary arms race for optimal reproductive potential. Because two parties are involved, what is good for one may not be good for the other. For example, males may be evolving sperm that are better and faster at fusing with the egg than all the other sperm, whereas females may be counteracting these more efficient sperm by evolving eggs that resist quick fertilization to prevent polyspermy.

From the standpoint of passing on the most and best genes, is it better to be a promiscuous Casanova or Cleopatra, or, as in Jane Austen's Pride and Prejudice, a faithful Mr. Darcy or Elizabeth Bennet? In promiscuous species, females mate with many partners during their estrous, or fertile, period. In monogamous species, females mate with one partner during their estrous period—so serial monogamists are included. Depending on a variety of factors, such as availability of resources, promiscuity or monogamy could be adaptive for a species. In highly promiscuous systems, males are free to sow their wild oats with many different females, giving them high reproductive potential. Although females don't have as high of a reproductive

potential as males, because they are limited by high female investment in eggs and young, they can maximize the continued proliferation of their genes by choosing the "perfect" mate(s) to fertilize their egg(s).

Because of their diverse mating systems, primates provide an excellent system to study how mating system differences have affected reproductive trait and protein evolution (See Figure 5). Even between closely related species, such as humans and bonobos, mating systems differ dramatically. Humans are mostly monogamous, and bonobos are some of the most promiscuous primates— a female bonobo may mate up to 50 times with different males during a single estrous period. Many studies have shown that in more promiscuous mating systems, size differences between the sexes, testes size in males, complex genital morphology and other morphological, behavioral and molecular traits are exaggerated. Richard Prum and colleagues at Yale University have shown that the complex maze of the female Pekin duck's vagina may have evolved so that she could confound sperm duds and select only the best of the best sperm from males she has mated with. The corkscrew-shaped male Pekin duck's penis enables him to penetrate the female, navigate the vaginal maze, and deposit his sperm in an advantageous location (See Figure 6). Work in Prum's lab demonstrated that nonoptimal spirals in a duck's penis have been shown to be less compatible with a female duck's vagina (2). Thus, promiscuity and competition between males influence the evolution of complex sexual structures.

Why is there so much diversity in reproductive traits and genes? Sexual selection favors genes that give a reproductive advantage, and these genes can increase over time in a given population. Sexual selection should not be confused with Charles Darwin's famous theory of natural selection, which emphasizes the survival of the fittest. In our peacock example, natural selection would not favor the male peacock tail, because it decreases survival by increasing

predation, but sexual selection through female preference favors an otherwise unfavorable trait. Although Darwin briefly discussed sexual selection in *On the Origin of Species*, modern interpretations of sexual selection group it into two parts: precopulatory sexual selection (before sex or mating occurs) and postcopulatory sexual selection (after sex occurs). Precopulatory sexual selection can occur within a sex (or between males) or between sexes (between males and females, see Figure 7). An example of within-sex competition is when males combat over access to females, as occurs in gorillas and big-horned sheep. The competition between males can increase the occurrence of "weaponry" in those males, such as antlers, horns, strength and body size. Precopulatory sexual selection between the sexes can occur when the female chooses her mate based on certain male qualities that are either behavioral or decorative. Behavior might include mating songs or dances, and decorative features include coloration or shape changes. For example, the male sage grouse has sacs on his chest that he inflates during the mating season to attract females. Change in such morphological features is driven by molecular changes within genes.

Why are we interested in these differences in mating strategies and reproductive traits and genes? The short answer is that many of these changes may have led to the evolution of new species in the past or could lead to the evolution of new species in the future (3). Studying these changes will lead to better understanding of the oddities seen in nature, such as outrageously long sperm and spiral-shaped penises. Many of the rapidly evolving reproductive genes are involved in more direct reproductive functions, such as sperm–egg binding, semen coagulation and fertilization. Although we do not know all the molecular mechanisms for sperm–egg interactions in every species, the puzzle pieces of how a sperm finds and fertilizes an egg are beginning to come together.

Sperm–Egg Interactions

Everyone learns in basic biology that the sperm has to reach the egg for fertilization to occur. It sounds simple enough. Among internal fertilizers, the male(s) will deposit semen into the female's reproductive tract, and the race to fertilize the egg begins. What about in external fertilizers where the sperm and eggs are released into the open environment? In both external and internal fertilizers, the egg releases chemicals to which sperm are attracted, and the millions or billions of sperm ejaculated or broadcasted follow this candy trail toward the egg. In a normal human male, it's quite surprising how many sperm are either dead or abnormal on ejaculation. Many evolutionary biologists have proposed that some sperm may have alternative functions during fertilization: Kamikaze sperm were thought to reduce the chances of rival sperm fertilizing the egg by blocking and killing them with enzymes. This notion has not panned out in subsequent studies in humans but is still an interesting hypothesis that might be relevant in other species. Simone Immler's lab at Uppsala University in Sweden showed that the hooked shape (Figure 1, middle of top row) of rodent sperm increased sperm swimming speed because the sperm hooked together to form clumps, which swam faster in a group than lone sperm. So maybe sperm work together for the good of the group.

The basic structures of sperm and eggs are similar across mammals and invertebrates (See Figure 8). Each sperm has a compartment called the acrosome that contains digestive enzymes. In mammals, sperm fully mature in the female reproductive tract, where certain molecules trigger them into full motility. Sperm are not alone on their journey; fructose-rich seminal fluid accompanies the sperm and provides a protective and nutritious environment in the female reproductive tract. Seminal fluid neutralizes the acidic vaginal environment to make it habitable for sperm. Seminal fluid proteins (SFPs) also have antimicrobial functions that may be

important for avoiding pathogens. In fruit flies, proteins in seminal fluid change female behavior after mating. In experiments conducted by Mariana Wolfner and her collaborators at Cornell University, specific seminal fluid proteins were "knocked down," or rendered inactive, in a group of females during mating. The females with unaltered semen deposited in their reproductive tracts lost interest in re-mating with other males and tended to lay more eggs than the females with "knocked-down" semen. This difference occurs because there are proteins in semen that facilitate egg laying and post-mating behavior in females, although the exact mechanisms remain unclear.

As Emily Martin of New York University has made clear in her study of language used to describe fertilization, many textbooks and articles describe the egg as passively awaiting the arrival of the sperm. But the egg has its own journey: Once mature, it will leave the ovary and travel into the fallopian tubes. The egg surrounds itself with multiple barriers (shown in Figure 9), including a collection of follicular cells called the cumulus oophorus; the glycogen- rich matrix of the egg envelope (or zona pellucida in mammals); and a plasma membrane enclosing the cytoplasm and nucleus of the egg (4,5). These barriers protect the contents of the egg but also slow down any sperm trying to penetrate the egg. When more than two sperm fuse with an egg, it results in cell death. Polyspermy, the fusion of multiple sperm with an egg, causes an unbalanced number of chromosomes, which eventually leads to disintegration of the cellular bodies. External fertilizers may have additional outer barriers to provide further protection from the environment.

The sperm and egg meet. Then what happens? The first and most well studied step in the fertilization process is the binding and passage of the sperm through the egg envelope. Many molecules on the sperm head and the egg's outer surface facilitate binding, but few of these

molecules have been identified. Abalone are used in many reproductive studies because as external fertilizers, their gametes are abundant and easily collected. In abalone, two interacting proteins have been identified on the sperm and egg, lysin and vitelline envelope receptor for lysin (VERL), respectively. The sperm protein lysin is found in the acrosome. Lysin is released when the sperm acrosomal contents are dispensed after binding to the egg in a process called the acrosome reaction. Lysin is then able to bind to VERL to create a hole through the egg envelope so that the sperm can pass through.

In other organisms, a variety of candidate proteins involved in binding the egg envelope have been proposed, but sperm proteins interacting with egg proteins only have been verified in abalone and sea urchins. Surprisingly, even in humans, no conclusive interacting sperm–egg proteins have been identified, despite the zona pellucida being well characterized. The proteins ZP3 and ZP2 are proposed to play an important role in binding sperm and are modified by parts of sugar molecules, which may be an important strategy for the egg to block nonoptimal sperm from binding. The sperm protein ZP3R/sp56 is proposed to bind with ZP3, but evidence for this binding remains contradictory. Protein complexes and parts of sugar molecules on the egg envelope may be involved in sperm–egg binding. In the next step to fertilization, the sperm fuses with the egg plasma membrane. This step does not seem as chemically specific as the sperm binding to the egg envelope. For example, mouse sperm are able to bind and fuse with human eggs when the zona pellucida is removed. (Of course, the fused mouse sperm and human egg are nonviable.) In humans, the egg protein CD9 is proposed to fuse with sperm protein Izumo. The interaction mechanisms between these proteins remain elusive, but both CD9 and Izumo show evidence of rapid evolution. Genes involved in sperm–egg fusion may be quite diverse between

taxa. Once the sperm and egg have fused, the resulting cell is called a zygote. The zygote undergoes rounds of mitosis and cellular specialization until an offspring is born.

Evolutionary Hypotheses

Studying rapid evolution brings into focus candidate genes that are changing quickly—change that may be important to successful reproduction or the rise of new species. Not all organisms can be brought into the lab and studied intensively. Detecting rapidly evolving sperm and egg proteins with DNA sequencing enables the study of organisms that cannot easily be included in experiments, such as humans and other long-lived organisms. As mentioned earlier, sexual selection influences egg and sperm diversity, as well as rapid evolution in reproductive proteins. In any given case of rapid evolution in a gene, natural selection and sexual selection may be acting alone or in concert with each other. Some hypothesized reasons for rapid evolution in reproductive genes include scenarios called sperm competition and sexual conflict (6).

Sperm are competitive. With millions of sperm trying to get to the egg first, there is no doubt that subtle changes in shape or energy stores will change a sperm's likelihood of successful fertilization. Adaptations to increase a male's chances of reproductive success are particularly important within promiscuous mating systems. Sometimes more is better. In some promiscuous species, males with more sperm and greater testes size have better chances at fertilizing multiple females. Males have evolved various tactics to reduce the chances of another male's sperm from fertilizing the egg, such as mate-guarding and copulatory plugs (7). One of the most promiscuous primate species is the bonobo. The bonobo has large testes size relative to body weight, and a thick copulatory plug will form from seminal fluids in the female's reproductive tract after mating. By contrast, the king-of-the-jungle gorilla has tiny testes and a

low sperm count, because the male gorilla has virtually no sperm competitors. After physically defending his territory and harem of females, the male gorilla will mate with all the females, under no pressure from other males.

As mentioned earlier, a more competitive sperm is not necessarily to the female's advantage. She may be looking for other qualities in a mate, or she may want to prevent polyspermy. Dissimilarities between the goals of the egg and those of the sperm may lead to sexual conflict. For example, the egg may be better off if the number of sperm that can enter is slowed through a mutation in an egg envelope protein. In such a situation, only sperm with a compatibly shaped protein would be able to bind and bypass the egg envelope. In turn, the sperm adapt to be more compatible, and thus competitive, with this new protein. We can think of this example as a lock and key, where the egg envelope is the lock and the sperm's interacting protein is the key (See Figure 10). By changing the shape of the lock, the egg eliminates the number of keys that fit (or sperm that bind), because all it requires is one sperm, and no more than that. In response to the lock change, sperm evolve to change the shape of the key to counteract the blocking strategy of the egg. All of these changes leave a mark on reproductive genes. Such interacting lock and key proteins should be changing rapidly. Research has shown that lysin and VERL are rapidly evolving in abalone populations, and evolutionary patterns that indicate sexual conflict may be occurring.

In addition to the hypotheses of sperm competition and sexual conflict, there are a variety of other hypotheses to explain the evolution of sexual traits and genes. In plants, the ability to recognize self from non-self plays an important role in fertilization, because self-fertilization will result in less diverse offspring than fertilization with pollen from another individual. Many genes with reproductive roles also have antibacterial and immune functions, which indicate that the

threat of microbial attack on the sperm or egg may be a major influence on rapid evolution during reproduction. Scenarios also exist where changes in a gene do not have a big effect, or where errors during DNA replication result in duplicating a gene on a chromosome, and as the twin genes change over generations, they specialize in their function. All or some of these pressures can act on a species' reproductive characteristics, and researchers use molecular evidence to disentangle the various hypotheses.

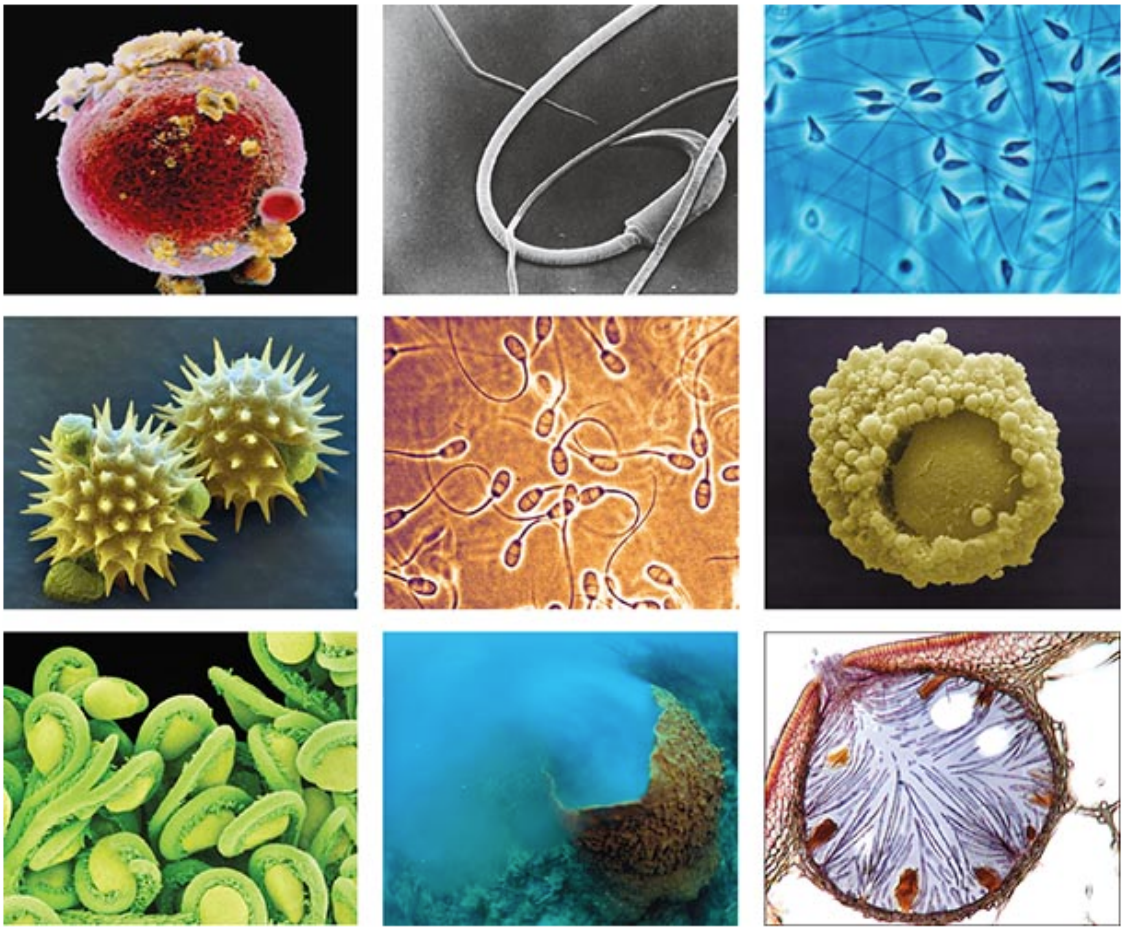


Figure 1. Sperm and egg morphological diversity. Sperm and eggs exhibit stunning morphological diversity across the plant, animal and fungal kingdoms. Clockwise from top left: False-color scanning electron microscope image (SEM) of a single human egg *(Homo sapiens)*; SEM of hamster sperm cell *(Cricetus* species); sea urchin sperm (Echinoidea); SEM of hamster egg *(Cricetus* species); light micrograph of conceptacle from female of the bladder wrack seaweed *(Fucus vesiculosus)*; male barrel sponge releasing sperm *(Xestospongia* species); SEM of ovules of cactus flower (Cactaceae); SEM of sunflower pollen *(Helianthus* species); and photomicrograph of rabbit sperm *(Lepus* species, *center)*.

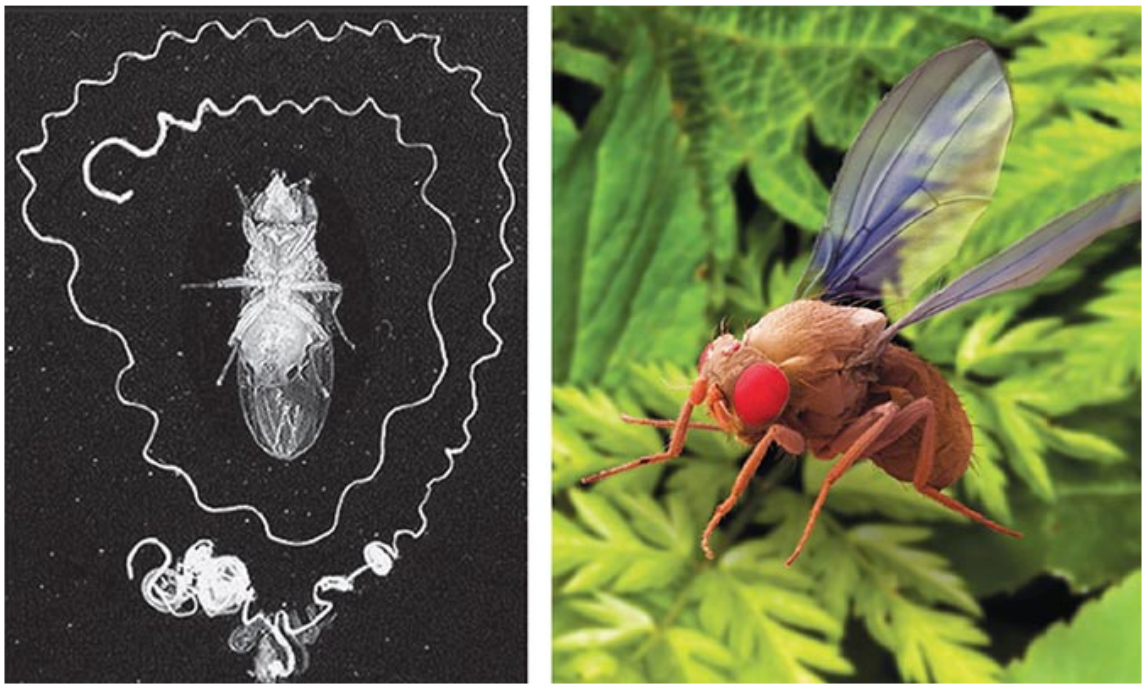


Figure 2. The fruit fly *Drosophila bifurca* lays claim to the superlative of biggest sperm, with a length of two inches. Shown here is a *D. bifurca* male encircled by the testis of another male *(left)*. This is even larger than the sperm of the common fruit fly *D. melanogaster,* pictured in a colored SEM at right, which measures two millimeters. (Image on left from T. Burkhead, *Promiscuity: An Evolutionary History of Sperm Competition,* Harvard University Press, 2002.)

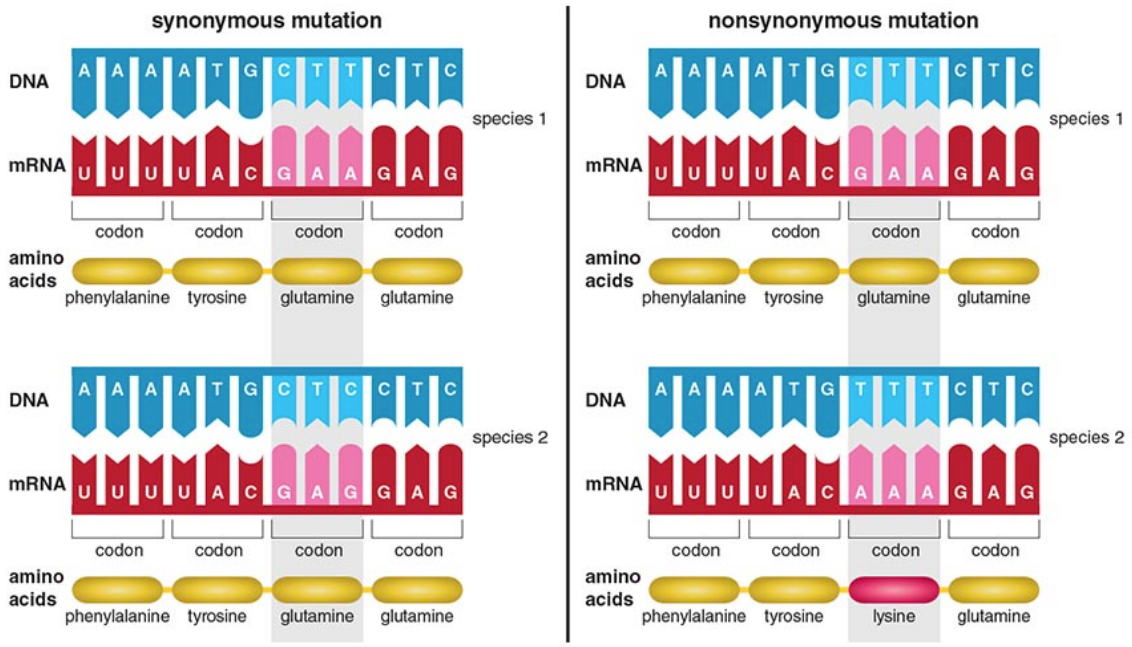


Figure 3. Detecting positive selection. Nucleotides, indicated here by the tabs labeled A, C, G, T and U are read in groups of three called codons. Messenger RNA (mRNA) makes a copy of the DNA and transports it to a cell's protein-making machinery. Each codon on the mRNA indicates an amino acid that will be added to the protein. Multiple codons can result in the same amino acid being coded. So if a copying mistake is made, and a nucleotide is changed, it does not necessarily mean that the amino acid (and thus the subsequent protein) will be changed. A nucleotide change that does not result in protein change is called a *synonymous mutation.* In the example on the left, the codon on the mRNA changes from GAA to GAG, but because these two codons both code for the amino acid glutamine, the protein remains the same. Conversely, a nucleotide change that does result in a change in the protein is called a *nonsynonymous mutation,* shown at right, where the codon GAA is changed to AAA, and thus the amino acid changes from glutamine to lysine, fundamentally changing the structure of the protein made. Rapid evolution is detected by comparing the number of synonymous changes to the number of nonsynonymous changes. If the ratio between these two is greater than one, the protein is rapidly evolving.

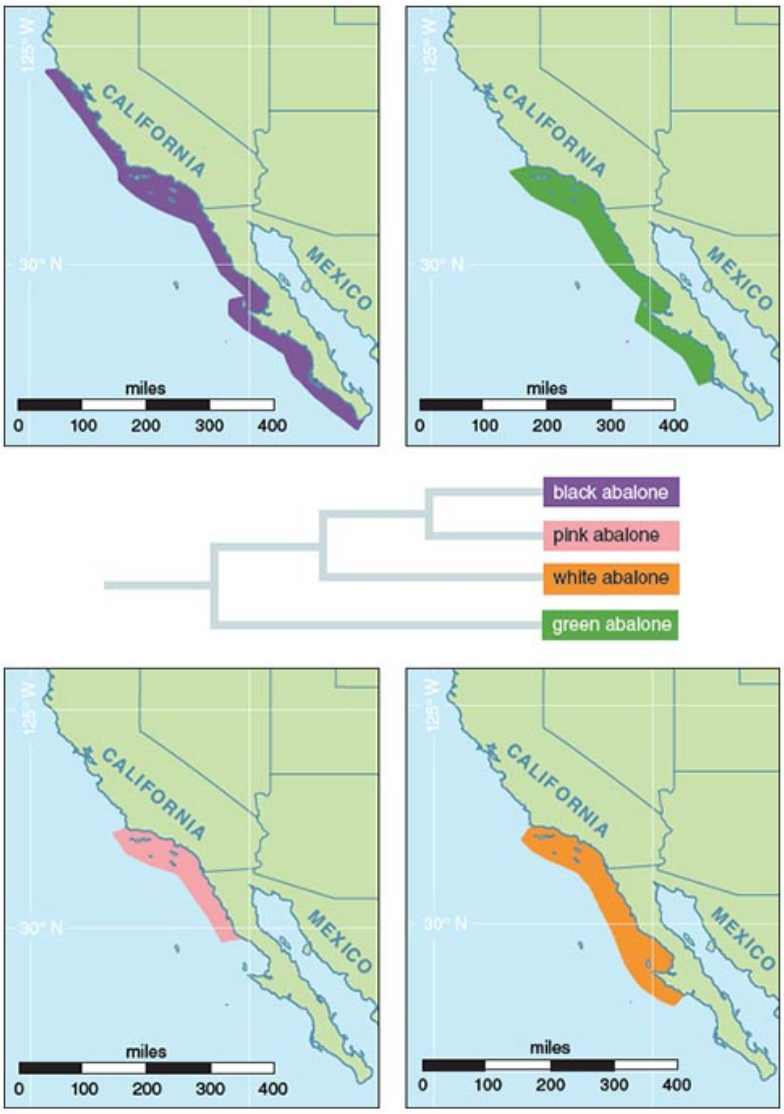


Figure 4. Abalone population distribution. Abalone historically have had overlapping ranges along the coast of California.Because they are external fertilizers, their sperm and eggs mix together in ocean water. Their sperm–egg protein interactions must be specific to each species to prevent hybridization. Rapid evolution has been detected in abalone reproductive genes, even within these closely related species, which has led to diversification.

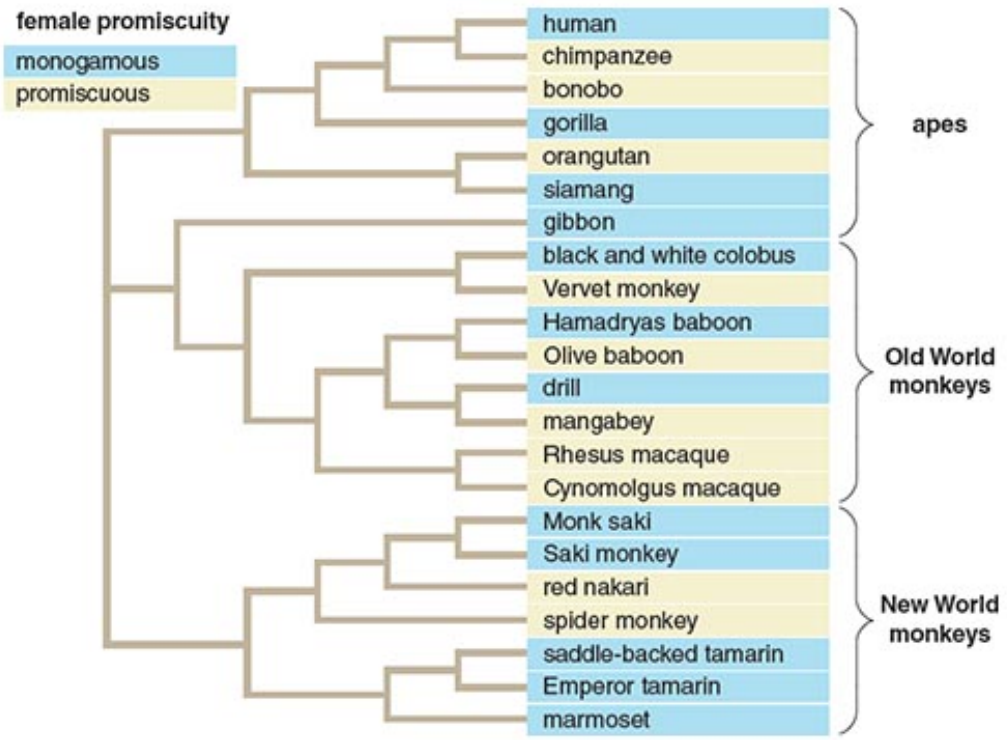


Figure 5. Primate mating systems. Primates have diverse mating systems, and closely related species have very different mating habits. This diversity provides an excellent platform for studying its effects on the evolution of reproductive traits and proteins.

Figure 6. Pekin duck genetalia. The spiral shape of the Pekin duck's genetalia is a result of promiscuity and male–male competition. Only the males with the most perfectly fitting spiral shape will successfully navigate the twists and turns of the female's vagina and fertilize her eggs. (Image from P. L. Brennan, et al. *Proceedings of the Royal Society B* 277:1309.)

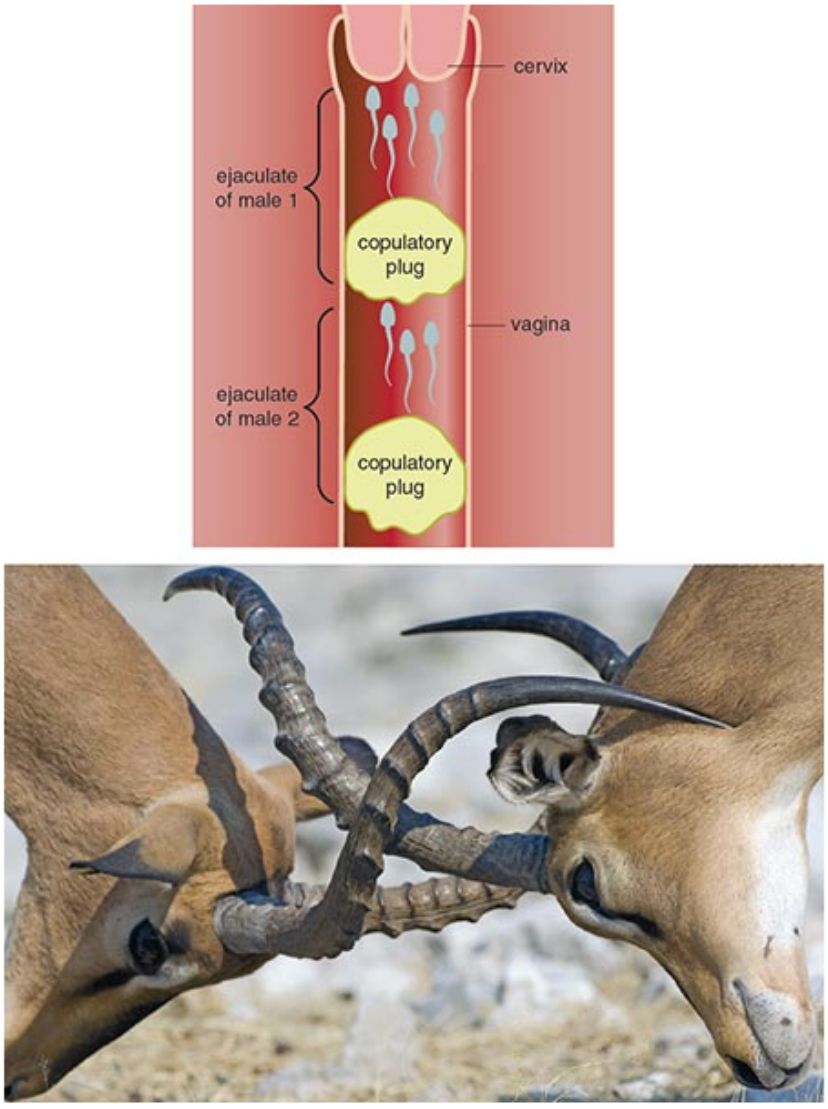


Figure 7. Precopulatory and postcopulatory sexual selection. In postcopulatory sexual selection, the selective process occurs after sex. For example, copulatory plugs form from proteins in semen and physically obstruct another male's sperm from advancing toward the egg *(left)*. By contrast, precopulatory sexual selection occurs before mating. Male courtship rituals and male–male combat over a female, such as the male impalas at right, are examples of precopulatory sexual selection.

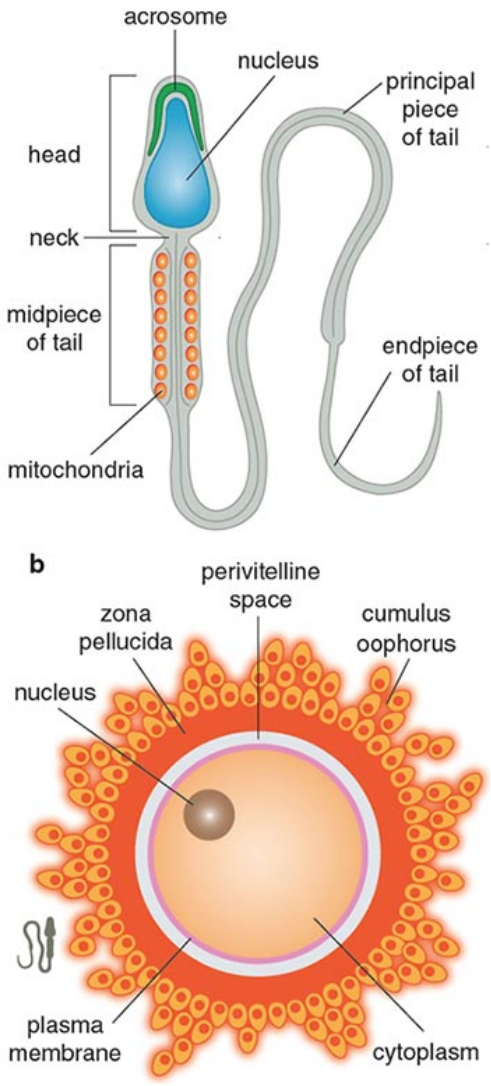


Figure 8. The basic stucture of sperm and eggs. The basic structure of sperm and eggs are conserved across animals. The sperm's *acrosome* is a compartment containing digestive enzymes to facilitate binding to the egg. The sperm swims toward the egg using the motile tail. The egg is surrounded by multiple protective layers. For successful fertilization to occur, the sperm must pass through the egg envelope (or *zona pellucida* in mammals) so that the nuclei of sperm and egg can fuse.

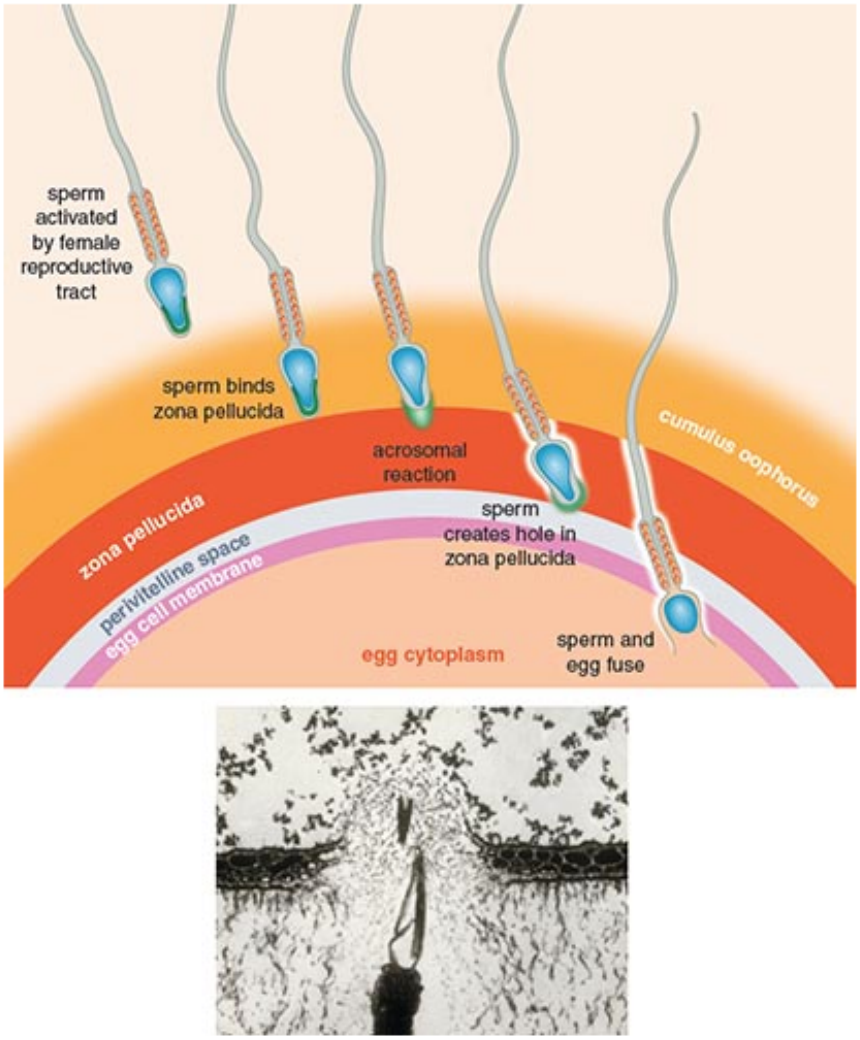


Figure 9. Sperm-egg protein interactions. Effective sperm–egg protein interactions are essential to each step of the fertilization process *(above)*: The sperm and egg attract one another; the outer proteins on the sperm bind to the egg's *cumulus oophorus*; the digestive enzymes from the acrosome allow the sperm through the egg envelope, or *zona pellucida* in mammals; and finally, cytoplasm and nuclei of sperm and egg fuse. The thin section transmission electron micrograph at left shows a sperm successfully gaining access through the egg envelope. (Image from N. Kresge, et al., *Bioessays* 23:95.)

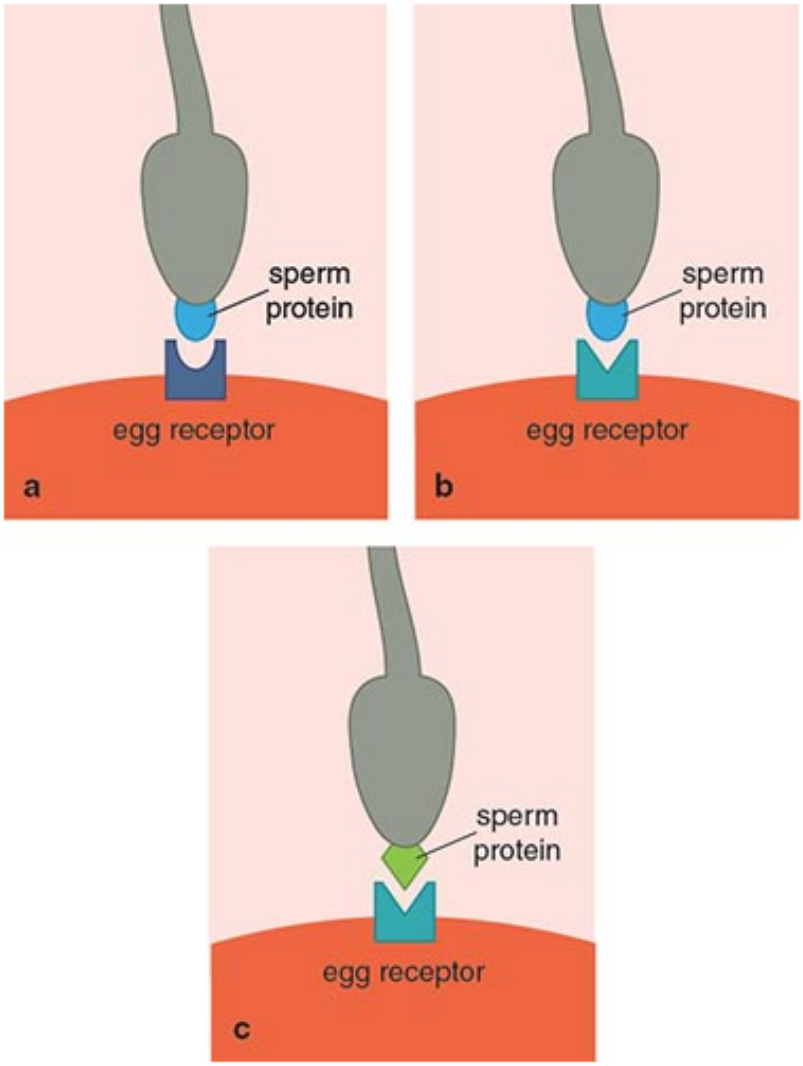


Figure 10. Sexual conflict, when the selective pressures on the egg differ from the selective pressures on the sperm, can lead to rapid evolution in reproductive genes. In this example, sperm and egg are only able to bind if their proteins are compatible, indicated here by matching shapes *(a)*. If the egg is under selective pressure to reduce *polyspermy,* or fertilization with multiple sperm, changes in egg proteins will be selected over generations to reduce the number of compatible sperm *(b)*. The sperm, however, are under selection for successful fertilization, so over subsequent generations, the new compatible sperm protein will increase in the population in response to the egg's change *(c)*.

Chapter 2: Evolution of the Egg: New Findings and Challenges

Introduction

The general structure of the egg has been conserved millions of years (8). Despite this conservation, specific proteins surrounding the egg are undergoing rapid evolution in many different species. It has been shown that many of these changes are under positive selection, and these egg proteins may be important for species-specificity during sperm-egg interactions and may establish of barriers to fertilization (6). Whether fertilization occurs externally (marine gastropods) or internally (most mammals), the definite players in sperm-egg binding and fusion still remain an enigma. The interacting proteins on the sperm and egg surfaces remain a puzzle for many species. Also, what is the mechanism of sperm-egg interactions? How do amino acid sequences, glycosylation, or receptor complexes play a role in enabling sperm-egg fusion? Disentangling the mechanism/s responsible for sperm-egg interactions remains a primary question in this field, and exciting new evidence from the composition and evolution of the egg provides important clues on how fertilization occurs in various species.

Biochemical and genetic methods have revealed many insights to the fertilization process and sperm-egg interactions. Knowledge gained about the mammalian system from the genetic manipulation of mice presents a complicated, and at times contrary, story of fertilization. However, the most extensively characterized reproductive systems are in marine invertebrates, the sea urchin (*Echinometra*) and abalone (*Haliotis*). Because of similarities in egg structures, invertebrates are very useful for characterizing reproductive systems in general. Newer and more precise technology, such as in the proteomics field, has enabled the identification of the complete repertoire of previously unknown reproductive proteins, many of which may prove to be major

players in sperm-egg interactions (9-11). The availability of whole genome sequences from multiple species has allowed sophisticated evolutionary methods to be developed, and links between genotype and phenotype may have important future implications.

In this review, we will discuss the current progress on the identification of egg components involved in sperm-egg interactions and reconcile the formation of egg extracellular barriers with evolutionary strategies (e.g. avoidance of polyspermy) with an emphasis on mammalian systems, though many of these processes are applicable to other vertebrate and invertebrate species. We discuss how the egg has evolved to become a specialized cell, and how interactions with the sperm have influenced egg protein evolution. We also examine the current fertilization paradigm and suggest future directions in the field to address unresolved questions.

Overview of fertilization process

There is considerable investment by organisms to create viable oocytes, which are often far outnumbered by spermatozoa, and the quality of the egg is directly related to reproductive outcome and number of offspring. Immature follicle cells transform into mature oocytes through the process of folliculogenesis, which includes oogenesis and maturation ((4) provides an extensive review of this). Eggs begin growing in female organisms before they are even born (12.5 days post-coitum in mice and 12-16 weeks in humans) and they enter and remain in prophase I of meiosis I until puberty (4). Meiosis resumes when females enter puberty where a surge of luteinizing hormone (LH) initiates ovulation and the mature egg is transported to the oviduct of the female reproductive tract, where sperm-egg interactions can occur.

Although there are many factors that play an important role in reproduction, we will focus specifically on the sperm-egg interactions in the fertilization process outlined below and in Figure 11. This process is can be generalized to both invertebrates and vertebrates. 1) Once

copulation and the release of millions of male sperm have occurred, the mobile sperm race into the vicinity of the egg. 2) The sperm pass through the cumulus oophorus, a matrix of cumulus cells and hyaluronic acid (HA) surrounding the egg (Figure 12.). 3) Sperm undergo the acrosome reaction (AR), where the acrosomal vesicle is exocytosed from tip of the sperm head. 4) Of the few sperm (~100) that reach the Zona Pellucida (ZP) extracellular matrix (another layer surrounding the egg), a sperm binds to the ZP through a complex interaction of sperm and egg receptors. 5) The AR enables the sperm to cross the ZP barrier and it enters the Perivitelline space. 6) The acrosome-reacted sperm binds to the egg plasma membrane. 7) The sperm's membrane and cytoplasm merges with the egg's and a zygote is formed.

While this is a complicated journey for the sperm and it undergoes a complex maturation process, for the purposes of this review, we will focus primarily on the egg structure and function and only mention specific sperm proteins if they interact with the egg/cumulus complex. For a review focusing on sperm structure and maturation and its role in fertilization, please refer to this review by Reid et al. (2011).

Basic Egg Structure

Mammalian and non-mammalian eggs have some similar structures and characteristics. The genetic components (chromosomes and mitochondria) of the egg are enclosed by a plasma membrane, which is surrounded by an extracellular matrix called the ZP, and an outer layer called the cumulus oophorus, Figure 12. The functions of these cellular barriers are to protect the egg, ensure proper maturation and implantation, crosstalk to sperm (sperm activation and recognition), and to serve as a block to polyspermy.

Although we see differences in reproductive behaviors and the interaction between molecules may vary, the basic structure of the egg has remained conserved for millions of years.

For example, many egg homologies are still retained between internally and externally fertilizing species. The mouse, sea urchin, and abalone diverged from each other millions of years ago, but they serve as model organisms in reproductive studies, and tell us a lot about the proteins and mechanisms involved in the sperm-egg interactions and fertilization in humans. There are also many species-specific differences that have evolved over time that we will discuss. Each section below will focus on the composition and properties of a specific egg extracellular barrier, discuss the mechanism by which molecules interact with sperm, and current debates. We will move in chronological order from the first barrier the sperm encounters on its journey from the male reproductive tract to the female egg: cumulus oophorus, the ZP, and the plasma membrane. In a later section, we will reconcile the evolution of the egg with the selective forces and focus on the egg barriers mentioned here.

The cumulus oophorus

The outer layer of a mature egg is surrounded by follicle-related cells that make up a matrix called the cumulus oophorus, Figure 12. (12). Approximately 3000 cumulus cells make up the cumulus oophorus, and these cells function to nurture and communicate with the egg and promote the process of ovulation (13). For a long time, the sperm AR was thought to be induced primarily by the interaction with ZP, and this focus on the ZP caused the cumulus cellular layer to remain an ambiguous extracellular barrier surrounding the egg. Subsequently, most *in vivo* studies of sperm-egg interactions have removed the cumulus cell layer prior to experiments. While eggs lacking the cumulus barrier are capable of being fertilized by sperm *in vitro*, it has been shown that having an intact cumulus layer promotes greater fertilization in mice and may play an important role in sperm-egg interactions, though this data varies (14-16). More recently, evidence suggests that cumulus cells promote the sperm AR, and this "premature" activation

may limit the number of ZP binding sperm from reaching the egg or it may play it's own role in preparing sperm for the AR (17-20).

Prior to ovulation and in response to LH, the cumulus cells surrounding the egg undergo mucification, or cumulus expansion. During cumulus expansion, the cumulus cells spread out and are connected together by HA oligosaccharide chains, proteoglycans, and hyaluronan binding proteins – all of which form a tight matrix of cells which interact with the egg (12). The cumulus matrix poses a formidable barrier to any approaching sperm, but it is also thought that chemoattractants emitted by these cells attract mammalian sperm (21-23). In early studies, the mechanism by which sperm bypassed the cumulus matrix was thought to be caused by the AR, where the AR exudates dissolved the HA matrix of cumulus cells and thus allowed sperm to pass freely. It was also suggested that acrosome-reacted sperm in the cumulus matrix would facilitate the passage of non-acrosome reacted sperm, because of the general view that acrosome-reacted sperm are unable to bind to the ZP. However, a large number of acrosome-reacted sperm have been found in the cumulus cell layer through live video imaging, and this raises the possibility that cumulus cells may play an important role in inducing the AR (17,20).

Nonetheless, the current mechanism for sperm passage through the cumulus cell layer is through the action of sperm surface proteins. Glycosylphosphatidylinositol (GPI)-anchored membranous proteins, Ph-20 and Hyal5, on the sperm membrane have been implicated in sperm penetration of the cumulus cell layer (24,25). Both of these proteins have hydrolic acid regions that have been proposed to have functional hyaluronidase activity which digests the HA outside cumulus cells and allows sperm penetration through cumulus layer (24). However, homozygous null mouse sperm lacking Ph-20 are still able to fertilize eggs even though they experience delayed penetration through the cumulus (26). This would suggest the presence of an additional

protein with dual functionality, perhaps Hyal5, but further functional studies will need to be done. In future studies, the phenotype of a double knockout of Ph-20 and Hyal5 could provide important insights into this process.

While the cumulus oophorus has not received a lot of attention, it may play an essential role in reducing polyspermy by trapping sperm and ensuring that only healthy sperm reach the egg. Alternatively, recent evidence has shown that most sperm reaching the ZP are already acrosome-reacted, which brings interesting challenges to the currently held view that the ZP induces the AR and that only acrosome-intact sperm can bind to the ZP (19,20). Inoue et al. (2011) show that not only are acrosome-reacted sperm able to penetrate the ZP, but these sperm are also able to fertilize normal offspring. This demonstrates that cumulus cells have some role in inducing the AR in sperm to facilitate fertilization. Future studies need to use the entire egg/cumulus complex to obtain accurate and meaningful results. In addition, while *in vitro* experiments can give us clues about the function of specific proteins, without competitive mating experiments we are unable to disentangle the complex interactions of sperm *in vivo*. Sutton et al. (2008) found that mice with a homozygous mutation of the Pkdrej gene were still able to fertilize the egg, but in competition trials with wildtype sperm, the Pkdrej mutant sperm took two hours longer than wildtype to reach the egg/cumulus complex. In the future, it would be useful to determine what other mechanisms induce the AR and how sperm interact with cumulus cells in competition experiments.

In marine invertebrates, a similar outer layer is found called the jelly coat, or the egg jelly (EJ). The sea urchin (*Strongylocentrotus purpuratus*) jelly coat is well studied; it is a transparent layer surrounding the egg that is composed of sulfated polysaccharides (27). Sulfated fucans on the EJ bind to a sperm surface receptor and induces a signal cascade that causes the AR (27).

Different proportions of the two main sulfated fucan isotypes in sea urchin, Sulfated fucan I and Sulfated fucan II, determine the reactivity of the sperm to egg (28,29). Egg jelly sulfated fucans are more receptive to sperm of the same or similar species. This provides much evidence that specific sulfation patterns and glycosidic linkages ensure species specificity in the sea urchin, and it is likely that we will encounter a similar case in mammals.

The Zona Pellucida

The extracellular matrix surrounding the egg is called the zona pellucida (ZP) in mammals, the vitelline envelope (VE) in amphibians, marine gastropods, and drosophila, the chorion in fish, and the perivitelline envelope (PE) in reptiles and birds (30). Despite the diversity in naming, egg structural and functional similarities suggest homology across vertebrates and invertebrates, Figure 13 (31,32). The ZP is composed of glycoproteins and is one of the main players in sperm-egg interactions and species-specificity. Competition experiments have shown that the ZP plays a role in limiting cross-species fertilization.

The ZP is composed of three to six zona pellucida glycoproteins (ZPGs). ZPGs are organized into 6 main subfamilies: ZP1, ZP2/ZPA, ZP3/ZPC, ZP4/ZPB, ZPAX, and ZPD; combinations of these proteins form the extracellular matrix surrounding the egg (31). ZP nomenclature is confusing, and we will use the simplified nomenclature referred to in Conner et al. (2005). All ZPGs have a 260 amino acid ZP domain, which is responsible for the polymerization of the ZPG. Figure 13 shows the domain structure across all ZPGs. The ZP domain has two subdomains called ZP-N (on the N-terminal portion of the ZP domain) and the ZP-C (on the C-terminus), each of which can exist as independent subdomains within the proteins. In fact, various ZPGs have multiple copies of ZP-N subdomains (31,32). Other common domains found in ZPGs include an N-terminal signal sequence, a C-terminal propeptide

(CTP) that is removed with at a consensus furin cleavage site (CFCS), a trefoil domain, and a transmembrane domain (TM). ZPG loci are found on different chromosomes and there is high homology between proteins (over 65-98% identical between ZP2 and ZP3) that suggests ZP proteins arose by gene duplication (8). Goudet et al. (2008) determined phylogenetically that ZP1 and ZP4 were the most recent duplications to occur, followed by the ZP2 and ancestral ZP of the ZP1 and ZP4 duplication (Figure 13). ZPAX and ZPD are pseudogenes in most mammals so they are excluded from further mention. The evolutionary origins of ZP3 is more complicated and it was proposed that ZP3 and the ancestral ZP gene to all other ZPGs was formed during the original duplication event (Goudet 2008), but this remains to be verified. Further characterization of ZP components in distantly related species would solve this question.

In humans, four ZPGs form the ZP: ZP1, ZP2, ZP3, and ZP4 (33,34). ZP2 and ZP3 make up the majority of the ZP (~80%) and form into long fibrils interconnected by the ZP1 and ZP4. ZP4 is found in humans, other non-human primates, and rats, and has been proposed to have similar functions as ZP1 (31). Evidence from various species suggests that the evolution of ZP genes occurs by gene duplication and pseudogenization. In particular, it is interesting to note that different organisms utilize various combinations of ZPGs to construct the ZP, which may have implications for species-specific fertilization. For instance, the mouse only has ZP1, ZP2, and ZP3 in the ZP. ZP2 and ZP3 are consistently found in the ZP of various species, and this suggests that these proteins are essential for ZP formation. In fact, further support for this in mouse models shows that when either ZP2 or ZP3 are mutated, a ZP barrier does not fully form (35,36).

One expects that proteins involved in crucial sperm-egg interactions and fertilization be evolutionarily conserved across species. Contrary to expectations, it has been shown that both

ZP2 and ZP3 are evolving rapidly (37). The rapid evolution of these proteins has promoted divergence between closely related species of primates, and positive selection on the sperm binding portion of ZP3 has been proposed to be caused by selective pressures related to sperm-egg interactions. Disentangling what selective forces play a role is a complicated problem, but there are some great examples in fruit flies, nematodes, and mammals where barriers to cross-species fertilization are established by allelic differentation; giving credence to adaptation by rapid evolution hypothesis (38-40). In particular, gamete recognition in sea urchins (*Echinometra*) is dependent on polymorphisms in the sperm protein binding, which allows sperm to attach to eggs (41). Interestingly, the eggs are very specific in the type of sperm that they allow to bind, and preferentially bind the sperm with similar genotypes. This provides evidence that specific allelic combinations can have strong affects on fertilization.

The mouse ZP is composed of only three ZPGs: ZP1, ZP2, and ZP3. Homozygous ZP2 and ZP3 knockouts of female mice fail to produce a ZP during oocyte growth and the oocytes are often infertile (8). Both ZP2 and ZP3 are important for proper ZP formation. Although homozygous ZP1 mutants produce a ZP, it is loose and not interconnected, and mutant females are not as fertile as wildtype mice (42). This suggests that the role of ZP1 is to interconnect ZP fibrils and demonstrates that the integral linkage of the ZP to fertilization. Intriguing results from sperm-binding assays show rat sperm can bind to mouse ZP but not human ZP (43). Recall that the rats contain the same type of ZPGs as humans, whereas the mouse has a pseudogenized ZP4 protein, demonstrating that rat sperm may recognize residues other than ZP structures. These findings have revealed many complex phenotypes which suggest two scenarios: 1) sperm recognition of the ZP is mediated by the species-specific glycosylation patterns, regardless of the amino acid sequence, or 2) that sperm have evolved to interact with a ZP complex composed of a

specific set of ZPGs (three, four, or more proteins). No doubt we will find that both of these mechanisms will play a role in sperm-egg interactions, with some species using more or less of each. In fact, glycosylation patterns do not appear to be important for sperm-egg interactions in the bonnet monkey (*Macaca radiata*) where amino acid sequence recognition by sperm was sufficient for the binding of recombinant ZP3 to sperm heads (44). The role of carbohydrate moieties on the ZP remains to be clarified, but significant work has been done toward identifying the specific oligosaccharide modifications.

As we alluded to previously, ZPGs are heterogeneously glycosylated with asparagine-linked (N-) and serine/threonine-linked (O-) oligosaccharides (45). These glycosylation sites are modified by sulfation, sialylation, and other moieties. Among ZPGs, specific glycosylation sites are conserved and this may relate to the availability of sperm binding sites. In particular, ZP3 has been implicated to be a major player in sperm binding and has been identified as a sperm binding receptor on the ZP (46). In mouse and human ZP3, two conserved serine residues (serine 332 and 334) located in the C-Terminus region make up the purported sperm-binding region (46). Recent studies have made strides in identifying the oligosaccharide modifications on human ZP3 with highly sensitive sample preparation and mass spectrometry on human oocytes (11). In this study, an abundance of sugar molecules called sialyl-Lewisx (SLeX) were found at the ends of ZP3 oligosaccharides, with sialic acid as the terminal sugar and combinations of galactose, fucose, and N-acetylgucosamine making up the internal tree structures (11). Mice completely lacking any O-linked oligosaccharides on the egg surface were unable to bind to any sperm (47). While the exact contribution of glycosylation remains to be determined, it is clear that modifications are important for fertilization.

It is quite surprising that that the crystal structure of these important proteins, ZPGs, have

remained elusive for so long. The general properties of ZPGs provide the answer to why a crystal structure has not been obtained: ZPGs are heavy and heterogeneously glycosylated proteins, they tend to aggregate when concentrated, and undergo a complex maturation (48,49). Three decades after the identification of ZP3, the crystal structure of the mouse ZP3 ZP-N domain (the N-terminal portion of the ZP domain) and the full structure of the avian precursor ZP3 (including the external hydrophobic patch (EHP) propeptide) were finally resolved using X-ray crystallography (30,49). Interestingly, the structure of the 2 subdomains (ZP-N and ZP-C) in the ZP domain revealed a new class of immunoglobulin (Ig)-like subtypes (30). The structure of each unique domain consisted of two beta sheets whose strands enclosed a hydrophobic core. The hydrophobic core contained either 8 (type I) or 10 (type 2) invariant cysteines.

Currently, two contrasting hypotheses exist concerning the location of sperm binding sites on ZP3. One hypothesis suggests that exon 7 of the ZP3 gene carries two active O-glycosylation sites (S332 and S334) that are under positive selection (Swanson 2001). These sites have been suggested to be the sperm-binding sites based on mouse ZP3 mutants in embryonic carcinoma cells (50). More recently, transgenic mice with mutated residues S332 and S334 show that these sites cannot be intrinsically involved in sperm binding because mutant female mice remained fertile (51). Additional proteomic evidence shows that the two sites are not modified *in vivo*, as one would expect if they were important glycosylation sites for sperm-binding (52). An alternative hypothesis has emerged in which two other conserved O-glycosylation sites (referred to as site 1 and site 2) are responsible for sperm binding (30). The crystal structure of ZP3 favors that site 1 and site 2 are the actual sperm binding sites because both are exposed on the protein surface of mouse ZP3, allowing interaction with sperm membrane proteins (49). Han et al. (2010) obtained the 3D structure of the precursor ZP3

protein that also contained the EHP, which inhibits polymerization and conformational changes. Further characterization of the mature ZP3 structure without the EHP may reveal additional binding sites. Future mutational studies need to be carried out on site 1 and site 2 to promote the alternative hypothesis. Despite this, the crystal structure has been valuable and enabled subsequent studies examining structural homology.

Amino acid sequence comparisons do not provide a clear picture that mammalian ZPGs are homologous to the VE proteins in marine species because of low sequence similarities, but structural homology of the ZP-N and N-terminal repeats in VERL suggest that this is the case (10,53). For example, the vitelline envelope receptor ligand (VERL) is a large glycoprotein with a ZP domain that makes up a portion the VE of abalone. Using shotgun proteomics on proteins isolated from the abalone VE, 29 additional proteins (including VERL paralog VEZP14) constitute the abalone VE, by far the most diverse egg coat characterized to date (10,53). The diversity of the abalone VE is in stark contrast to the relative simplicity of the vertebrate ZP composed of 3-4 egg proteins. The abalone system is unique in that it is one of a three model animal systems in which interacting egg and sperm proteins have been identified and functionally characterized (drosophila and sea urchin are the other two) (54-56). An affinity binding approach was used to identify VERL as the binding partner to the sperm protein Lysin (57). VERL contains 22 tandem repeat structure, each of which consists of 153 amino acid sequences, that have been undergoing concerted evolution, with the exception of the first two repeated structures (6,10,58). While most of VERL is evolving neutrally, positive selection acts in the two terminal N-repeats (10). The sperm is able to penetrate the VE through a non-enzymatic process by competing sperm lysin with VERL fibrils to unravel a portion of the egg VE to allow sperm to cross.

The crystal structure of ZP3 has enabled subsequent studies that suggest a link between human ZPGs and yeast proteins. Swanson et al. (2011) used protein structural homology to show that human ZPGs, abalone VERL and VEZP14, and the ZP-N domain of Sag1p, a yeast protein alpha-agglutinin, had the same 3D protein folds. This suggests that the same basic domains are at play in vertebrates, invertebrates, and unicellular eukaryotes, suggesting commonality over 0.6-1 billion years of evolution. Basic cellular interactions appear to be ubiquitous across cell-types and there is no reason we should expect the sperm and egg cell adhesion to be any different.

As we have seen, the mechanism for sperm-egg interactions (binding and fusion) with the ZP involves the sperm binding to specific egg receptors, but this is not the case in many other species. Various mechanisms of sperm-egg interactions have evolved because of differences in behavior, environment, and morphology. As such, we briefly discussed allelic specific interactions seen in abalone and sea urchins, but there are also physical mechanisms that have evolved. For example, in most fishes and drosophila the VE serves as a structural barrier, and the only entrance for the sperm to bypass the VE is through the micropyle, a physical opening in the ZP where only one or a few sperm can enter at a time (59). Very little is known about the sperm-egg fusion event in species with the micropyle.

The Plasma Membrane

The plasma membrane is the last barrier that the sperm has to bind and fuse to in order for fertilization to occur. This membrane is composed of lipids and membrane proteins, similar to many other cellular membranes, with the unique property of facilitating sperm binding and membrane and cytoplasmic integration. In mammals, tetraspanins and integrin receptors on eggs have been implicated in the sperm binding and fusion process (60). There are multiple tetraspanins found in many mammalian tissues and they have four transmembrane domains that

span the plasma membrane. In particular, CD9, a member of the tetraspanin family, is also found on the egg membrane surface and is essential for proper egg function (61). Homozygous CD9 mutant mice are severely affected and have much lower fertility than their wildtype litter-mates, demonstrating it's importance in sperm binding (62). Null CD9 mice maintain some fertility, perhaps due to the presence and functional redundancy of CD81, which is 45% identical to CD9 on the sequence level, and also found on the egg plasma membrane. Further strength for this hypothesis was provided when female mice with a double knockout of CD9 and CD81 were shown to be completely infertile (63). These data suggests that combinations of tetraspanins on the plasma membrane in different species may contribute to species-specificity. Other membrane proteins such as integrins or GPI-anchored proteins may also facilitate this process.

Integrins are membrane receptor proteins that have been well characterized throughout the body, and their primary functions are to mediate attachment between a cell-cell and cell-tissue and to initiate cell signaling within and outside the cell. Integrins are composed of α and β subunits; different heterodimeric combinations of which result in 10 unique integrin pairs in mice eggs (61). Mice knockouts studies have shown that integrins are not essential to fertilization and sperm-egg interactions, but in combination, there complex structures may play an important role in sperm binding to the egg as demonstrated in a time-lapse video where integrin-deficient eggs (deficient in ITGB1) show a delay in sperm binding (64). In particular, the ITGA9-ITGB1 (α9β1) pair has shown decreased fertility in mutant mice, and may be important for facilitating the sperm binding process (64).

Rapid Evolution of Egg proteins

We return to the original puzzling question of why the egg proteins are undergoing rapid evolution despite displaying structural conservation across taxa. Rapidly evolving proteins

experience greater amino acid substitutions between species. One can test for rapid evolution by using a robust way to test for positive selection, which doesn't require any a priori knowledge, by calculating the ratio of the number of nonsynonymous substitutions per nonsynonymous sites (d_N) to the number of synonymous substitutions per synonymous sites (d_S) for each SFP (65). The ratio of $d_N/d_S = 1$ indicates that neutral evolution is occurring. When d_N/d_S is less than 1, this indicates that purifying selection (conserved evolution) is occurring. When d_N/d_S is greater than 1, this indicates that positive Darwinian selection (rapid evolution) is occurring. The genome-wide d_N/d_S average for protein coding genes is 0.6. By comparing the likelihood ratios (LR) between neutral (M1, M7, M8a) models and positive selection models (M2, M8), one is able to identify positive selection acting on genes. M8 also allows one to identify specific codon sites under selection, which may be led to potential functional sites within a gene. Using these tests, we can determine if the rapid evolution is adaptive. The evolution of the egg is dynamic, and, as we mentioned before, many egg proteins are under rapid evolution that may be adaptive.

Selective forces such as sexual selection, sexual conflict, reinforcement, and avoidance of pathogens have likely contributed to the variation and evolution of egg coat proteins and structures. In fact, the multitude of barriers that the egg has accumulated (cumulus cells, ZP) suggests that these might have evolved as an egg defense strategy against polyspermy. Remarkably, despite rapid divergence of their constituent proteins, animal egg coats share a common molecular basis (66-68). The rapidly evolving egg coat proteins may suggest that they are continually evolving in a race with sperm membrane proteins, and also to limit the genetic contribution from non-compatible sperm or cryptic female choice or polyspermy. In the following sections, we will disentangle how these various selective forces may contribute to the evolution of the egg.

ZP2 and ZP3 on the Zona Pellucida

We would predict that many female proteins involved sperm-egg interactions would be undergoing rapid evolution, since many male sperm proteins involved in reproduction are also undergoing rapid evolution. Proteins on the extracellular barriers of the ZP are great candidates, and both ZP2 and ZP3 have been shown to be evolving rapidly (37). This study identified specific amino acid residues in the proteins that were under intense selective pressure, and in particular, found positively selected sites in the sperm combining region of ZP3. These findings promote the hypotheses that sperm competition, sexual conflict, and cryptic female choice drive the evolution of these egg proteins found on the ZP.

CD9 on the plasma membrane

We also find evidence that a protein found on the plasma membrane is under positive selection. In primates, CD9 diverges rapidly between closely related species and is subject to positive selection (R. George, unpublished). This suggests that the forces of selection are at work at the plasma membrane also, and promote the hypotheses mention above. But, this remains a little studied area, and future directions should take into consideration not only sperm candidates, but also their binding partners on the egg. We would expect to see concerted evolution of both interacting proteins if they are evolving in response to each other. In fact, Clark et al. (2009) used a likelihood method to show that abalone VERL and Lysin had correlated evolutionary rates (69). This type of analysis could be done on potential sperm and egg binding proteins.

Mechanisms to limit polyspermy

Polyspermy occurs when more than one sperm enters an egg. Some species have evolved mechanisms that allow polyspermy and these species are able to form a healthy gamete despite

multiple sperm (70). However, in most cases, polyspermy is not acceptable and is often detrimental to the zygote because of centromere function. The egg has evolved multiple barriers to prevent polyspermy that are induced by the sperm's passage across the various egg barriers we discuss above. Following sperm penetration, signal transduction events result in a large efflux in Ca2+ from cellular stores that cause fast (changes in the membrane potential) and slow blocks (physical modification of egg's extracellular matrix) to polyspermy, which involve both chemical and physical modifications/formation of extracellular barriers. These blocks include such processes as the cortical reaction, the zona reaction, and modifications of the zona pellucida or plasma membrane, which are presumed to serve as blocks to polyspermy (71,72).

Physical barriers also play a large role in limiting polyspermy; these include the female reproductive tract and the cumulus cell layer. The female reproductive tract poses a formidable physical barrier for sperm, and only 100-200 sperm from the 50 million sperm that are ejaculated into the female during mating ever make it into the vicinity of the egg (73). Even if these sperm are lucky enough to get to the oviduct, they will encounter the cumulus cell layer surrounding the egg, which has been shown to trap sperm and promote premature AR that may prevent subsequent ZP binding (74). In addition, the body's natural defenses are turned on when foreign substances enter the female reproductive tract. The innate immunity may play a role in limiting sperm migration.

The cortical reaction is characterized as a fast block to polyspermy, and occurs when special organelles in the egg, cortical granules (CG), are exocytosed into the perivitelline space of the egg. These CG exudates form a CG envelope in the perivitelline space that may play a role in blocking polyspermy from occurring (75). The CG envelope, or fertilization envelope, has been exceptionally well characterized in the sea urchin (species), and it forms within 30-60

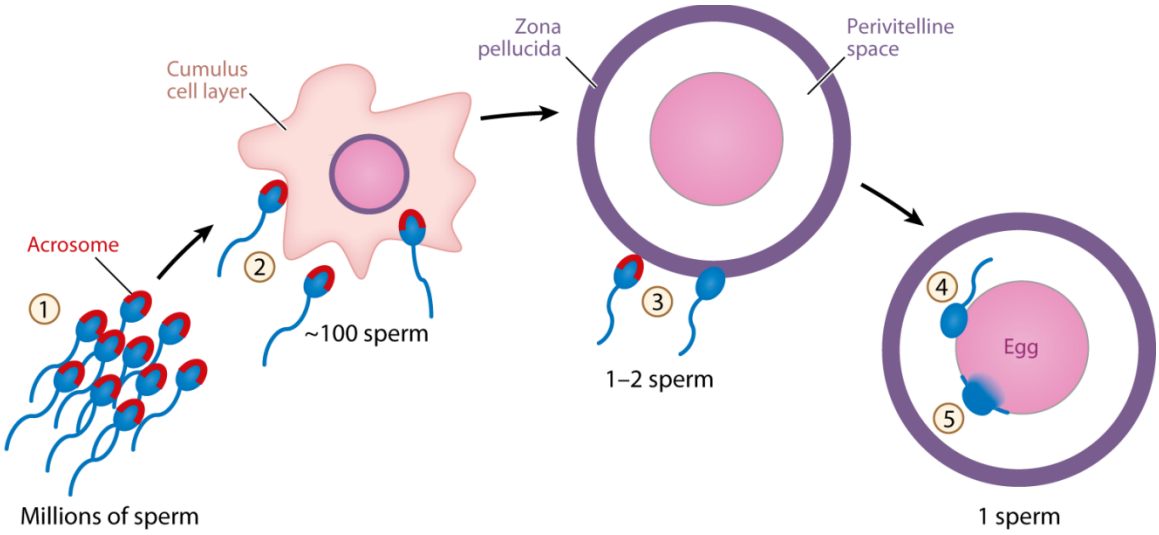
seconds post-penetration. One of its primary functions seems to be to prevent multiple sperm from entering the egg.

A slower block to polyspermy is the zona reaction, and it refers to the "hardening" of the ZP that renders it inaccessible to sperm. Proteinases, ovoperoxidase, and N-acetylglucosaminidase are thought to be the main players that bring about this change (74). Proteinases have been implicated in cleaving the propeptide on ZP2 and this cleavage causes structural modification of the ZP making it insoluble and no longer accessible by other sperm (74). In addition, it has been suggested that the membrane fusion event between the sperm and the egg results in cytoplasmic rearrangements that have sperm blocking properties. The physical modification of the plasma membrane may be the final block to polyspermy. In most cases, fast and slow blocks to polyspermy involve the cortical and zona reactions.

Conclusion

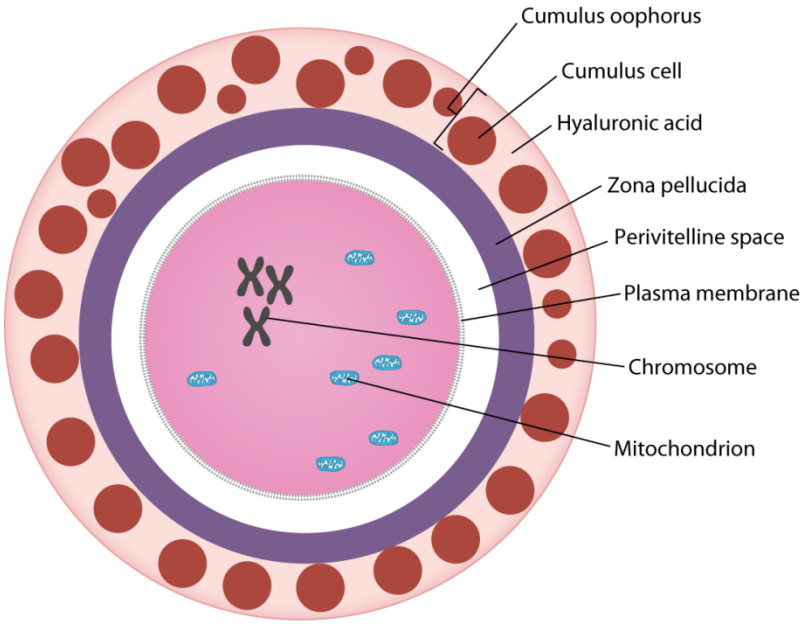
In terms of the evolution of the egg and sperm-egg interactions, we know a lot about the composition of egg coats and this has given us a lot of insight into the evolutionary forces driving the divergence of reproductive proteins. But, there remain many unanswered questions, and what about other stages in fertilization? For instance, we know that chemoattractants are important for allowing the sperm find the egg in the first place, but we still don't know a lot about the molecules or molecular interactions involved. Competition is fierce not only from an individual's own multitude of sperm but also from other males sperm. The molecular mechanism of sperm chemotaxis is still not fully understood, but it is an important guidance system for sperm in both invertebrates and mammals. Chemoattractants secreted by the egg and surrounding cumulus cells have not been evolutionarily characterized – are there species-specific chemoattractants? Also, we know that sperm-egg fusion must be an important process as both

sp18 in abalone and CD9 in primates evolves so rapidly, but what are the molecular interactions occurring between the sperm and egg?



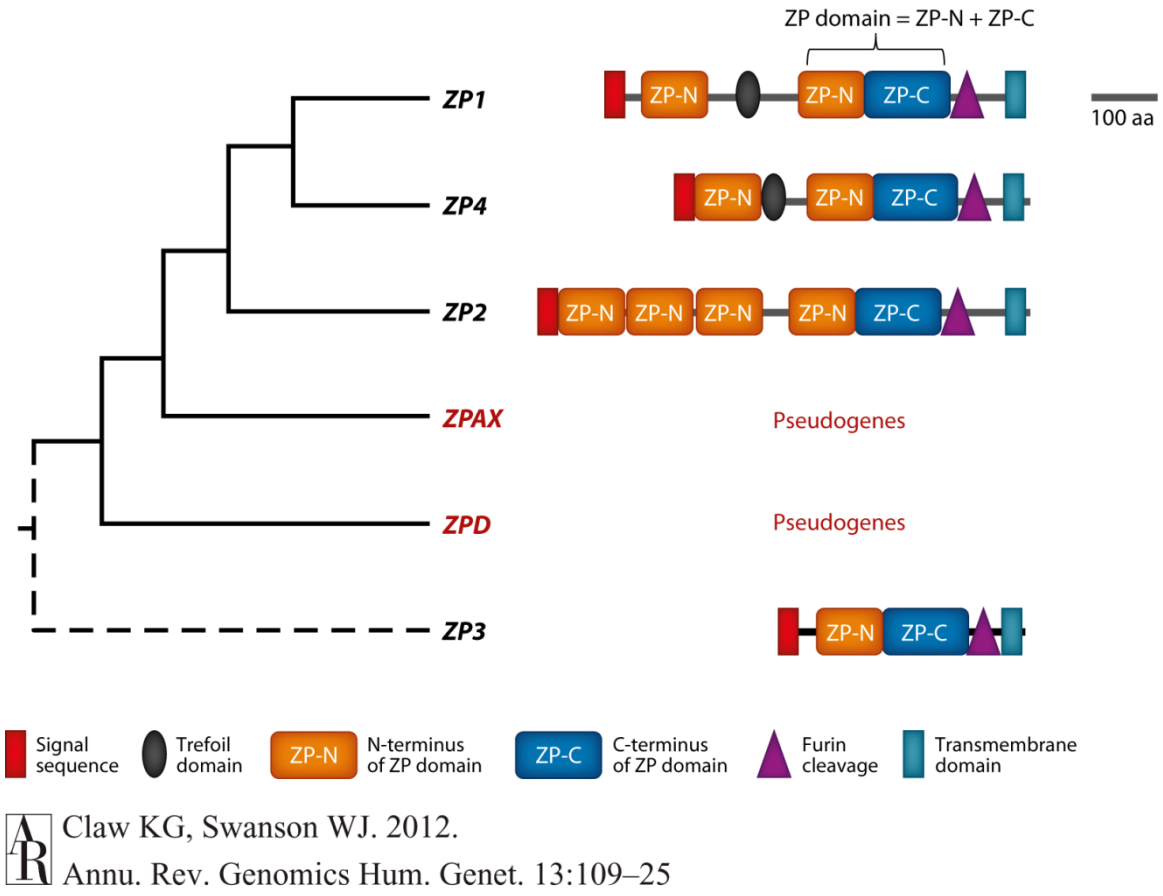
Claw KG, Swanson WJ. 2012.
Annu. Rev. Genomics Hum. Genet. 13:109–25

Figure 11. Process of mammalian fertilization. Sperm enter the female reproductive tract. Sperm pass through the cumulus cell layer. The first viable sperm to reach the egg's extracellular barrier binds to the zona pellucida, undergoes the acrosome reaction, crosses the zona pellucida barrier, and enters the perivitelline space. This sperm binds to the egg plasma membrane. The sperm and egg become cytoplasmically contiguous.



Claw KG, Swanson WJ. 2012.
Annu. Rev. Genomics Hum. Genet. 13:109–25

Figure 12. Basic structure of the egg.



Claw KG, Swanson WJ. 2012.
Annu. Rev. Genomics Hum. Genet. 13:109–25

Figure 13. Evolution of zona pellucida (ZP) glycoproteins and structure. The dashed line indicates the uncertain evolutionary origins of the *ZP3* gene. Abbreviation: aa, amino acid.

Chapter 3: Comparative proteomics and evolution of primate seminal fluid proteins

Introduction

Various forms of sexual selection have been proposed to explain the rapid evolution of reproductive proteins: sexual conflict, reinforcement, female *cryptic* choice, and sperm competition to name a few. The direct effect of sexual selection on genome-level changes and the molecular evolution of reproductive proteins remain ambiguous despite well-established evidence of sexual phenotypes varying with mating systems in primates (76). In primate species, correlations between sexual promiscuity and sexual phenotypes abound: larger sex to size ratios between males and females (sexual size dimorphism) and higher volumes and number of sperm in the ejaculates of males is common in promiscuous species (77-79). Currently, no conclusive evidence of correlations between the evolutionary rate of reproductive genes and mating systems exists, and no one has studied variation in reproductive protein abundance between primate species (76).

Primate species have diverse mating systems that evolved between closely related lineages and provide an ideal system to study the effects of mating systems on the evolution of reproductive proteins. To distinguish mating systems based on female promiscuity, we will refer to females who mate with a single male as "uni-male" mating systems and females who mate with multiple males as "multi-male" mating systems (Figure 14.). Semen, or ejaculates, is the product of five main organs in primates: the testes, epididymis, prostate, seminal vesicles, and the Cowper gland. Seminal fluid, the liquid portion separated from the spermatozoa, is easily collected and affects various physiological characteristics during reproduction, including: sperm motility, female immunological suppression, sperm competition, female receptivity, ovulation,

oogenesis, sperm storage, and copulatory plug formation. Recent studies in insects suggest that different mating systems can exert dramatically different selective pressures on seminal fluid proteins (SFPs) (80). In particular, SFPs involved in the formation and dissolution of the copulatory plug are under positive selection and show evidence of gene losses in less promiscuous species (81-83). One copulatory plug protein, SEMG2, has previously shown a correlation between evolutionary rate and mating system, with more promiscuous species having higher evolutionary rates (81). This suggests that mating systems may play an important role in the evolution of SFPs. We hypothesize that the selective forces that drive reproductive protein divergence differ between primates with different mating systems.

Wong et al. (2010) found that the genome-wide rate of nonsynonymous substitutions was greater in chimpanzees than in humans. Recently, Good et al. (2013) sequenced 285 ejaculate proteins from multiple individuals (n=20) from various primate populations, including gorilla, human, chimp, and bonobo. They did not find strong evidence that any of the ejaculate proteins were driven by sperm competition, and concluded that genetic variation was more greatly affected by gene function and effective population sizes than sexual selection. Sexual selection may be harder to detect because of independent contrasts and various pressures acting on proteins simultaneously.

Shotgun proteomics is increasingly being used to characterize the protein constituents in complex mixtures and has been used to identify proteins from male and female reproductive fluids, including seminal fluid, prostatic secretions, and follicular fluid, among others (84-86). Methods are improving so that we can quantify absolute and relative protein abundances of complex mixtures such as seminal fluid to make inferences about protein abundances, which can be complemented with evolutionary sequence analysis. Within primates, human seminal fluid is

the only primate in which the proteome has been comprehensively characterized (84,87). The role of sexual selection on protein regulation has not been studied extensively as most studies have focused on protein coding changes. Proteomic studies show that in general SFPs seem to vary widely between individuals in humans and rodents, but comprehensive quantification of SFPs across primate species remains to be done (88,89).

Here, we use a unique combination of evolutionary genomics and comparative proteomics to study the evolution of SFPs in human and non-human primates. We identify and quantify a large proportion of uncharacterized SFPs from the diverse mating systems of 8 primate lineages (Figure 14, samples used in proteomic analyses highlighted in grey). Using evolutionary methods, we show that a high proportion of primate SFPs are evolving under positive selection and use lineage-specific selection and Bayesian models to test for correlations between evolutionary rates and mating systems. In addition, we find evidence of a higher amount of pseudogenization in uni-male mating systems compared to multi-male mating systems. We have identified proteins that vary significantly in abundance between uni-male and multi-male mating systems and have broadly characterized protein abundance variation in primates. We show how the combination of genomics and proteomics can be a powerful tool for evolutionary studies.

Materials and Methods

Primate Samples

Semen samples were collected at various institutions, in a manner that conformed to animal and human subjects protocols. Collection of the non-human primate samples was performed at the Yerkes Primate Center (*Pan troglodytes troglodytes*/chimpanzee), Wake Forest

University (*Chlorocebus aethiops sabaeus*/vervet monkey and *Macaca fascicularis*/cynomolgus macaque), California National Primate Research Center (*Macaca mulatta*/rhesus macaque), Southwest Primate Research Center (*Callithrix jacchus*/marmoset and *Papio anubis*/baboon), and the San Diego Zoo's Institute for Conservation Research (*Mandrillus leucophaeus*/drill). Human semen samples were purchased from Lee Biosolution's (http://www.leebio.com/). Electroejaculation was performed to collect samples from the following primates (following protocol in (90)): rhesus macaque, vervet monkey, cynomolgus macaque, marmoset, baboon, and drill. An artificial vagina was used to collect samples from the chimpanzee (following protocol in (91)). Human samples were anonymously donated to Lee Biosolution's for research purposes. In total, eight primate samples with a minimum of two biological individuals per species (with the exception of the chimpanzee) comprised the dataset, including *species* (biological replicate): *Homo sapiens* (8), *Pan troglodytes troglodytes* (1)*, Macaca mulatta* (8), *Macaca fascicularis* (2)*, Papio Anubis* (2)*, Mandrillus leucophaeus* (2), *Chlorocebus aethiops sabaeus* (2), and *Callithrix jacchus* (2).

Sample preparation and Mass Spectrometry

After collection, samples were immediately frozen and shipped on dry ice to minimize any proteolysis. During sample preparation, semen samples were thawed at room temperature for 10 minutes, 300 uL (if possible) of the liquefied portion of the sample was separated, and centrifuged initially at 3000 x g for 10 minutes to separate the sperm from the seminal fluid. Samples were then centrifuged a second time at 10,000 x g for 20 minutes to ensure the complete separation of sperm. When a thick copulatory plug was present (chimpanzee), samples were thawed for an additional 30 minutes at 37°C. The proteins were quantified with BCA Protein Assay (Pierce) kit. Samples were randomized into batch groups of 10 to eliminate any sample

preparation bias. 200 femtomoles of horse myoglobin protein was spiked into each sample before digestion as a standard. 50 μg of each sample with the horse myoglobin standard was prepared for trypsin digestion (10).

After digestion, samples were cleaned up with MCX columns to remove detergents and glycerol contaminants. All batch samples were aggregated and the 3 technical replicates per sample were randomized in the order of loading onto the mass spectrometer. The digested samples were loaded onto a High-performance Liquid Chromatography (HPLC) column 30 cm in length and 75 nm in internal diameter. The column was packed with 30 cm of C-12 reverse phase material (Jupiter C12). The capillary column was then placed on-line to a LTQ-FT ion-trap mass spectrometer and eluted over a 3-hour gradient with increasing salt concentration in 3 random technical replicates of 5 μg each. Throughout mass spectrometry (MS) data collection, BSA peptides were used as controls and peptides abundance was measured using selected reaction monitoring (SRM) techniques. Mass spectra data was collected using data-dependent acquisition and MS peptide spectra were searched against their respective sequence databases using the Sequest algorithm (92). Species with no genomic sequences available were searched against the closest evolutionary relative (i.e. drill MS data was searched against the rhesus macaque coding reference sequences).

To improve discrimination between true and false positive identifications and to set an empirical false discovery rate, the Percolator algorithm was used (93). The MSDaPl software in the MacCoss lab, a protein inference program, was used to store and visualize proteomics results. MSDaPl infers parsimonious proteins based on the IDPicker algorithm (94). Because of the exploratory nature of this project and the high error threshold, a minimum of 1 peptide hit in a

run was used to identify a SFP. Using these filtering methods, a parsimonious list of inferred SFPs was generated for each species.

Normalization and quantification of relative protein abundance

Relative isotope abundances (RIA) were calculated for individual peptides detected in MS experiments using the program TOPOGRAPH (95). RIAs were normalized by first calculating the geometric mean of internal standard peptides across all samples (horse myoglobin and trypsin) to reduce the bias of noise or errors from ion abundances. Then, a geometric mean ratio was calculated for each MS run, and used to normalize all peptides in the run. Peptides with a Coefficient of Variation (CV) equal or less than 25%, as is the standard cut-off in clinical MS experiments, between technical replicates were included in further analyses.

To explore relative abundance variability, the %CV was calculated for all peptides within species (between mean biological replicates) and between species (between the overall means of biological replicates for each species). Peptides with high or low CV based on a 95% Confidence Interval were used to identify conserved and variable abundances between individuals/species.

A nonparametric test, Wilcoxon rank-sum test, was used to compare the relative peptide abundances from uni-male mating and multi-male mating groups. We performed a 2-sided test since we have no prior expectations, and pvalues were calculated to show evidence of a difference in the means between the two mating groups. Greater and less Wilcoxon rank-sum tests were used to detect the direction of the differences between the means.

Within the MS data, proteins often have multiple unique peptides, and we were interested in seeing how the concordance of peptides within the same protein measured. We plotted the relative abundance of 2 peptides from the same protein across multiple individuals/species, and used R^2 values to look for significant concordance between them. The average R^2 was taken from

proteins with 3 or more peptides. Although it is known that peptide modifications and inherent differences in ionization during MS scans can affect the calculated RIA.

Coding sequences and multiple sequence alignments

Coding sequences were obtained from publicly available reference assemblies of human (hg19), chimpanzee (panTro3), orangutan (ponAbe2), gorilla (gorGor3), Northern White-cheeked gibbon (nomLeu1), rhesus macaque (rheMac2), hamadryas baboon (papHam1), marmoset (calJac3), mouse lemur (micMur1), and bushbaby (otoGar1), Additional coding sequences for colobus, tamarin, and vervet/African Green monkey were obtained from assembled exomes as referred to in George et al. (2011). Coding sequences and orthologous alignments were filtered and assembled using the methods in (96). Orthologous coding sequence alignments were generated for 13 primate species (where possible) of 1,170 human seminal fluid proteins (this study), 1090 human saliva proteins (97), and 1338 human plasma proteins (98).

Evolutionary analysis

A robust method was used to test for positive selection, which does not require any *a priori* knowledge by calculating the ratio of the number of nonsynonymous substitutions per nonsynonymous sites (d_N) to the number of synonymous substitutions per synonymous sites (d_S) (65). The ratio of $d_N/d_S = 1$ indicates that neutral evolution is occurring. When $d_N/d_S < 1$, this indicates that purifying selection (conserved evolution) is occurring. When $d_N/d_S > 1$, this indicates that positive selection (rapid evolution) is occurring. This method effectively distinguishes between drift and selection scenarios. The genome-wide d_N/d_S average for protein coding genes is 0.6. Maximum-likelihood analysis from the *codeml* program in the PAML

package were used to calculate d_N/d_S for seminal fluid, saliva, and plasma. Likelihood ratios (LR) were compared between neutral (M1, M7, M8a) and selection models (M2, M8) to identify positive selection acting on genes, and calculated p-values with FDR < 0.01. M8 identified specific codon sites under selection.

Analogous to identifying codon sites under selection, the branch-sites test was used to detect positive selection along particular lineages (foreground branches) (99,100). A LR test between an alternative model where the d_N/d_S ratio is fixed at 1 and a null model where the d_N/d_S ratio is fixed at 0 was used to detect selection. With branch-specific codon models, we grouped uni-male and multi-male mater lineages, and allowed the two groups to have different d_N/d_S values within our model. We alternated multi-male lineages as foreground and background branches, and calculated p-values with FDR < 0.01.

Evolutionary correlation

Two methods were used simultaneously to detect if a correlation between protein evolutionary rates and mating type exists: the branch-sites test and a phylogenetic model for estimating correlations. Measurements of continuous phenotypic characters were used to quantify primate mating types: binary classification into uni-male and multi-male mating systems, relative testis size (79), sexual size dimorphism (77), semen coagulation rating (7), and the mean number of sexual partners during an estrous period (81). Orthologous sequence alignments of the seminal fluid genes and mating behavior characters were the inputs for the correlation analysis. Non-reproductive datasets, saliva and plasma, were used as null comparisons. The branch-sites test is described above (Evolutionary analysis).

The phylogenetic model for estimating correlations was done with the software package Coevol 1.1 from (101). The Coevol program jointly models evolutionary rates of substitution

and phenotypic characters (i.e. relative testis size) as a multivariate Brownian diffusion process along the branches of a phylogenetic tree. The Coevol program corrects for the uncertainty about branch lengths and substitution history. Using Bayesian MCMC method and correcting for phylogenetic inertia using the independent contrast method, correlations between the rates of substitution and phenotypic characters are estimated with posterior probabilities (between 0 to 1). Posterior probabilities (pp) close to 0 indicated a negative correlation and close to 1 indicate a positive correlation. Strict cutoffs ($pp < 0.025$ and $pp > 0.975$) were used to reduce false positives. Summary statistics for all dataset results were analyzed with the R program.

Gene Loss

Comparing SFPs across species as an initial test identified putative lineage-specific gene loss events that may have occurred. Seminal fluid proteins not detected by MS does not indicate a gene loss event in itself because of the difficulty of detecting low abundance and hydrophobic proteins with MS. Putative gene loss events were validated by analyzing gene orthologs for early stop codons, frameshifts, and nonsense mutations. In addition, we searched for putative pseudogenization events by checking for frameshifts and premature stop codons in the orthologous sequence alignments for seminal fluid, saliva, and plasma. The number of putative pseudogenes between uni-male and multi-male lineages was used to calculate test statistics on the differences between the two groups.

Gene Ontology analysis

Two sources were used to determine Gene Ontological groupings, the in-house MSDaPl program (MacCoss lab) and the online source FatiGO (102,103). Each GO source had unique applications. Using MSDaPl, we were able to use GO Slim analysis on our identified proteins to

assess molecular function, biological process, and cellular component. Using FatiGO, we tested for overrepresentations of specific gene functions in human SFPs compared to the whole genome. We used orthologous human gene names to test for overrepresentations in the other primate species.

Results/Discussion

Seminal fluid protein composition and gene ontology

Seminal fluid proteins were identified in 8 primate species from diverse mating systems utilizing Liquid Chromatography-Mass Spectrometry (LC-MS) (Figure 14, samples used in proteomic analyses are highlighted). Two biological replicates per species were collected, with the exception of humans and rhesus macaques, where 8 biological replicates per species were collected. Three technical replicates per biological sample were run to avoid sampling bias, and the number of unique proteins identified in each biological replicate varied widely (mean peptide =1748, sd=±943, mean protein = 361, sd=±149) (Table 1). Overlapping proteins between biological replicates was consistent across all species (mean of overlap between biological replicates = 70%, sd=±7.24) (Figure 15). Overall, we identified the greatest number of unique proteins within humans (1136 proteins), and the least amount within the drill (157 proteins). Protein identification variation may be due to sample preparation, proteolysis, or MS instrumentation detection limits, but may also reflect inherent abundance differences in primate seminal fluid proteins. In particular, the drill samples had previously been cryogenically preserved which left an excess of glycogen and the MCX clean-up protocol may have eluted a significant amount of SFPs.

To investigate the gene ontology of the seminal fluid datasets, we used the in-house MSDaPl program (MacCoss lab). A large proportion of the SFPs' molecular function corresponds to the gene ontology (GO) terms binding (50.8%), protein binding (33.8%), and catalytic activity (27.5%). Using another online program FatiGO, we compared human SFPs to the whole genome to test for proteins with overrepresented molecular functions. 87.52% (1024/1170) of SFPs were annotated. SFPs were significantly overrepresented for the GO molecular functions: hydrolase activity, calcium ion binding, and carbohydrate binding (adjusted pvalue < 0.05). We used the orthologous gene names from other primates to also assess protein content and function in other species. Overall, the GO results were consistent with our human results and previous studies which showed that seminal fluid is a complex mixture of secreted proteins involved in binding and catalytic activity. In addition, using the online server SignalP 4.1 (http://www.cbs.dtu.dk/services/SignalP/), we detected 493 proteins with a signal peptide, 38 proteins with a transmembrane domain, and 134 proteins with a mitochondrion peptide.

Protein abundance within and between species

To ensure the accuracy of the RIA, as in many clinical studies to date, we used a 25% CV cutoff for each biological sample, each of which has 1-3 technical replicates. If only 1 technical replicate was present or the %CV was greater than 25%, the sample was excluded in subsequent analyses. We explored both interspecies and intraspecies variation. For both type of analyses, RIA values from identical internal standards across species were used to normalize the overall RIA data, which excluded some samples and species. For interspecies analysis, the RIA of identical internal standard peptides was detected in 5 species (human, drill, rhesus macaque, cynomolgus macaque, and vervet monkey) and, in total, we detected 7418 unique peptides with peaks present in 1 or more species. For the 3 remaining species (marmoset, baboon,

chimpanzee), we were not able to detect the same internal standards and therefore could not normalize them with the other species to calculate relative abundance. However, we were still able to use the species-specific data for intraspecies analysis. For intraspecies analysis, RIA values from identical internal standards were used to normalize the biological replicates within each species. Primarily, the human and rhesus macaque samples were analyzed since there were more biological replicates for each of these species. The number of quantifiable peptides identified in each species is shown in Table 1.

We identified the top 5 most abundant proteins in each of 8 primates species identified by Normalized Spectral Abundance Factor (NSAF) calculated with the MSDaPl program and also by relative abundance quantification scores. Interestingly, the most abundant proteins in all the primate species are involved in the copulatory plug pathway (SEMG1, SEMG2, TGM4, KLK3, ACPP). SEMG1, SEMG2, and TGM4 are involved in the formation of the copulatory plug, while KLK3 and ACPP are involved in the dissolution of the copulatory plug. At least one protein of the top 5 from each species is involved in the copulatory plug pathway so it appears that these proteins remain important constituents of seminal fluid even in uni-male mating systems. Another commonly abundant protein found was albumin (4 out of 8 species). Albumin is a major component of seminal fluid and is involved in preserving the sperm motility after ejaculation. A protein involved in immunosuppression, PIP, was also found in high abundance in multiple primate species. The copulatory plug pathway, immune response, and sperm motility are among the proteins most abundant in all primates we analyzed.

Within species, there was a lot of variation in peptide abundance between individuals. Primarily, we refer to human and rhesus macaque samples because we identified 8 individual's seminal fluid proteomes from both species. Using %CV as an indicator of high variation between

individuals, for human, the mean %CV of peptide abundance = 75%, sd = 26%. For rhesus macaque, mean %CV of peptide abundance = 116%, sd = 44%. While many of the SFPs had high variation between individuals, we were able to consistently identify proteins with conserved variation (i.e. QSOX1, CV=21%).

To assess protein abundance that varied significantly between humans and rhesus macaques, we used the Wilcoxon rank-sum test. This test reveled significant differences between humans and rhesus macaque abundances in 12 seminal fluid peptides, corresponding to 12 proteins, with all but 1 peptide that had human abundances larger than rhesus macaque abundances (Wilcoxon pvalues < 0.05). In particular, AZGP1 had a significantly greater abundance in humans compared to rhesus macaques. AZGP1 is involved in immune regulation, and has a similar structure to MHC-I and binds to many different substrates (104). Another protein, ACPP, had a greater abundance in human than rhesus macaque and, as previously mentioned, is involved in dissolving the copulatory plug. This is surprising, as humans do not have a prominent copulatory plug as do rhesus macaques. Perhaps ACPP ensures that seminal fluid retains a liquefied state upon ejaculation to ensure that the sperm is able to reach the egg.

In addition, we tested whether the two mating systems, uni-male and multi-male, protein abundance values were distributed differently from each other. Using the Wilcoxon rank-sum test, a nonparametric test, we compared the relative peptide abundances from uni-male and multi-male mating groups. This test revealed that out of the 7418 unique peptides across species, 40 peptides had abundances that are distributed differently between uni-male and multi-male mating groups (Wilcoxon pvalues < 0.05). Of the 40 significant peptides, 26 peptides had uni-male abundances that tended to be smaller than multi-male abundances (Wilcoxon pvalues < 0.05) and 14 peptides had uni-male abundances that tended to be larger than multi-male

abundances (Wilcoxon pvalues < 0.05). The 40 unique peptides correspond to 32 unique proteins. In particular, we highlight the TGM4 protein, which showed uni-male peptide abundances lower than the multi-male abundances (Figure 17A). TGM4 forms the copulatory plug along with the semenogelin proteins, and the protein is significantly lower in abundance in uni-male species. Four other proteins had more than one peptide per protein, including TGM4, which had 6 unique peptides. The Wilcoxon rank-sum test results were concordant for all peptides from the same protein, and the same pattern of relative peptide abundance between species was observed between different peptides from the same protein (Figure 17B). This gives us confidence that the ionization of peptides is not affecting RIA greatly.

Seminal fluid proteins are subject to rapid evolution

51 of the 1161 seminal fluid genes analyzed showed robust evidence of positive selection with significant LR test corrected for multiple testing (M8 vs. M8a; FDR < 0.01) (Table 3 and Supplemental Table 1). 1110 seminal fluid genes analyzed did not show any evidence of positive selection. Of the 51 genes under positive selection, 7 genes were previously shown to be positively selected in the rhesus macaque genome sequencing project (105). 6 of the 51 positively selected genes were also found among the top 5 highly abundant proteins in the 8 primate species (PIP, SLPI, SEMG2, MSMB, ACPP, and KLK3).

With the branch-sites test, we varied d_N/d_S between uni-male and multi-male mating lineages (Yang 2002, Zhang 2005). This partition reflects the relative intensity of sperm competition between species, with multi-male species generally experiencing more competition than uni-male species. We partitioned branches into foreground branches (multi-male) and background branches (uni-male). With this method, we identified 23 genes with significantly $d_N/d_S > 1$ on the multi-male lineages and $d_N/d_S = 0$ on uni-male lineages (pvalue < 0.01) (Table

3, Table 4, and Supplemental Table 1).

Is there a correlation between evolutionary rates and mating systems?

We jointly estimated the correlation of evolutionary rates to sexual characters using the program coevol (101). Posterior probabilities for each correlation were returned, and we used the stringent cutoffs for positive correlations ≥ 0.975 and for negative correlations ≤ 0.025 posterior probabilities. We identified 34 candidate genes with significant positive and negative correlations between d_N/d_S and 5 sexual characters (mating system, mean number of partners per estrous period, semen coagulation rating, relative testis size, and sexual size dimorphism)(Table 4). Many of the same genes were identified across sexual characters. 9/14 seminal fluid genes with positive correlations overlap with 3-4 other sexual characters. 15/21 with negative correlations overlap with 3-4 other sexual characters.

The molecular function of these candidate genes varies. There are candidate genes with clear reproductive functions, such as CRISP1, PATE, and AKAP4. CRISP1 is expressed in the testes and is a component of seminal fluid and sperm heads (106). CRISP family proteins include CRISP1, CRISP2, and CRISP3 and have been suggested to play an important role in sperm binding (107). PATE is a sperm-associated protein involved in sperm maturation and AKAP4 is found in the sperm flagellum involved in sperm motility (108,109). AKAP4 was found to be one of the most highly abundant protein in the rat and rhesus macaque sperm proteomes (110,111). This suggests that sperm proteins directly involved in sperm motility may experience elevated evolutionary rates, consistent with a previous study which showed that sperm swimming speed increases in more promiscuous primate species compared to monogamous primates (112). Many of the proteins identified in our evolutionary screen are associated with the sperm, consistent with the view that SFPs can have multiple uses on the sperm and in the seminal fluid.

Sample collection, shipping, or sperm-seminal fluid separation methods may have contaminated the seminal fluid with some sperm proteins. However, we identified highly abundant seminal fluid proteins and not abundant sperm proteins, so the additional proteins do not affect our abundance measurements.

Pseudogenization in seminal fluid proteins

Using the multiple gene alignments from seminal fluid, saliva, and plasma proteins, we searched for premature stop codons and frameshifts in each primate sequence compared to the human reference sequence. Due to the poor coding sequence quality in multiple species, alignments with a frameshift detected were discarded and only the premature stop codon data was used for further analysis. Despite limiting our analysis to gene alignments with no frameshifts, we found 32 putative pseudogenes in seminal fluid (Table 5). Compared to the number of putative pseudogenes in plasma (9) and saliva (9) proteins, we see a greater number of putative pseudogenes in the seminal fluid dataset. In order to determine if more of the pseudogenes occurred on the uni-male or multi-male lineages, we counted the number of pseudogenes occurring on each lineage and applied a correction for the number of species in each group. There was a significant overrepresentation of pseudogenes on the uni-male lineages compared to the multi-male lineages within the seminal fluid proteins when compared to saliva proteins (Pearson Chi-square test, $p<0.01$). We did not detect a significant difference in the number of pseudogenes between the uni-male and multi-male lineages with either the plasma or saliva datasets.

This suggests that uni-male lineages have more overall pseudogenization events than multi-male lineages. In the copulatory plug pathway, there is evidence that gene pseudogenization events have occurred in the TGM4, SEMG1, and SEMG2 proteins in the

gorilla and gibbon (both uni-male species) (82,113,114). Seminal fluid proteome-wide, we see this same general pattern of gene loss.

Conclusion

We present an example of the value of the combination of proteomics and evolutionary analysis in studying the effect of mating systems on seminal fluid protein evolution. Broadly, our study is the first to comprehensively characterize and compare seminal fluid proteins from a variety of different primates. We were then able to quantify and compare relative protein abundances across and within species. With our evolutionary analysis, we narrowed down candidate genes that show a correlation between evolutionary rates and sexual promiscuity. The general effect of sexual selection on seminal fluid protein regulation and expression has not been specifically studied in the context of mating system variation before, and we provide strong evidence that highly abundant proteins are also rapidly evolving in primates, and may be important indicators for how selection is acting on SFPs. In particular, we saw abundance patterns in which specific SFPs were low in abundance in uni-males compared to multi-males. This indicates that protein abundance differs between uni-males and multi-males and selective pressures may influence regulatory regions as opposed to coding regions. This novel insight into the regulatory mechanisms effect on seminal fluid protein abundances between species provides new candidate genes and approaches to studying the evolution of reproductive proteins.

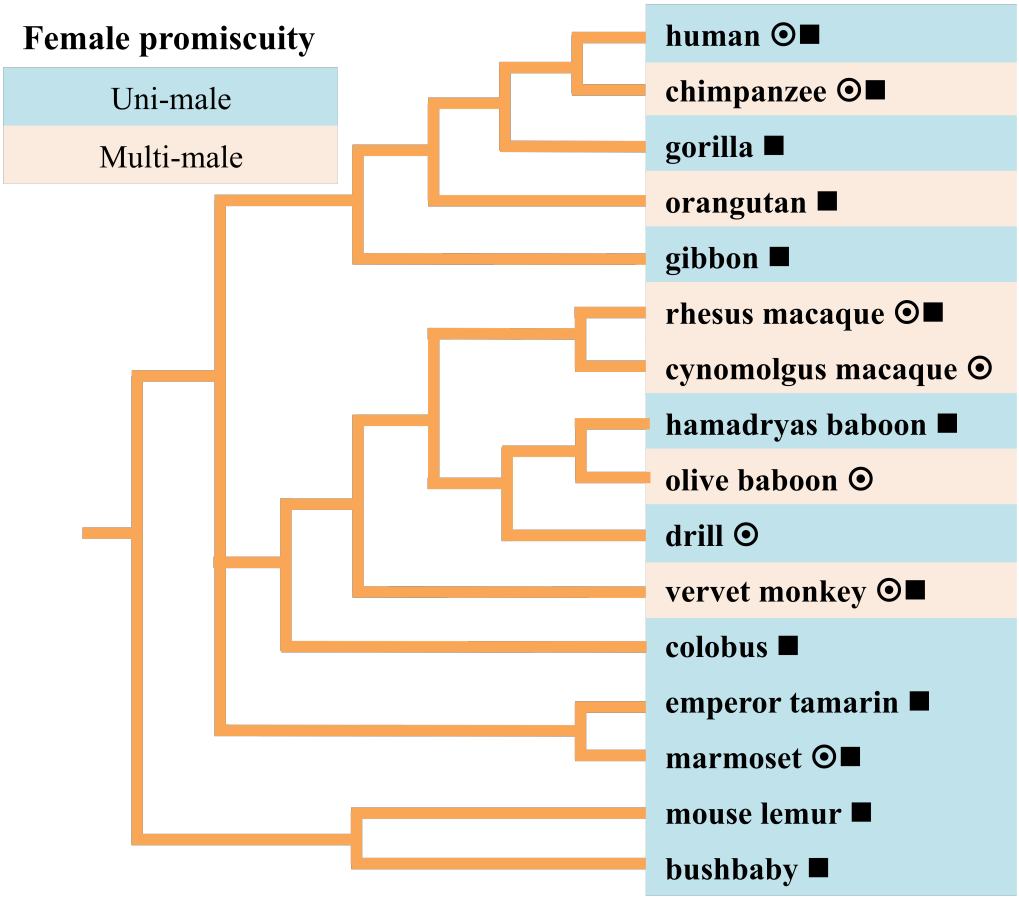


Figure 14. Primate mating systems and seminal fluid samples. Shading indicates uni-male or multi-male mating system designation. ⊙ indicates inclusion as a proteomic sample. ■ indicates inclusion in multiple sequence alignment from either genome reference coding sequence or exome sequencing in George et al. (2012).

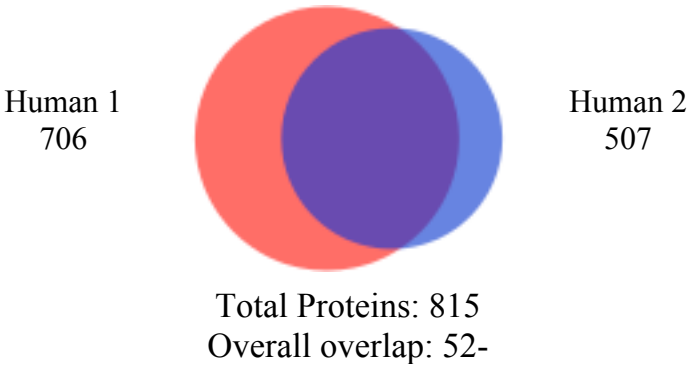


Figure 15. Venn diagram of protein overlap between 2 human biological samples.

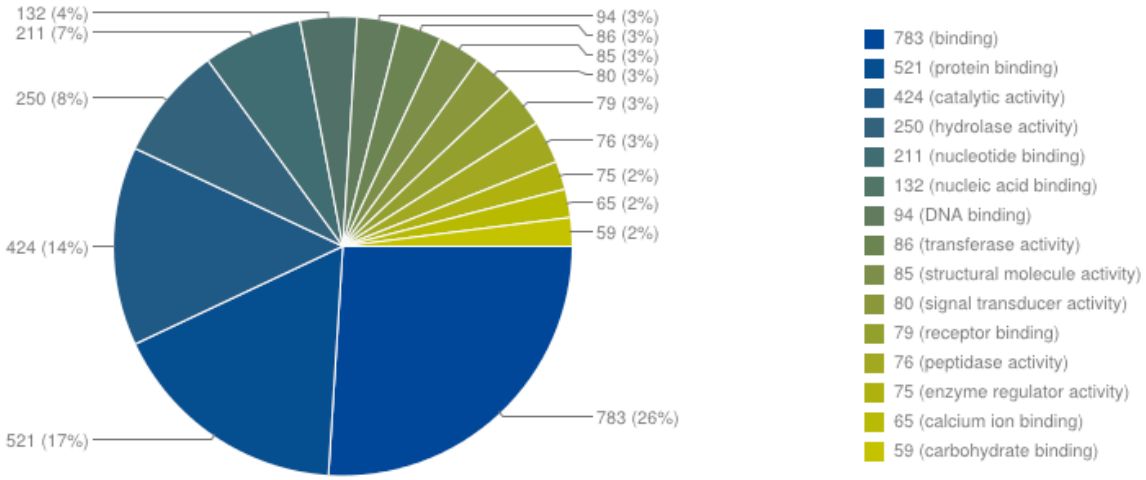


Figure 16. Gene Ontology of the molecular function of human seminal fluid. Total protein identifications taken from 8 biological human samples. 925 proteins were annotated. The top 15 terms are displayed, and there 41 terms total.

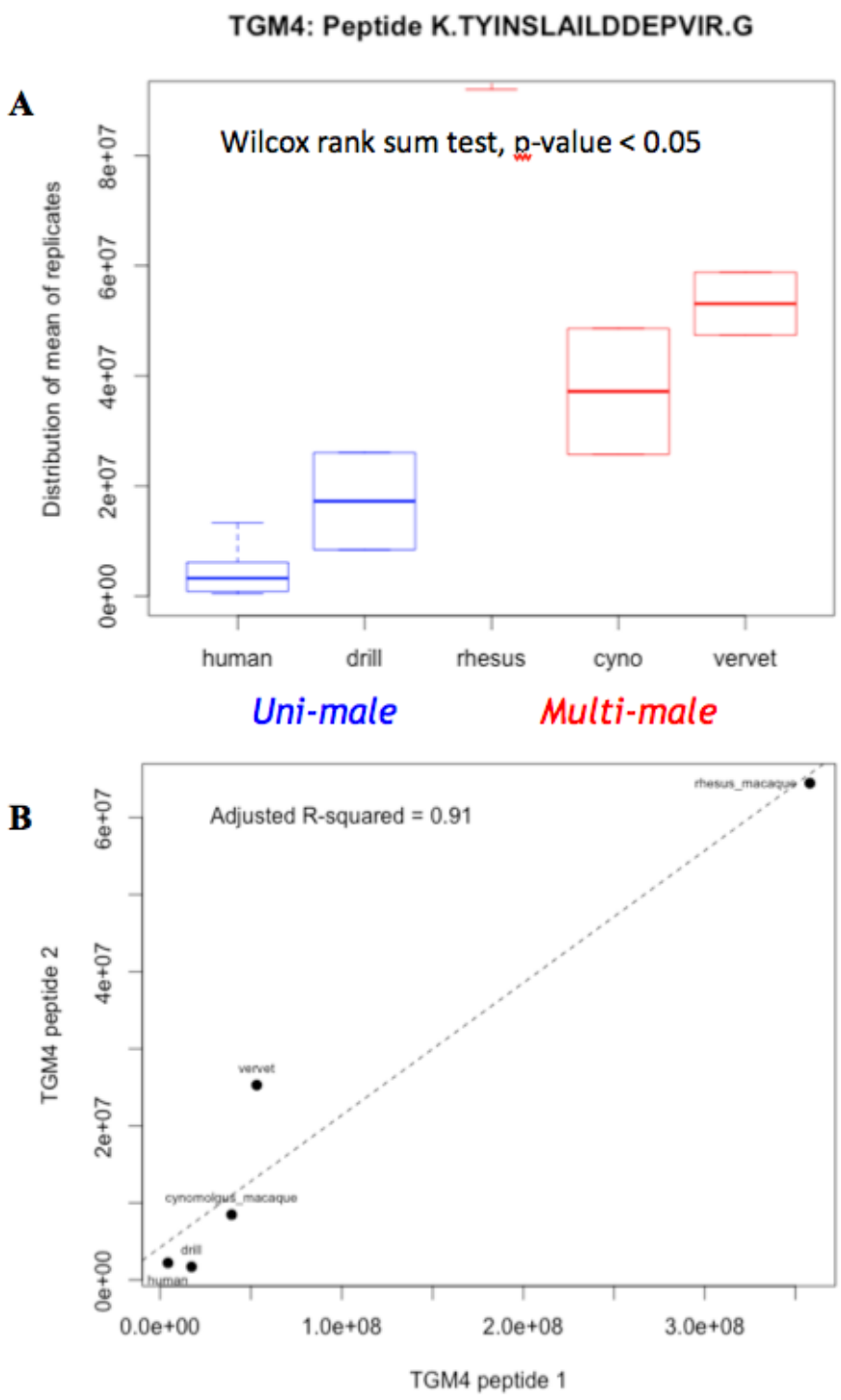


Figure 17. Quantitative proteomics. A) Significant differences in TGM4 protein abundances between uni-male and multi-male mating systems. B) TGM4 peptide concordance across species. Similar concordance seen in 5 other TGM4 peptides.

Table 1. Mass Spectrometry protein identification results. The total number of peptides/proteins identified includes all runs from each biological sample with a minimum of 1 peptide per protein with a high false discovery rate. Each biological replicate consists of 3 separate technical replicate runs. * Proteins were quantified using the TOPOGRAPH program (95).

Species	biological replicates	total number of runs	Total peptides identified	Total proteins identified	Total quantified peptides *
human	8	25	5707	1136	2188
chimpanzee	1	7	2275	464	1120
rhesus macaque	8	24	4480	736	5306
cynomolgus macaque	2	6	1338	331	357
drill	2	6	638	157	667
baboon	2	6	2003	437	2268
vervet monkey	2	6	1806	373	2012
marmoset	2	6	2311	441	24

Table 2. Top 5 abundant proteins in primate species.

Species	Transcript ID	Common Name	Coverage	NSAF	# Peptide
human	ENST00000372781	SEMG1	81.17	0.073588	247
	ENST00000372769	SEMG2	79.73	0.056651	288
	ENST00000291009	PIP	77.4	0.047171	51
	ENST00000351273	ACPP	68.9	0.032555	129
	ENST00000326003	KLK3	86.21	0.030361	115
chimpanzee	ENSPTRT00000025194	SEMG1	74.94	0.159899	109
	ENSPTRT00000027722	TGM4	83.92	0.053026	104
	ENSPTRT00000061981	SEMG2	57.35	0.028585	42
	ENSPTRT00000030078	ALB	68.1	0.027359	87
	ENSPTRT00000036677	PIP	71.23	0.024896	14
rhesus macaque	ENSMMUT00000046047	TGM4	78.12	0.060405	192
	ENSMMUT00000009192	NPC2	66.89	0.024963	42
	ENSMMUT00000005416	ALB	71.88	0.024152	121
	ENSMMUT00000041537	LCN2	58.5	0.022282	34
	ENSMMUT00000015692	SERPINA5	74.69	0.018228	59
cynomolgus macaque	ENSMMUT00000046047	TGM4	69.16	0.105581	97
	ENSMMUT00000038399	KLK3	80.84	0.068511	38
	ENSMMUT00000014459	SLPI	44.7	0.043827	7
	Contaminant	Trypsin	25.97	0.043257	20
	ENSMMUT00000012553	LYZ	40.54	0.033505	11
drill	ENSMMUT00000041537	LCN2	48	0.056088	16
	ENSMMUT00000015692	SERPINA5	63.14	0.053816	33
	ENSMMUT00000005416	ALB	65.3	0.046235	66
	ENSMMUT00000046047	TGM4	63.73	0.04525	49
	ENSMMUT00000009192	NPC2	61.59	0.033181	16
baboon	ENSMMUT00000046047	TGM4	71.07	0.095019	187
	ENSMMUT00000022739	ZG16B	57.99	0.034851	26
	ENSMMUT00000038399	KLK3	54.41	0.026286	41
	ENSMMUT00000005416	ALB	66.28	0.019534	74
	ENSMMUT00000008353	PIP	65.81	0.017841	13
vervet monkey	CCDS7235.1_1	MSMB	56.14	0.077368	19
	CCDS13346.1_1	SEMG2	53.95	0.061224	93
	CCDS12807.1_1	KLK3	71.65	0.03582	50
	CCDS13345.1_1	SEMG1	32.61	0.035368	56
	CCDS3540.1_1	SMR3B	65.82	0.035354	17
marmoset	ENSCJAT00000037357	SCGB2A1	75.27	0.05271	17
	ENSCJAT00000004068	LTF	74.37	0.037167	121

	ENSCJAT00000007443	DEFB1	58.82	0.033653	12
	ENSCJAT00000034191	SLPI	58.33	0.03303	20
	ENSCJAT00000034200	SEMG2	51.52	0.030814	57

Table 3. Summary of tests for positive selection. The Sites-test shows the results from the *codeml*'s Model 8a vs. Model 8 with a false discovery rate calculated by qvalues. The Branch-sites test shows the results from a likelihood ratio test where foreground and background branches are compared.

		Sites-test	Branch-sites test	
Dataset	**Total genes**	**M8a vs. M8 (FDR <0.01)**	**Foreground (Multi-male)**	**pvalue <0.01**
seminal fluid	1170	51	4	23
saliva	1080	66	4	24
plasma	1287	87	4	31

Table 4. Candidate genes from correlation and branch-sites analyses. Positive posterior probabilities from the coevol program are highlighted in 'green' and negative posterior probabilities are highlighted in 'yellow'. Genes with a significant branch-sites test are indicated with an 'x'. Non-significant values are marked with 'ns'.

CCDS	Transcript ID	Gene name	mating type (uni or multi)	mean number of partners	semen coagulat-ion	relative testis size	sexual size dimorp-hism	branch-sites
CCDS13927.1	ENST00000216181	MYH9	0.004	1.000	1.000	1.000	ns	x
CCDS4932.1	ENST00000335847	CRISP1	0.001	1.000	0.990	1.000	ns	ns
CCDS11192.1	ENST00000327031	FLII	0.016	0.990	ns	1.000	ns	ns
CCDS11061.1	ENST00000225655	PFN1	0.025	0.990	ns	1.000	ns	ns
CCDS2885.1	ENST00000295956	FLNB	0.010	0.990	0.980	0.990	ns	ns
CCDS10869.1	ENST00000268794	CDH1	0.003	0.990	0.980	0.990	ns	ns
CCDS4022.1	ENST00000261416	HEXB	0.007	1.000	ns	0.990	ns	ns
CCDS840.1	ENST00000369709	RAP1A	ns	ns	ns	0.980	ns	ns
CCDS11788.1	ENST00000269321	ARHGDIA	ns	ns	ns	0.980	ns	ns
CCDS31584.1	ENST00000378024	AHNAK	0.005	1.000	0.990	ns	ns	x
CCDS8440.1	ENST00000227378	HSPA8	0.011	0.990	0.980	ns	ns	x
CCDS1585.1	ENST00000366667	AGT	ns	0.980	ns	ns	ns	ns
CCDS32883.1	ENST00000245907	C3	ns	0.980	ns	ns	ns	ns
CCDS34209.1	ENST00000261483	MAN2A1	0.980	0.025	0.025	ns	ns	ns
CCDS8464.1	ENST00000305738	PATE	0.980	0.013	ns	ns	ns	ns
CCDS3125.1	ENST00000337777	PLS1	0.980	0.018	ns	ns	ns	ns
CCDS11400.1	ENST00000167586	KRT14	ns		ns	0.025	0.990	ns
CCDS2762.1	ENST00000296435	CAMP	ns	0.020	ns	0.025	ns	ns
CCDS31035.1	ENST00000366869	CAPN2	0.990	0.007	0.024	0.017	ns	ns
CCDS7299.1	ENST00000373232	PPA1	ns	ns	ns	0.016	ns	ns
CCDS12385.1	ENST00000222271	COMP	ns	ns	ns	0.015	ns	ns
CCDS33524.1	ENST00000284984	ADAMTS1	0.980	0.015	ns	0.015	ns	ns
CCDS34632.1	ENST00000381083	IGFBP3	0.990	0.013	ns	0.015	ns	ns
CCDS9927.1	ENST00000298841	SERPINA4	1.000	0.000	0.005	0.013	ns	ns
CCDS14330.1	ENST00000376064	AKAP4	0.980	0.015	0.018	0.012	ns	ns
CCDS9456.1	ENST00000377453	CLN5	0.990	0.011	0.015	0.011	ns	ns
CCDS10856.1	ENST00000268793	DPEP3	0.990	0.006	0.021	0.010	ns	ns
CCDS42064.1	ENST00000220166	CTSH	0.990	0.009	ns	0.009	ns	ns
CCDS1721.1	ENST00000380649	HADHA	1.000	0.002	0.009	0.008	ns	ns

CCDS10356.1	ENST00000300060	ANPEP	ns	0.016	ns	0.007	ns	ns
CCDS2991.1	ENST00000273371	PLA1A	0.990	0.007	ns	0.007	ns	ns
CCDS42992.1	ENST00000248923	GGT1	0.990	0.011	ns	0.007	ns	ns
CCDS6828.1	ENST00000373818	GSN	ns	0.017	ns	0.005	ns	ns
CCDS10721.1	ENST00000299138	VPS35	ns	0.024	ns	0.000	ns	ns
CCDS30861.1	ENST00000388718	FLG2	ns	ns	ns	ns	ns	x
CCDS7472.1	ENST00000266066	SFRP5	ns	ns	ns	ns	ns	x
CCDS42353.1	ENST00000333412	LRRC37A2	ns	ns	ns	ns	ns	x
CCDS34640.1	ENST00000275603	CCT6A	ns	ns	ns	ns	ns	x
CCDS14124.1	ENST00000217939	MXRA5	ns	ns	ns	ns	ns	x
CCDS11257.1	ENST00000225719	CPD	ns	ns	ns	ns	ns	x
CCDS3280.1	ENST00000232003	HRG	ns	ns	ns	ns	ns	x
CCDS8103.1	ENST00000301873	LTBP3	ns	ns	ns	ns	ns	x
CCDS34768.1	ENST00000291009	PIP	ns	ns	ns	ns	ns	x
CCDS33564.1	ENST00000332149	TMPRSS2	ns	ns	ns	ns	ns	x
CCDS93.1	ENST00000377493	PARK7	ns	ns	ns	ns	ns	x
CCDS10659.1	ENST00000308713	SEZ6L2	ns	ns	ns	ns	ns	x
CCDS11791.1	ENST00000331285	PCYT2	ns	ns	ns	ns	ns	x
CCDS43896.1	ENST00000372080	CEL	ns	ns	ns	ns	ns	x
CCDS13245.1	ENST00000216951	GSS	ns	ns	ns	ns	ns	x
CCDS2976.1	ENST00000273398	ATP6V1A	ns	ns	ns	ns	ns	x
CCDS3421.1	ENST00000281243	QDPR	ns	ns	ns	ns	ns	x
CCDS6545.1	ENST00000379405	PRSS3	ns	ns	ns	ns	ns	x
CCDS9557.1	ENST00000326783	FAM12B	ns	ns	ns	ns	ns	x
CCDS3508.1	ENST00000248701	SPINK2	ns	ns	ns	ns	ns	x

Table 5. Gene loss in seminal fluid, saliva, and plasma proteomes. Counts of the total number of premature stop codons are shown in the dataset. # indicates that these counts have been corrected for the number of lineages in each mating group.

Dataset	Total genes	uni-male lineages #	multi-male lineages #
seminal fluid	32	268.75	187.5
plasma	9	62.5	75
saliva	9	75	62.5

Chapter 4: Detecting coevolution in mammalian sperm-egg fusion proteins

Introduction

Mammalian fertilization is a complex process, involving direct and indirect interactions between proteins from male and female reproductive tracts. Sperm-egg interactions between proteins can occur on the sperm-zona pellucida (ZP) barrier and the sperm-egg cell membrane barrier (Figure 18). While candidates for interacting sperm-egg proteins abound, there are still no functionally confirmed interacting proteins on sperm and egg surfaces in mammals. Recent studies suggest that fertilization proteins may have compensatory functions, rendering proteins important but not essential for fertilization, which explains the lack of definitive functional evidence from gene deletion studies. Reproductive proteins, including those involved in sperm-egg interactions (ZP3, ADAM1, ADAM2, ACR, Fertilin α, Fertilin β, and CD9), have been shown to evolve rapidly by adaptive evolution between species (6). Studying the variation in the evolutionary rate of fertilization proteins can inform us about genetic interactions between these proteins, and in particular, can provide evidence of coevolution or shared function/expression over time between interacting sperm-egg fusion proteins within mammals.

Evolutionary rate covariation reflects the covariation of a pair of proteins over time, and has been often assumed to be a phylogenetic signature of physically interacting proteins. Clark et al. (2012) showed that a strong correlation in covariation can occur between proteins if they share a biological function or have coevolving expression levels (as can happen in specific tissues). The study also found strong correlations between functionally-verified physically interacting proteins, but these were a small subset of the overall protein data. More commonly found were strong correlations in evolutionary rates between proteins of similar function and

shared expression. Also, evolutionary rate covariation was uniformly distributed along the entire protein sequence, rather than clustering in specific interaction domains. When a similar test of covariation was applied to known interacting and rapidly evolving sperm-egg proteins in abalone, evidence of correlated evolutionary rates between the sperm protein lysin and egg protein VERL was observed (69). Previous studies show that lysin and VERL are rapidly evolving and this process may be driven by the compensatory mutations between the interacting proteins (6).

Fertilization proteins may experience constrained evolution in order to maintain the fidelity of the molecular interactions. Sexual conflict predicts that adaptive mutations in one interacting protein may select for compensatory mutations in its interacting partner. This coevolutionary process predicts a long-term correlation of evolutionary rates along phylogenetic lineages between interacting sperm and egg cell receptors. Thus, correlated rates of evolution can inform us *a prior* of which proteins interact physically or genetically, and can be a powerful statistical test to identify putative interacting proteins within the primate clade. Based on previous studies, we predict that rapid evolution and correlated evolutionary rates between sperm and egg proteins as an indicator of genetic interactions. We look for evidence of positive selection and test for molecular coevolution between putative interacting sperm-egg proteins using primate sequence divergence data.

On the egg cell membrane, CD9, has been proposed as a sperm receptor (115). On the sperm surface, epididymal protein DE (CRISP1), TXP2 (CRISP2), and Izumo have been identified as putative egg receptors (116-118). Gene deletion studies in mice have shown that these receptors show reductions in fertility, implying that they play an important role in the sperm-egg fusion process (119,120). The only receptors shown to be essential to sperm-egg

fusion by gene deletion are CD9 and Izumo (62,121-123). These results indicate that CD9 and Izumo are likely interactors, though there is no experimental data demonstrating that they physically interact, and CRISP1 and CRISP2 play an as yet fully characterized role in the sperm-egg fusion process.

CD9, CRISP1, CRISP2, and Izumo are strong candidate sperm and egg proteins involved in the sperm-egg fusion step of mammalian fertilization. CD9 is a tetraspanin containing four transmembrane domains and two extracellular loops of unequal size (124). It is highly expressed on the egg cell surface as well as several other types of hematopoietic and epithelial cells (125). CRISP1 and CRISP2 are members of the cysteine-rich secretory proteins, antigen 5 and pathogenesis-related 1 proteins (CAP) superfamily and each contain 16 conserved cysteine residues. CRISP1 is expressed exclusively in the epididymus and attaches loosely and tightly to the spermatozoa during epididymal maturation, and is proposed to be involved in both sperm-ZP binding and sperm-egg fusion through an egg-binding site, Signature 2 (106,126). CRISP2 is expressed in developing spermatids in the testes, and is able to bind to the egg cell surface (106,127). Izumo is a member of the immunoglobulin superfamily and contains an extracellular immunoglobulin domain. It is expressed specifically in the sperm and testis tissue (123). *CD9, CRISP1, CRISP2,* and *Izumo* were chosen as candidate interacting genes for our tests of selection and coevolution.

Methods

Sequencing and identification of sites under positive selection

The coding exons of *CD9, CRISP1, CRISP2, and Izumo* were Sanger sequenced from a panel of 12 primates, which included several species from the hominid, old world monkey and

new world monkey clades: *Pan troglodytes*, *Pan paniscus*, *Homo sapiens*, *Gorilla gorilla*, *Pongo pygmaeus*, *Macaca mulatta*, *Macaca nemestrina*, *Saguinus labiatus*, *Ateles geoffroyi*, *Lagothrix lagotricha*, *Erythrocebus patas*, and *Callithrix jacchus*. The proteins were first screened for evidence of positive selection, which is an observation seen in many other reproductive proteins, using d_N/d_S ratios (6). We used maximum likelihood methods from *codeml* of the PAML package (Model M7 vs. Model M8 and Model M8a vs. Model M8) to compare neutral models of codon evolution to selection models (128). Model M8 allows one to identify specific codon sites under selection. Codeml was run three times for each gene with different starting d_N/d_S values to account for gene convergence.

Correlation of evolutionary rates to predict interacting proteins

A predicted signature of coevolution is a long-term correlation of evolutionary rates between two genes (129,130). If divergence of one protein selects for a compensatory change in its partner, then an increase in the evolutionary rate in one protein should be matched by an increase along the same lineage in its partner to maintain their joint interaction during divergence. This process will result in a correlation of evolutionary rates when comparing the corresponding branches of each protein's phylogenetic tree. The similarity of the branch lengths can then be statistically evaluated and used to infer coevolution (Figure 19). This is referred to as the "mirror tree" approach and has been supported by multiple studies (130-132). The mirror tree method evaluates the intensity of coevolutionary behavior as Pearson's correlation coefficient, ñ, between a pair of distance matrices for two proteins (131). One of the main problems with this method is that a large number of false positives are often predicted. This is because the phylogenetic relationship behind the matrices can cause a high correlation between non-interacting proteins due to shared ancestry.

To control for correlations due to shared ancestry, we employ the method of (69). In this method, the degree of correlation is evaluated between d_N/d_S values, rather than genetic distance, for corresponding branches of the phylogenetic trees of two proteins. By dividing by the rate of synonymous substitutions, we can normalize for neutral divergence and in effect remove correlation due to shared ancestry. This method was evaluated using the well-characterized biochemical interaction between the sperm-egg proteins lysin and VERL in abalone (69).

Coevolution between CD9-Izumo, CD9-CRISP1, and CD9-CRISP2 was evaluated by testing for a correlation of evolutionary rates along phylogenetic lineages in our panel of 12 primates. Acceleration in the rate of change in *CD9* predicts a similar burst of selection in *Izumo* along the same lineage if they are under constraint to maintain their interaction. Hence, evolutionary rates would correlate between corresponding branches of *CD9* and interacting sperm protein phylogenies. The evolutionary rate for each branch was measured as the d_N/d_S ratio.

We first tested for a correlation of d_N/d_S values between egg and sperm proteins using a simple linear regression analysis with unweighted and weighted d_N/d_S values. We then tested for a correlation of d_N/d_S values between egg and sperm proteins using likelihood models to control for undefined branches. We compared a correlated model, which constrains the d_N/d_S values to fall on a best-fit line of regression, with a null model, which sets the slope of this line to zero to represent an uncorrelated relationship. In addition, we compared a free model, which allows d_N/d_S values to vary between the proteins, with the correlated model. As controls, we also tested for a correlation of d_N/d_S values between CD9 and two other rapidly evolving male

reproductive proteins, PKDREJ and ZAN (133,134). PKDREJ and ZAN are thought to be involved in ZP binding and so should be unable to physically bind CD9 during sperm-egg fusion.

Results

Evolutionary analysis of sperm-egg fusion genes

Two of the genes (*CRISP1* and *CRISP2*) screened showed evidence of positive selection with significant LTRs for Model M8a vs. M8. *CRISP1* showed evidence of sites under selection with $d_N/d_S > 1$, and these sites were resolved with Bayesian posterior probabilities > 95%. For *CRISP1*, we find evidence of positive selection for 7% of sites ($d_N/d_S = 3.77$), which corresponded to a site in exon 2, amino acid 80K. To determine if this d_N/d_S value of 3.77 is significantly different from 1, we used a neutral model where a class of sites is fixed at $d_N/d_S = 1$ (Model M8a) and compared that to a model of selection and find that this value is significantly different than $d_N/d_S = 1$ (Chi-square = 9.00, pvalue < 0.01) (135,136). *CRISP2* is evolving under positive selection with 4% of sites under selection ($d_N/d_S = 5.48$, Chi-square = 6.95, pvalue < 0.01), but no specific sites under selection were identified with Model M8. For our controls, we find evidence of positive selection for ZAN (Chi-square = 11.94, pvalue < 0.01) but not for PKDREJ (Chi-square = 2.58, pvalue = not significant). Previously, both genes were reported to be under selection, but this may change because of the number of species included in the analysis.

Analogous to identifying codon sites under selection, a Likelihood ratio test between a constant model where one d_N/d_S ratio is defined for all branches (fixed, Model 0) and free branch model with multiple d_N/d_S values per branch (free-ratio model) can identify lineages

under positive selection. This test identified two candidate genes in which the free model indicated that branches had multiple d_N/d_S values (*CD9* pvalue < 0.05 and *CRISP1* pvalue <0.01).

Coevolution between sperm-egg fusion genes

We tested for a correlation of d_N/d_S values for CD9-Izumo, CD9-CRISP1, and CD9-CRISP2 using a simple linear regression analysis. Branches with no synonymous substitutions were rounded up to the highest integer in the dataset. The linear regression produced the following correlations between the branch specific d_N/d_S values between CD9-CRISP1, CD9-CRISP2, and CD9-Izumo, respectively (Adjusted r^2 = -0.02, Adjusted r^2=0.06, and Adjusted r^2 = 0.11) (Figure 20, not showing CD9-CRISP2). A weighted linear regression was also done in which the total number of substitutions per codon values for specific branches were compared, thereby limiting the effects of undefined branches. The weighted linear regression produced the following correlations between CD9-CRISP1, CD9-CRISP2, and CD9-Izumo, respectively (Adjusted r^2 = -0.05, Adjusted r^2=0.17, and Adjusted r^2 = 0.31) (Figure 20, not showing CD9-CRISP2). T-tests indicated that only the unweighted CD9-Izumo (pvalue = 0.08) and the weighted CD9-CRISP2 (pvalue = 0.04) and CD9-Izumo (pvalue = 0.005) correlations predicted a correlation above the level of chance.

To improve the correlations without having to correct for the branches with no synonymous sites (undefined), we next used likelihood models to test the correlation of d_N/d_S values. We compared a correlated model, which constrains the d_N/d_S values to fall on a best-fit line of regression, with a null model, which sets the slope of this line to zero to represent an uncorrelated relationship. The correlated model was found to fit the data significantly better than

the null for CD9-CRISP1 and CD9-Izumo (pvalue < 0.001 and pvalue < 0.05, respectively) (Table 6). This is consistent with the idea that CD9 and IZUMO are coevolving, likely interacting, and may be subject to a similar selective pressures. CD9 and CRISP1 show a strong correlation between the proteins, and this supports the hypothesis that CRISP1 is also involved in the sperm-egg fusion process. None of our candidate genes showed significant evidence that the free model (separate branch models of codeml for each gene) fit the data better than the correlated model. Our controls showed no significant correlation of evolutionary rates for CD9 with either PKDREJ or ZAN (Table 6 and Figure 20C). This indicates that the correlated evolution between CD9 and the sperm proteins is not simply a general correlation seen between all reproductive proteins. We can use this type of coevolution test to investigate different stages of sperm-egg interaction with the ZP and the egg cell membrane, as both PKDREJ and ZAN are involved in the sperm-ZP binding. To further support that no correlation exists between the CD9 and ZAN, we found significant evidence that the free model fit the data better than the correlated model (pvalue <0.05). Our data implies that ZP interaction and fusion may be under different selective pressures.

Discussion

We identified positive selection acting on two of our candidate proteins (CRISP1 and CRISP2), and have added to the list of rapidly evolving reproductive proteins. Previously, CD9 had been identified as being under positive selection, but this signal may have been detected because of the low number of species used in the analysis (n=5) (66). With our inclusion of 12 primate species, we expect to have better resolution than the previous study. It is interesting that CD9 exhibited significant evidence of multiple primate lineages having different d_N/d_S values (free-ratio model), and this may indicate that CD9 is more important to fertilization in some

primates species than in others. As for Izumo, our results correspond to a previous study that found no evidence of selection for Izumo in the primate lineage, though evidence of positive selection was found in the Laurasiatheria group (137). The same study showed that Izumo gene family showed no consistent pattern of selection across the four Izumo genes, but Izumo3 gene showed a bout of positive selection within primates. It is unclear what drives the patterns of evolution of reproductive proteins, but it is likely that species-specific selection patterns for reproductive proteins have evolved over time. With the addition of more species and population data, we will be able to better study the long-term evolution of reproductive proteins.

CD9, CRISP1, CRISP2, and IZUMO interact in a complex network of proteins that facilitates sperm-egg fusion. We show evidence that the egg protein CD9 is interacting with sperm proteins, CRISP1 and Izumo, using a test of evolutionary rate correlation. It still remains uncertain how the proteins are genetically interacting together as previous studies have shown that correlated evolutionary rates can occur between proteins with the same function, tissue-specific expression, or which are physically interacting. Coevolution driven by sexual conflict between CD9 and CRISP1-Izumo predicts a strong correlation in evolutionary rate covariation. It may be that CRISP1 assists Izumo (or vice versa) in binding to the egg cell membrane, and CRISP1 is also required to change whenever Izumo changes in response to the amino acid changes in CD9. Alternatively, CD9 may be interacting separately with both CRISP1 and Izumo. We saw evidence of positive selection in CRISP1, and this can inform us about the rate of compensatory changes occurring between CD9 and CRISP1. We would predict that for every one amino acid change that occurs in CD9, there are multiple amino acid changes that occur in the CRISP1 protein based on the selection analysis and also the slope of the correlation line ($m = 0.67$). In contrast, Izumo is not evolving under positive selection and has a lower correlation

slope with CD9 (m = 0.25).

Our analysis suggests that CD9 is not interacting with sperm protein CRISP2 or the sperm proteins involved in ZP-binding (PKDREJ and ZAN). CRISP2 has been shown to be involved in the sperm-egg fusion process in previous studies and we show that it is evolving under positive selection. It is clear that some proteins may not be under long-term coevolution with another protein, or perhaps CRISP2 interacts with anther egg cell membrane protein. In addition, some reproductive proteins may experience bouts of selection along certain lineages and not overall within an entire clade. This may alter the signature of coevolution that we are measuring. It is also interesting to note that CD9 showed no evidence of correlation with the ZP binding sperm proteins. This suggests that no coevolution is occurring between CD9 and PKDREJ-ZAN, and more importantly, that sperm-egg fusion and ZP-binding proteins may be evolving at different rates. Indeed, for mammalian systems, evidence suggests that the ZP may play an important role in species-specificity because it is one of the first barriers the sperm needs to cross. In future studies, it would be of interest to compare the rate of evolution of ZP-binding proteins and sperm-egg fusion proteins. Within mammalian systems, it is difficult to elucidate interacting proteins functionally. Our method uses evolutionary correlations to predict interacting proteins, and provides candidates for further characterization.

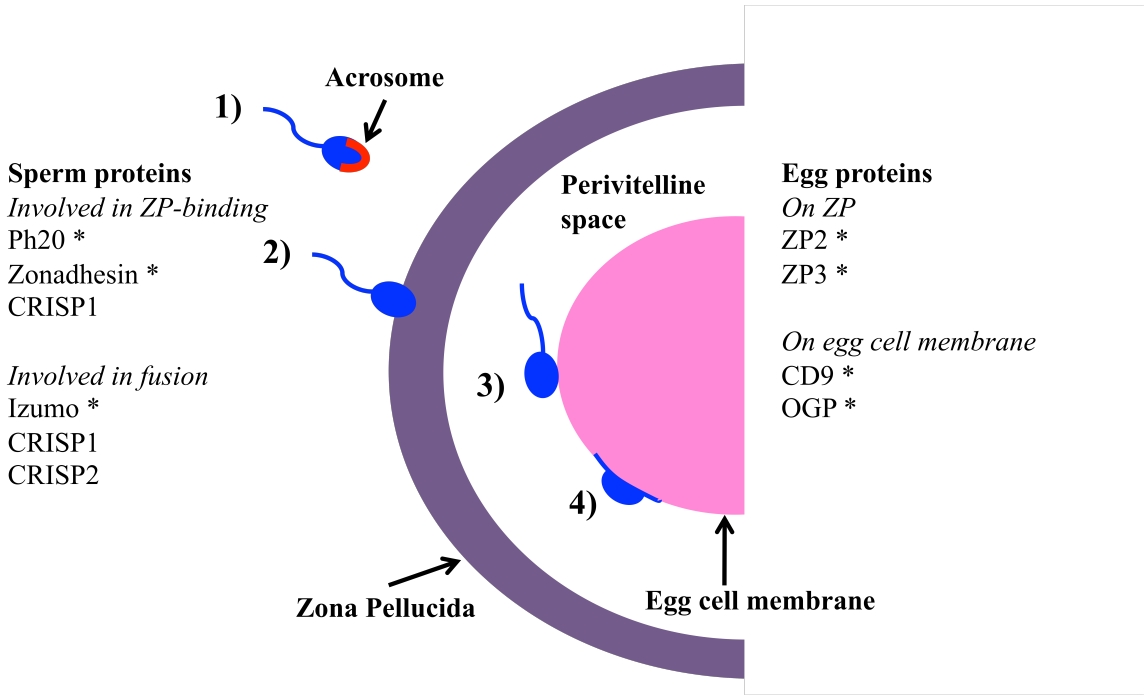


Figure 18. Process of mammalian fertilization. 1) Sperm swims toward the egg. 2) Sperm undergoes the acrosome reaction where the proteins in the acrosome are released and sperm is able to then cross the zona pellucida (ZP). 3) Sperm enters the perivitelline space and binds to the egg cell membrane. 4) Sperm and egg become cytoplasmically contiguous. * Indicates protein that has previously been shown to be under positive selection

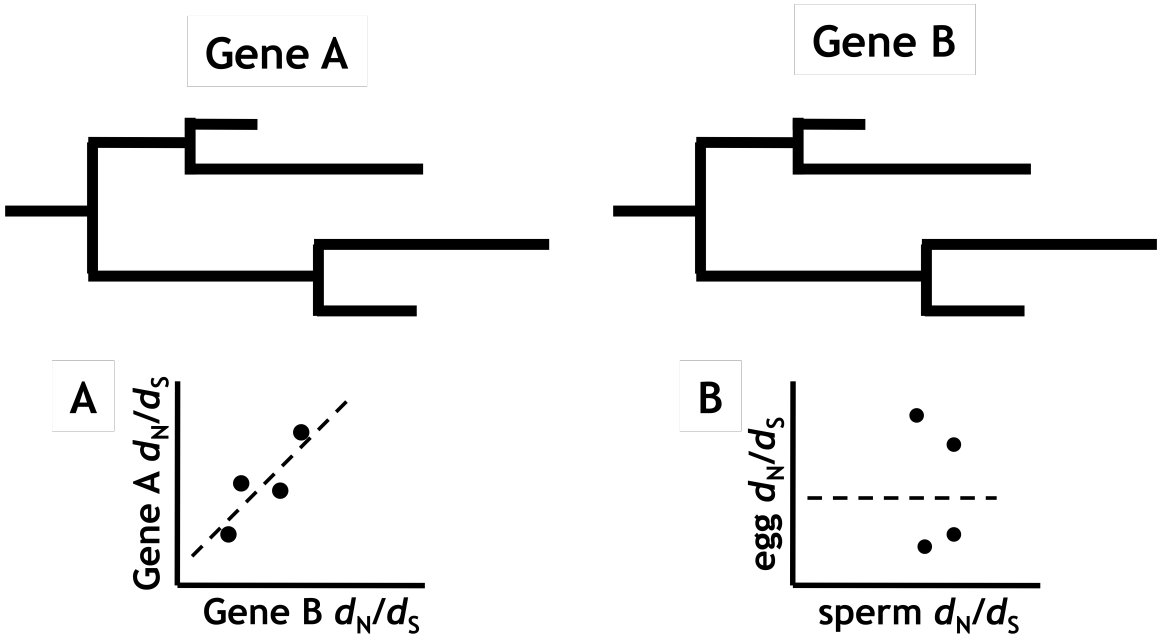
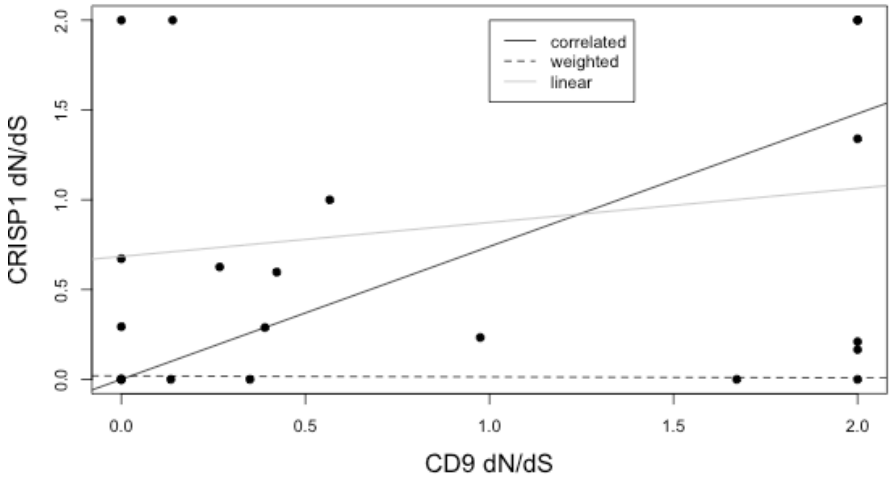
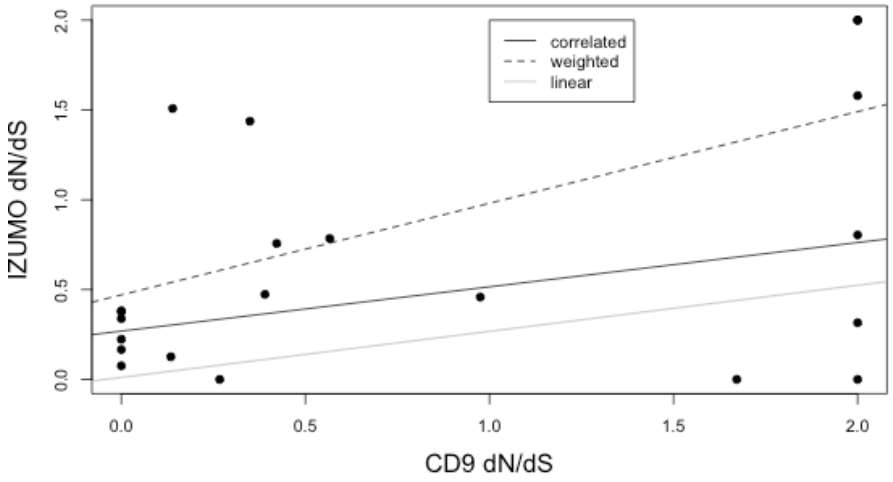


Figure 19. Detecting long-term coevolution using phylogenetics. Gene A and Gene B are potentially interacting proteins and their hypothetical phylogenetic tree determined by multiple sequence alignment is shown. A) Correlated branch lengths for interacting proteins. B) Uncorrelated branch lengths for non-interacting proteins.


correlated
weighted
linear
CRISP1 dN/dS
CD9 dN/dS
0.0
0.5
1.0
1.5
2.0


correlated
weighted
linear
IZUMO dN/dS
CD9 dN/dS
0.0
0.5
1.0
1.5
2.0

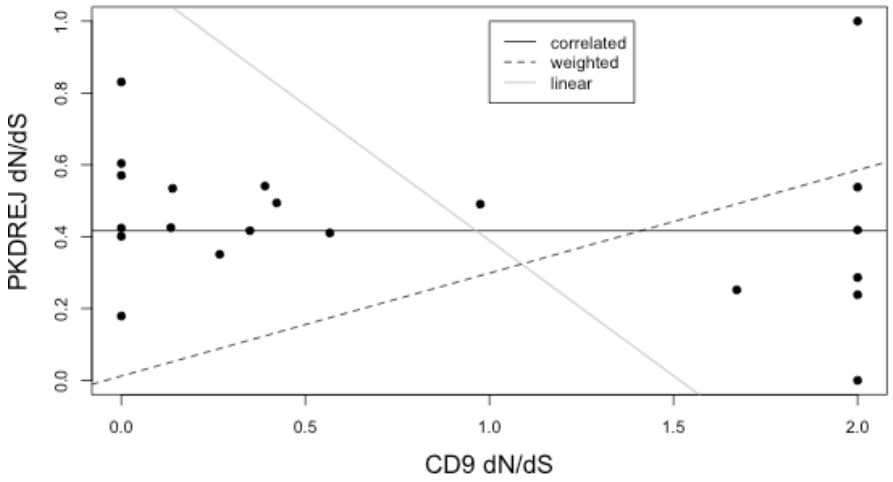

correlated
weighted
linear
PKDREJ dN/dS
CD9 dN/dS
0.0
0.2
0.4
0.6
0.8
1.0
0.5
1.0
1.5
2.0

Figure 20. Linear regressions of sperm-egg correlations. A) CD9 and CRISP1 d_N/d_S values were estimated on each branch of a 12 species phylogeny, plotted in the points. Linear regression for each correlation model are plotted: Correlation likelihood model (black line), standard linear regression (red line), weighted linear regression (green line) B) CD9 and Izumo d_N/d_S values were estimated on each branch of a 12 species phylogeny, plotted in the points. C) CD9 and PKDREJ d_N/d_S values were estimated on each branch of a 12 species phylogeny.

Table 6. Coevolution analysis results

		Likelihood model (lnL)			chi-squared (2*(lnL))	
	n	**free**	**correlated**	**null**	**free vs. correlated**	**correlated vs. null**
CD9-CRISP1	12	-2836.24	-2847.44	-2853.4	22.4	11.92 ***
CD9-CRISP2	12	-2619.51	-2627.95	-2627.36	16.88	1.18
CD9-Izumo	12	-3618.56	-3629.26	-3631.79	21.4	5.06 *
CD9-PKDREJ	12	-16131.3	-16142.9	-16139.6	23.2	ns
CD9-ZAN	11	-19493.5	-19509.6	-19509.6	32.2 *	0.00

pvalue < 0.05 *, pvalue < 0.01 **,pvalue < 0.001***

Chapter 5: Conclusions and Future Directions

Advances in genomic and proteomic technologies have provided new insights in reproductive research. Genomic technology enables us to sequence the entire genetic blueprint of organisms, from bacteria and viruses to platypuses and humans (138). The genome is the information storehouse for our bodies and encodes all the information we need to function. If DNA is the storehouse, then proteins are the workers that make things function. Proteomic technology such as mass spectrometry provides tools to identify proteins in specific tissues or cells and confirms the types of interactions that occur (139). The increasing ease of genomic sequencing has made many genomes available to test evolutionary hypotheses and study interesting organisms. Proteomic methods have enabled researchers to identify new proteins in sperm, egg, seminal fluid and follicular fluid, which may be important for identifying the causes of infertility.

In this dissertation, I combined proteomics and evolutionary analyses to study the evolution of reproductive proteins in primates. I used proteomics to identify and quantify known and novel seminal fluid proteins, and identified candidate proteins affected by sperm competition and involved in the sperm-egg fusion process. To elucidate the effects of sexual conflict effect, I further studied the coevolution of sperm-egg fusion proteins. In this chapter, I comment on how we can use genetic data to disentangle the various hypotheses that attempt to explain why rapid evolution is prevalent in reproductive proteins, how evolutionary proteomics can be of value to evolutionary studies, and future directions for the functional verification of interacting reproductive proteins.

Detecting the causes behind rapidly diverging reproductive proteins

As the genomes of many organisms were sequenced across distantly diverged taxa, it became apparent that many rapidly evolving proteins had reproductive functions (113,140,141). This presented an interesting paradox: reproductive proteins should be extremely conserved because of their fertilization and protective properties, but a high proportion of reproductive proteins are rapidly evolving when compared to general protein coding genes. Various hypotheses were posed to explain the rapid evolution in these proteins (sexual conflict, sexual selection, sperm competition, and hybrid reinforcement, among others). We can use coding sequence data to test if specific scenarios that have occurred over evolutionary time. In particular, within sperm-egg fusion it is predicted that intersexual conflict occurs when the mating rate differs between the sexes and it thus results in a continuous adaptive struggle between males and females. This is analogous to the Red Queen hypothesis proposed for host-pathogen interactions because the interaction between males and females (mating rate) does not change despite the rapid evolution of each sex in response to conflict.

We address hypotheses about how sperm competition affects the rapid evolution of seminal fluid proteins and how sexual conflict can cause sperm-egg fusion protein coevolution. Rapid evolution can also result in fixed amino acid changes or population-level polymorphisms. Future research should use more primate sequences to improve the detection of positive selection, and incorporate more population-level data. With the current number of available genome sequences from different primate species, there is still low detection of positively selected genes. Having higher quality and better-annotated genome/exome sequences would aid in our search for putative pseudogenization events. The previous gene loss analysis in Chapter 3 was severely limited by the presence of frameshifts in the low quality primate sequences and

those sequences were excluded because of potentially false pseudogenization events. Having higher quality sequences would enable us to identify more pseudogenization events in primate lineages. In addition, some of the primates in our proteomic protein identification analysis did not have coding sequence available to search the peptide identifications against. Instead, I used the next closest evolutionary relative to the species in question. Although searching with a related relative does not lower the number of protein identifications significantly, having the coding sequences available for each primate species ensures that we identify the maximum amount of proteins in the proteomic data, specifically from proteins that are unique within a species or rapidly evolving.

Using proteomics for comparative studies

As the identification of large numbers of proteins becomes easier with the development of superior sample isolation and high-throughput separation and identification techniques, the need to develop methods for quantifying the absolute abundance of proteins in complex mixtures becomes more important (142-145). Various methods have been developed for the absolute quantification of proteins, global stable isotope labeling approach and isotope-coded affinity tag (reviewed in (146)), but these methods are costly and inefficient when trying to quantify complex or whole proteomes.

When studying the long-term evolution of a specific protein, coding sequence evolution is frequently the only parameter considered. Proteomics can give a wealth of information about the abundance and regulation of the protein. Proteomics can directly measure the abundance of a protein in a cell or mixture. While measuring mRNA transcript levels and expression can give us some idea of the amount of protein in a cell, these methods do not directly measure the amount of protein in a sample, as it is known that translation levels do not directly reflect the amount of a

protein. With protein abundance measurements, we can assess if amino acid coding changes have affected protein abundance levels across species. If changes at the amino acid level in coding sequence are absent but we see drastic differences in the protein abundances across species, this indicates regulatory changes occurring.

Using a relative quantification method such as we have presented, we can identify candidate genes that show significant abundance differences between species and study regulatory regions instead of the coding regions for certain proteins. Absolute abundance methods can then be used to target a smaller subset of these candidate genes.

Future directions for functionally verifying interacting reproductive proteins

It is clear that the field of reproductive biology has made significant strides in the past decade with the availability of whole genome sequences and improvements in technology (gene expression, bioinformatics, in vitro fertilization, follicular culture and in vitro maturation, proteomics). The fertilization paradigm involving sperm-egg interactions has become more complicated, and new evidence suggests that a multitude of proteins are involved in facilitating sperm-egg interactions and this varies considerably between species. In terms of the evolution of the egg, its interaction with the sperm plays a large part in its evolution. Revealing the proteins composing the egg structure is only the first step; functional characterization is crucial for understanding interactions.

We identified coevolving and putatively interacting sperm-egg fusion genes. In order to verify the genetic interactions, functional tests are necessary. Verification of gene expression and localization in sperm or seminal fluid, identification of any interacting proteins from males or females, potential protein modifications on the mature protein, and identifying the functional consequences of amino acid changes across species are all future studies that can be done on

candidate proteins. Competition assays with other species gene variants would be useful to study species-specificity.

This dissertation work points to several directions for future study and raises some questions about interacting reproductive proteins. The methods I use can comprehensively identify reproductive proteins from multiple species or tissues. In our study, we are only looking at one side of the puzzle, the male side, and there is a need to study both female and male reproductive proteins. To this end, there has been work to characterize human follicular fluid (86). It will be important to not only identify new proteins but to study how male and female proteins interact in the reproductive tract, not only during sperm-egg fusion, but also in chemotaxis, sperm storage, and throughout the fertilization process. We are only at the tip of the iceberg in understanding how sex influences biological diversity.

References

(1) Pitnick S, Spicer GS, Markow TA. How long is a giant sperm? Nature 1995 May 11;375(6527):109.

(2) Brennan PL, Clark CJ, Prum RO. Explosive eversion and functional morphology of the duck penis supports sexual conflict in waterfowl genitalia. Proc Biol Sci 2010 May 7;277(1686):1309-1314.

(3) Smith RL. Sperm competition and the evolution of animal mating systems. Orlando, Florida: Academic Press, Inc.; 1984.

(4) Zuccotti M, Merico V, Cecconi S, Redi CA, Garagna S. What does it take to make a developmentally competent mammalian egg? Hum Reprod Update 2011 Jul-Aug;17(4):525-540.

(5) Claw KG, Swanson WJ. Evolution of the egg: new findings and challenges. Annu Rev Genomics Hum Genet 2012;13:109-125.

(6) Swanson WJ, Vacquier VD. The rapid evolution of reproductive proteins.. Nat Rev Genet 2002 Feb;3(2):137-44.

(7) Dixson AL, Anderson MJ. Sexual selection, seminal coagulation and copulatory plug formation in primates. Folia Primatol (Basel) 2002 Mar-Jun;73(2-3):63-69.

(8) Wassarman PM, Litscher ES. Mammalian fertilization: the egg's multifunctional zona pellucida. Int J Dev Biol 2008;52(5-6):665-676.

(9) Findlay GD, MacCoss MJ, Swanson WJ. Proteomic discovery of previously unannotated, rapidly evolving seminal fluid genes in Drosophila. Genome Res 2009 May;19(5):886-896.

(10) Aagaard JE, Yi X, MacCoss MJ, Swanson WJ. Rapidly evolving zona pellucida domain proteins are a major component of the vitelline envelope of abalone eggs. Proc Natl Acad Sci U S A 2006 Nov 14;103(46):17302-17307.

(11) Pang PC, Chiu PC, Lee CL, Chang LY, Panico M, Morris HR, et al. Human sperm binding is mediated by the sialyl-Lewis(x) oligosaccharide on the zona pellucida. Science 2011 Sep 23;333(6050):1761-1764.

(12) Russell DL, Salustri A. Extracellular matrix of the cumulus-oocyte complex. Semin Reprod Med 2006 Sep;24(4):217-227.

(13) Myles DG, Primakoff P. Why did the sperm cross the cumulus? To get to the oocyte. Functions of the sperm surface proteins PH-20 and fertilin in arriving at, and fusing with, the egg. Biol Reprod 1997 Feb;56(2):320-327.

(14) Salustri A, Garlanda C, Hirsch E, De Acetis M, Maccagno A, Bottazzi B, et al. PTX3 plays a key role in the organization of the cumulus oophorus extracellular matrix and in in vivo fertilization. Development 2004 Apr;131(7):1577-1586.

(15) Shimada M, Yanai Y, Okazaki T, Noma N, Kawashima I, Mori T, et al. Hyaluronan fragments generated by sperm-secreted hyaluronidase stimulate cytokine/chemokine production via the TLR2 and TLR4 pathway in cumulus cells of ovulated COCs, which may enhance fertilization. Development 2008 Jun;135(11):2001-2011.

(16) Tanii I, Aradate T, Matsuda K, Komiya A, Fuse H. PACAP-mediated sperm-cumulus cell interaction promotes fertilization. Reproduction 2011 Feb;141(2):163-171.

(17) Sun TT, Chung CM, Chan HC. Acrosome reaction in the cumulus oophorus revisited: involvement of a novel sperm-released factor NYD-SP8. Protein Cell 2011 Feb;2(2):92-98.

(18) Inoue N, Satouh Y, Ikawa M, Okabe M, Yanagimachi R. Acrosome-reacted mouse spermatozoa recovered from the perivitelline space can fertilize other eggs. Proc Natl Acad Sci U S A 2011 Dec 13;108(50):20008-20011.

(19) Avella MA, Dean J. Fertilization with acrosome-reacted mouse sperm: implications for the site of exocytosis. Proc Natl Acad Sci U S A 2011 Dec 13;108(50):19843-19844.

(20) Jin M, Fujiwara E, Kakiuchi Y, Okabe M, Satouh Y, Baba SA, et al. Most fertilizing mouse spermatozoa begin their acrosome reaction before contact with the zona pellucida during in vitro fertilization. Proc Natl Acad Sci U S A 2011 Mar 22;108(12):4892-4896.

(21) Jeon BG, Moon JS, Kim KC, Lee HJ, Choe SY, Rho GJ. Follicular fluid enhances sperm attraction and its motility in human. J Assist Reprod Genet 2001 Aug;18(8):407-412.

(22) Oren-Benaroya R, Orvieto R, Gakamsky A, Pinchasov M, Eisenbach M. The sperm chemoattractant secreted from human cumulus cells is progesterone. Hum Reprod 2008 Oct;23(10):2339-2345.

(23) Wang Y, Storeng R, Dale PO, Abyholm T, Tanbo T. Effects of follicular fluid and steroid hormones on chemotaxis and motility of human spermatozoa in vitro. Gynecol Endocrinol 2001 Aug;15(4):286-292.

(24) Lin Y, Mahan K, Lathrop WF, Myles DG, Primakoff P. A hyaluronidase activity of the sperm plasma membrane protein PH-20 enables sperm to penetrate the cumulus cell layer surrounding the egg. J Cell Biol 1994 Jun;125(5):1157-1163.

(25) Kim E, Yamashita M, Kimura M, Honda A, Kashiwabara S, Baba T. Sperm penetration through cumulus mass and zona pellucida. Int J Dev Biol 2008;52(5-6):677-682.

(26) Baba D, Kashiwabara S, Honda A, Yamagata K, Wu Q, Ikawa M, et al. Mouse sperm lacking cell surface hyaluronidase PH-20 can pass through the layer of cumulus cells and fertilize the egg. J Biol Chem 2002 Aug 16;277(33):30310-30314.

(27) Alves AP, Mulloy B, Moy GW, Vacquier VD, Mourao PA. Females of the sea urchin Strongylocentrotus purpuratus differ in the structures of their egg jelly sulfated fucans. Glycobiology 1998 Sep;8(9):939-946.

(28) Vilela-Silva AC, Castro MO, Valente AP, Biermann CH, Mourao PA. Sulfated fucans from the egg jellies of the closely related sea urchins Strongylocentrotus droebachiensis and Strongylocentrotus pallidus ensure species-specific fertilization. J Biol Chem 2002 Jan 4;277(1):379-387.

(29) Vilela-Silva AC, Alves AP, Valente AP, Vacquier VD, Mourao PA. Structure of the sulfated alpha-L-fucan from the egg jelly coat of the sea urchin Strongylocentrotus franciscanus: patterns of preferential 2-O- and 4-O-sulfation determine sperm cell recognition. Glycobiology 1999 Sep;9(9):927-933.

(30) Monne M, Han L, Schwend T, Burendahl S, Jovine L. Crystal structure of the ZP-N domain of ZP3 reveals the core fold of animal egg coats. Nature 2008 Dec 4;456(7222):653-657.

(31) Goudet G, Mugnier S, Callebaut I, Monget P. Phylogenetic analysis and identification of pseudogenes reveal a progressive loss of zona pellucida genes during evolution of vertebrates. Biol Reprod 2008 May;78(5):796-806.

(32) Swanson WJ, Aagaard JE, Vacquier VD, Monne M, Sadat Al Hosseini H, Jovine L. The molecular basis of sex: linking yeast to human. Mol Biol Evol 2011 Jul;28(7):1963-1966.

(33) Harris JD, Hibler DW, Fontenot GK, Hsu KT, Yurewicz EC, Sacco AG. Cloning and characterization of zona pellucida genes and cDNAs from a variety of mammalian species: the ZPA, ZPB and ZPC gene families. DNA Seq 1994;4(6):361-393.

(34) Hughes DC, Barratt CL. Identification of the true human orthologue of the mouse Zp1 gene: evidence for greater complexity in the mammalian zona pellucida? Biochim Biophys Acta 1999 Oct 28;1447(2-3):303-306.

(35) Rankin T, Familari M, Lee E, Ginsberg A, Dwyer N, Blanchette-Mackie J, et al. Mice homozygous for an insertional mutation in the Zp3 gene lack a zona pellucida and are infertile. Development 1996 Sep;122(9):2903-2910.

(36) Liu C, Litscher ES, Mortillo S, Sakai Y, Kinloch RA, Stewart CL, et al. Targeted disruption of the mZP3 gene results in production of eggs lacking a zona pellucida and infertility in female mice. Proc Natl Acad Sci U S A 1996 May 28;93(11):5431-5436.

(37) Swanson WJ, Yang Z, Wolfner MF, Aquadro CF. Positive Darwinian selection drives the evolution of several female reproductive proteins in mammals. Proc Natl Acad Sci U S A 2001 Feb 27;98(5):2509-2514.

(38) Higgie M, Chenoweth S, Blows MW. Natural selection and the reinforcement of mate recognition. Science 2000 Oct 20;290(5491):519-521.

(39) Hill KL, L'Hernault SW. Analyses of reproductive interactions that occur after heterospecific matings within the genus Caenorhabditis. Dev Biol 2001 Apr 1;232(1):105-114.

(40) O'Rand MG. Sperm-egg recognition and barriers to interspecies fertilization. Gamete Res 1988 Apr;19(4):315-328.

(41) Palumbi SR. All males are not created equal: fertility differences depend on gamete recognition polymorphisms in sea urchins. Proc Natl Acad Sci U S A 1999 Oct 26;96(22):12632-12637.

(42) Rankin T, Talbot P, Lee E, Dean J. Abnormal zonae pellucidae in mice lacking ZP1 result in early embryonic loss. Development 1999 Sep;126(17):3847-3855.

(43) Hoodbhoy T, Joshi S, Boja ES, Williams SA, Stanley P, Dean J. Human sperm do not bind to rat zonae pellucidae despite the presence of four homologous glycoproteins. J Biol Chem 2005 Apr 1;280(13):12721-12731.

(44) Gahlay GK, Srivastava N, Govind CK, Gupta SK. Primate recombinant zona pellucida proteins expressed in Escherichia coli bind to spermatozoa. J Reprod Immunol 2002 Jan;53(1-2):67-77.

(45) Wassarman PM, Jovine L, Litscher ES, Qi H, Williams Z. Egg-sperm interactions at fertilization in mammals. Eur J Obstet Gynecol Reprod Biol 2004 Jul 1;115 Suppl 1:S57-60.

(46) Wassarman PM. Mammalian fertilization: molecular aspects of gamete adhesion, exocytosis, and fusion. Cell 1999 Jan 22;96(2):175-183.

(47) Florman HM, Wassarman PM. O-linked oligosaccharides of mouse egg ZP3 account for its sperm receptor activity. Cell 1985 May;41(1):313-324.

(48) Bleil JD, Wassarman PM. Structure and function of the zona pellucida: identification and characterization of the proteins of the mouse oocyte's zona pellucida. Dev Biol 1980 Apr;76(1):185-202.

(49) Han L, Monne M, Okumura H, Schwend T, Cherry AL, Flot D, et al. Insights into egg coat assembly and egg-sperm interaction from the X-ray structure of full-length ZP3. Cell 2010 Oct 29;143(3):404-415.

(50) Chen J, Litscher ES, Wassarman PM. Inactivation of the mouse sperm receptor, mZP3, by site-directed mutagenesis of individual serine residues located at the combining site for sperm. Proc Natl Acad Sci U S A 1998 May 26;95(11):6193-6197.

(51) Boja ES, Hoodbhoy T, Fales HM, Dean J. Structural characterization of native mouse zona pellucida proteins using mass spectrometry. J Biol Chem 2003 Sep 5;278(36):34189-34202.

(52) Gahlay G, Gauthier L, Baibakov B, Epifano O, Dean J. Gamete recognition in mice depends on the cleavage status of an egg's zona pellucida protein. Science 2010 Jul 9;329(5988):216-219.

(53) Aagaard JE, Vacquier VD, MacCoss MJ, Swanson WJ. ZP domain proteins in the abalone egg coat include a paralog of VERL under positive selection that binds lysin and 18-kDa sperm proteins. Mol Biol Evol 2010 Jan;27(1):193-203.

(54) Kamei N, Glabe CG. The species-specific egg receptor for sea urchin sperm adhesion is EBR1,a novel ADAMTS protein. Genes Dev 2003 Oct 15;17(20):2502-2507.

(55) Swanson WJ, Vacquier VD. The abalone egg vitelline envelope receptor for sperm lysin is a giant multivalent molecule. Proc Natl Acad Sci U S A 1997 Jun 24;94(13):6724-6729.

(56) Lewis CA, Talbot CF, Vacquier VD. A protein from abalone sperm dissolves the egg vitelline layer by a nonenzymatic mechanism. Dev Biol 1982 Jul;92(1):227-239.

(57) Swanson WJ, Vacquier VD. The abalone egg vitelline envelope receptor for sperm lysin is a giant multivalent molecule. Proc Natl Acad Sci U S A 1997 Jun 24;94(13):6724-6729.

(58) Galindo BE, Vacquier VD, Swanson WJ. Positive selection in the egg receptor for abalone sperm lysin. Proc Natl Acad Sci U S A 2003 Apr 15;100(8):4639-4643.

(59) Chapman RF. *The Insects: Structure and Function*. Third ed. Cambridge, MA: Harvard University Press; 1982.

(60) Wassarman PM, Jovine L, Litscher ES. A profile of fertilization in mammals. Nat Cell Biol 2001 Feb;3(2):E59-64.

(61) Evans JP. Sperm-Egg Interaction. Annu Rev Physiol 2011 Feb 15.

(62) Le Naour F, Rubinstein E, Jasmin C, Prenant M, Boucheix C. Severely reduced female fertility in CD9-deficient mice. Science 2000 Jan 14;287(5451):319-321.

(63) Rubinstein E, Ziyyat A, Prenant M, Wrobel E, Wolf JP, Levy S, et al. Reduced fertility of female mice lacking CD81. Dev Biol 2006 Feb 15;290(2):351-358.

(64) Baessler KA, Lee Y, Sampson NS. Beta1 integrin is an adhesion protein for sperm binding to eggs. ACS Chem Biol 2009 May 15;4(5):357-366.

(65) Yang Z. PAML: a program package for phylogenetic analysis by maximum likelihood. Comput Appl Biosci 1997 Oct;13(5):555-556.

(66) Swanson WJ, Nielsen R, Yang Q. Pervasive adaptive evolution in mammalian fertilization proteins. Mol Biol Evol 2003 Jan;20(1):18-20.

(67) Turner LM, Hoekstra HE. Causes and consequences of the evolution of reproductive proteins. Int J Dev Biol 2008;52(5-6):769-780.

(68) Calkins JD, El-Hinn D, Swanson WJ. Adaptive evolution in an avian reproductive protein: ZP3. J Mol Evol 2007 Nov;65(5):555-563.

(69) Clark NL, Gasper J, Sekino M, Springer SA, Aquadro CF, Swanson WJ. Coevolution of interacting fertilization proteins.. PLoS Genet 2009 Jul;5(7):e1000570.

(70) Snook RR, Hosken DJ, Karr TL. The biology and evolution of polyspermy: insights from cellular and functional studies of sperm and centrosomal behavior in the fertilized egg. Reproduction 2011 Dec;142(6):779-792.

(71) Glahn D, Nuccitelli R. Voltage-clamp study of the activation currents and fast block to polyspermy in the egg of Xenopus laevis. Dev Growth Differ 2003 Apr;45(2):187-197.

(72) Shapiro BM, Somers CE, Weidman PJ. Extracellular remodeling during fertilization. In: Schatten H, Schatten G, editors. The Cell Biology of Fertilization San Diego: Academic Press; 1989. p. 251-276.

(73) Wassarman PM. Fertilization: Welcome to the fold. Nature 2008 Dec 4;456(7222):586-587.

(74) Sun QY. Cellular and molecular mechanisms leading to cortical reaction and polyspermy block in mammalian eggs. Microsc Res Tech 2003 Jul 1;61(4):342-348.

(75) Talbot P, Dandekar P. Perivitelline space: does it play a role in blocking polyspermy in mammals? Microsc Res Tech 2003 Jul 1;61(4):349-357.

(76) Wong A. Testing the Effects of Mating System Variation on Rates of Molecular Evolution in Primates. Evolution 2010 May 21.

(77) Lindenfors P. Sexually antagonistic selection on primate size. J Evol Biol 2002;15:595-607.

(78) Dixson AF. Sexual selection and evolution of the seminal vesicles in primates. Folia Primatol (Basel) 1998;69(5):300-306.

(79) Harcourt AH, Harvey PH, Larson SG, Short RV. Testis weight, body weight and breeding system in primates. Nature 1981 Sep 3;293(5827):55-57.

(80) Baer B, Heazlewood JL, Taylor NL, Eubel H, Millar AH. The seminal fluid proteome of the honeybee Apis mellifera. Proteomics 2009 Apr;9(8):2085-2097.

(81) Dorus S, Evans PD, Wyckoff GJ, Choi SS, Lahn BT. Rate of molecular evolution of the seminal protein gene SEMG2 correlates with levels of female promiscuity. Nat Genet 2004 Dec;36(12):1326-1329.

(82) Jensen-Seaman MI, Li WH. Evolution of the hominoid semenogelin genes, the major proteins of ejaculated semen. J Mol Evol 2003 Sep;57(3):261-270.

(83) Carnahan SJ, Jensen-Seaman MI. Hominoid seminal protein evolution and ancestral mating behavior. Am J Primatol 2008 Oct;70(10):939-948.

(84) Pilch B, Mann M. Large-scale and high-confidence proteomic analysis of human seminal plasma. Genome Biol 2006;7(5):R40.

(85) Drake RR, Elschenbroich S, Lopez-Perez O, Kim Y, Ignatchenko V, Ignatchenko A, et al. In-depth proteomic analyses of direct expressed prostatic secretions.. J Proteome Res 2010 May 7;9(5):2109-16.

(86) Ambekar AS, Nirujogi RS, Srikanth SM, Chavan S, Kelkar DS, Hinduja I, et al. Proteomic analysis of human follicular fluid: A new perspective towards understanding folliculogenesis. J Proteomics 2013 Jul 11;87:68-77.

(87) Fung KY, Glode LM, Green S, Duncan MW. A comprehensive characterization of the peptide and protein constituents of human seminal fluid.. Prostate 2004 Oct 1;61(2):171-81.

(88) Martinez-Heredia J, de Mateo S, Vidal-Taboada JM, Ballesca JL, Oliva R. Identification of proteomic differences in asthenozoospermic sperm samples.. Hum Reprod 2008 Apr;23(4):783-91.

(89) Ramm SA, McDonald L, Hurst JL, Beynon RJ, Stockley P. Comparative proteomics reveals evidence for evolutionary diversification of rodent seminal fluid and its functional significance in sperm competition. Mol Biol Evol 2009 Jan;26(1):189-198.

(90) Sarason RL, VandeVoort CA, Mader DR, Overstreet JW. The use of nonmetal electrodes in electroejaculation of restrained but unanesthetized macaques. J Med Primatol 1991 May;20(3):122-125.

(91) Fussell EN, Franklin LE, Frantz RC. Collection of chimpanzee semen with an artificial vagina. Lab Anim Sci 1973 Apr;23(2):252-255.

(92) Eng JK, McCormack AL, Yates JR. An Approach to Correlate Tandem Mass Spectral Data of Peptides with Amino Acid Sequences in a Protein Database. J Am Soc Mass Spectrom 1994;5:976-989.

(93) Kall L, Canterbury JD, Weston J, Noble WS, MacCoss MJ. Semi-supervised learning for peptide identification from shotgun proteomics datasets.. Nat Methods 2007 Nov;4(11):923-5.

(94) Tabb DL, McDonald WH, Yates JR,3rd. DTASelect and Contrast: tools for assembling and comparing protein identifications from shotgun proteomics. J Proteome Res 2002 Jan-Feb;1(1):21-26.

(95) Hsieh EJ, Shulman NJ, Dai DF, Vincow ES, Karunadharma PP, Pallanck L, et al. Topograph, a software platform for precursor enrichment corrected global protein turnover measurements. Mol Cell Proteomics 2012 Nov;11(11):1468-1474.

(96) George RD, McVicker G, Diederich R, Ng SB, MacKenzie AP, Swanson WJ, et al. Trans genomic capture and sequencing of primate exomes reveals new targets of positive selection. Genome Res 2011 Oct;21(10):1686-1694.

(97) Denny P, Hagen FK, Hardt M, Liao L, Yan W, Arellanno M, et al. The proteomes of human parotid and submandibular/sublingual gland salivas collected as the ductal secretions. J Proteome Res 2008 May;7(5):1994-2006.

(98) Omenn GS, States DJ, Adamski M, Blackwell TW, Menon R, Hermjakob H, et al. Overview of the HUPO Plasma Proteome Project: results from the pilot phase with 35 collaborating laboratories and multiple analytical groups, generating a core dataset of 3020 proteins and a publicly-available database. Proteomics 2005 Aug;5(13):3226-3245.

(99) Zhang J, Nielsen R, Yang Z. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. Mol Biol Evol 2005 Dec;22(12):2472-2479.

(100) Yang Z, Swanson WJ. Codon-substitution models to detect adaptive evolution that account for heterogeneous selective pressures among site classes.. Mol Biol Evol 2002 Jan;19(1):49-57.

(101) Lartillot N, Poujol R. A phylogenetic model for investigating correlated evolution of substitution rates and continuous phenotypic characters. Mol Biol Evol 2011 Jan;28(1):729-744.

(102) Al-Shahrour F, Diaz-Uriarte R, Dopazo J. FatiGO: a web tool for finding significant associations of Gene Ontology terms with groups of genes. Bioinformatics 2004 Mar 1;20(4):578-580.

(103) Al-Shahrour F, Minguez P, Vaquerizas JM, Conde L, Dopazo J. BABELOMICS: a suite of web tools for functional annotation and analysis of groups of genes in high-throughput experiments. Nucleic Acids Res 2005 Jul 1;33(Web Server issue):W460-4.

(104) Hassan MI, Kumar V, Singh TP, Yadav S. Purification and characterization of zinc alpha2-glycoprotein-prolactin inducible protein complex from human seminal plasma. J Sep Sci 2008 Jul;31(12):2318-2324.

(105) Rhesus Macaque Genome Sequencing and Analysis Consortium, Gibbs RA, Rogers J, Katze MG, Bumgarner R, Weinstock GM, et al. Evolutionary and biomedical insights from the rhesus macaque genome. Science 2007 Apr 13;316(5822):222-234.

(106) Cohen DJ, Busso D, Da Ros V, Ellerman DA, Maldera JA, Goldweic N, et al. Participation of cysteine-rich secretory proteins (CRISP) in mammalian sperm-egg interaction. Int J Dev Biol 2008;52(5-6):737-742.

(107) Kratzschmar J, Haendler B, Eberspaecher U, Roosterman D, Donner P, Schleuning WD. The human cysteine-rich secretory protein (CRISP) family. Primary structure and tissue distribution of CRISP-1, CRISP-2 and CRISP-3. Eur J Biochem 1996 Mar 15;236(3):827-836.

(108) Soler-Garcia AA, Maitra R, Kumar V, Ise T, Nagata S, Beers R, et al. The PATE gene is expressed in the accessory tissues of the human male genital tract and encodes a secreted sperm-associated protein. Reproduction 2005 Apr;129(4):515-524.

(109) Moretti E, Scapigliati G, Pascarelli NA, Baccetti B, Collodel G. Localization of AKAP4 and tubulin proteins in sperm with reduced motility. Asian J Androl 2007 Sep;9(5):641-649.

(110) Skerget S, Rosenow M, Polpitiya A, Petritis K, Dorus S, Karr TL. The Rhesus Macaque (Macaca mulatta) Sperm Proteome. Mol Cell Proteomics 2013 Jul 1.

(111) Baker MA, Hetherington L, Reeves G, Muller J, Aitken RJ. The rat sperm proteome characterized via IPG strip prefractionation and LC-MS/MS identification. Proteomics 2008 Jun;8(11):2312-2321.

(112) Nascimento JM, Shi LZ, Meyers S, Gagneux P, Loskutoff NM, Botvinick EL, et al. The use of optical tweezers to study sperm competition and motility in primates. J R Soc Interface 2008 Mar 6;5(20):297-302.

(113) Clark NL, Swanson WJ. Pervasive adaptive evolution in primate seminal proteins. PLoS Genet 2005 Sep;1(3):e35.

(114) Kingan SB, Tatar M, Rand DM. Reduced polymorphism in the chimpanzee semen coagulating protein, semenogelin I. J Mol Evol 2003 Aug;57(2):159-169.

(115) Chen MS, Tung KS, Coonrod SA, Takahashi Y, Bigler D, Chang A, et al. Role of the integrin-associated protein CD9 in binding between sperm ADAM 2 and the egg integrin alpha6beta1: implications for murine fertilization. Proc Natl Acad Sci U S A 1999 Oct 12;96(21):11830-11835.

(116) Rochwerger L, Cohen DJ, Cuasnicu PS. Mammalian sperm-egg fusion: the rat egg has complementary sites for a sperm protein that mediates gamete fusion. Dev Biol 1992 Sep;153(1):83-90.

(117) Busso D, Cohen DJ, Maldera JA, Dematteis A, Cuasnicu PS. A novel function for CRISP1 in rodent fertilization: involvement in sperm-zona pellucida interaction. Biol Reprod 2007 Nov;77(5):848-854.

(118) Okabe M, Adachi T, Takada K, Oda H, Yagasaki M, Kohama Y, et al. Capacitation-related changes in antigen distribution on mouse sperm heads and its relation to fertilization rate in vitro. J Reprod Immunol 1987 Jun;11(2):91-100.

(119) Nishimura H, Cho C, Branciforte DR, Myles DG, Primakoff P. Analysis of loss of adhesive function in sperm lacking cyritestin or fertilin beta. Dev Biol 2001 May 1;233(1):204-213.

(120) Da Ros VG, Maldera JA, Willis WD, Cohen DJ, Goulding EH, Gelman DM, et al. Impaired sperm fertilizing ability in mice lacking Cysteine-RIch Secretory Protein 1 (CRISP1). Dev Biol 2008 Aug 1;320(1):12-18.

(121) Miyado K, Yamada G, Yamada S, Hasuwa H, Nakamura Y, Ryu F, et al. Requirement of CD9 on the egg plasma membrane for fertilization. Science 2000 Jan 14;287(5451):321-324.

(122) Kaji K, Oda S, Shikano T, Ohnuki T, Uematsu Y, Sakagami J, et al. The gamete fusion process is defective in eggs of Cd9-deficient mice. Nat Genet 2000 Mar;24(3):279-282.

(123) Inoue N, Ikawa M, Isotani A, Okabe M. The immunoglobulin superfamily protein Izumo is required for sperm to fuse with eggs. Nature 2005 Mar 10;434(7030):234-238.

(124) Wright MD, Tomlinson MG. The ins and outs of the transmembrane 4 superfamily. Immunol Today 1994 Dec;15(12):588-594.

(125) Rubinstein E, Benoit P, Billard M, Plaisance S, Prenant M, Uzan G, et al. Organization of the human CD9 gene. Genomics 1993 Apr;16(1):132-138.

(126) Ellerman DA, Cohen DJ, Da Ros VG, Morgenfeld MM, Busso D, Cuasnicu PS. Sperm protein "DE" mediates gamete fusion through an evolutionarily conserved site of the CRISP family. Dev Biol 2006 Sep 1;297(1):228-237.

(127) Busso D, Cohen DJ, Hayashi M, Kasahara M, Cuasnicu PS. Human testicular protein TPX1/CRISP-2: localization in spermatozoa, fate after capacitation and relevance for gamete interaction. Mol Hum Reprod 2005 Apr;11(4):299-305.

(128) Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol 2007 Aug;24(8):1586-1591.

(129) Pazos F, Helmer-Citterich M, Ausiello G, Valencia A. Correlated mutations contain information about protein-protein interaction. J Mol Biol 1997 Aug 29;271(4):511-523.

(130) Goh CS, Cohen FE. Co-evolutionary analysis reveals insights into protein-protein interactions. J Mol Biol 2002 Nov 15;324(1):177-192.

(131) Pazos F, Valencia A. Similarity of phylogenetic trees as indicator of protein-protein interaction. Protein Eng 2001 Sep;14(9):609-614.

(132) Ramani AK, Marcotte EM. Exploiting the co-evolution of interacting proteins to discover interaction specificity. J Mol Biol 2003 Mar 14;327(1):273-284.

(133) Hamm D, Mautz BS, Wolfner MF, Aquadro CF, Swanson WJ. Evidence of amino acid diversity-enhancing selection within humans and among primates at the candidate sperm-receptor gene PKDREJ. Am J Hum Genet 2007 Jul;81(1):44-52.

(134) Gasper J, Swanson WJ. Molecular population genetics of the gene encoding the human fertilization protein zonadhesin reveals rapid adaptive evolution. Am J Hum Genet 2006 Nov;79(5):820-830.

(135) Yang Z, Swanson WJ, Vacquier VD. Maximum-likelihood analysis of molecular adaptation in abalone sperm lysin reveals variable selective pressures among lineages and sites.. Mol Biol Evol 2000 Oct;17(10):1446-55.

(136) Swanson WJ. Adaptive evolution of genes and gene families.. Curr Opin Genet Dev 2003 Dec;13(6):617-22.

(137) Grayson P, Civetta A. Positive Selection and the Evolution of izumo Genes in Mammals. Int J Evol Biol 2012;2012:958164.

(138) Collins FS, Green ED, Guttmacher AE, Guyer MS, US National Human Genome Research Institute. A vision for the future of genomics research. Nature 2003 Apr 24;422(6934):835-847.

(139) Liska AJ, Shevchenko A. Expanding the organismal scope of proteomics: cross-species protein identification by mass spectrometry and its implications. Proteomics 2003 Jan;3(1):19-28.

(140) Karn RC, Clark NL, Nguyen ED, Swanson WJ. Adaptive evolution in rodent seminal vesicle secretion proteins. Mol Biol Evol 2008 Nov;25(11):2301-2310.

(141) Schmid KJ, Tautz D. A screen for fast evolving genes from Drosophila. Proc Natl Acad Sci U S A 1997 Sep 2;94(18):9746-9750.

(142) Helinski ME, Hood RC, Gludovacz D, Mayr L, Knols BG. A 15N stable isotope semen label to detect mating in the malaria mosquito Anopheles arabiensis Patton. Parasit Vectors 2008 Jul 1;1(1):19.

(143) Helinski ME, Hood-Nowotny R, Mayr L, Knols BG. Stable isotope-mass spectrometric determination of semen transfer in malaria mosquitoes. J Exp Biol 2007 Apr;210(Pt 7):1266-1274.

(144) Wu CC, MacCoss MJ. Shotgun proteomics: tools for the analysis of complex biological systems. Curr Opin Mol Ther 2002 Jun;4(3):242-250.

(145) Shevchenko A, Jensen ON, Podtelejnikov AV, Sagliocco F, Wilm M, Vorm O, et al. Linking genome and proteome by mass spectrometry: large-scale identification of yeast proteins from two dimensional gels. Proc Natl Acad Sci U S A 1996 Dec 10;93(25):14440-14445.

(146) MacCoss MJ, Matthews DE. Quantitative MS for proteomics: teaching a new dog old tricks. Anal Chem 2005 Aug 1;77(15):294A-302A.

Vita

Katrina Garnet Claw was born in Fort Defiance, Arizona and grew up in Many Farms, Arizona on the Navajo Nation. She graduated from Chinle High School in 2001 and then moved to Tempe where she received her Bachelor of Science in Biology and her Bachelor of Arts in Anthropology from Arizona State University in 2006. Following her undergraduate studies, she became a post-baccalaureate research fellow at Arizona State University. In 2008, she began her graduate work at the University of Washington. In 2013, she received a Doctor of Philosophy in Genome Sciences. In her free time, Katrina enjoys reading, running, hiking, and traveling.