Molecular Diversification and Species-Specific Interactions of Gamete Recognition Proteins

Jolie Anna Carlisle

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2021

Reading Committee:

Willie J. Swanson, Chair

Harmit Malik

Douglas Fowler

Program Authorized to Offer Degree:

Genome Sciences

University of Washington

Abstract

Molecular Diversification and Species-Specific Interactions of Gamete Recognition Proteins

Jolie Anna Carlisle

Chair of the Supervisory Committee:
Professor Willie J. Swanson
Genome Sciences

Reproductive proteins mediating fertilization commonly exhibit rapid sequence diversification driven by positive selection. This pattern has been observed among nearly all taxonomic groups, including mammals, invertebrates, and plants, and is remarkable given the essential nature of the molecular interactions mediating fertilization. In chapters two and three of my thesis, I discuss how variation in reproductive gene content between species can be generated through subfunctionalization or gene loss. In chapter four, I discuss how sequence diversification of gamete recognition proteins can create boundaries to hybridization.

The marine gastropod mollusk abalone (genus *Haliotis*) is a classic model for fertilization. Its two acrosomal proteins (lysin and sp18) are ancient gene duplicates with unique gamete recognition

functions. Through detailed genomic and bioinformatic analyses we show how duplication events followed by sequence diversification has played an ongoing role in the evolution of abalone acrosomal proteins. The common ancestor of abalone had four members of its acrosomal protein family in a tandem gene array that repeatedly experienced positive selection. Further, a more recent species-specific duplication of both lysin and sp18 in the European abalone *H. tuberculata* is described. We hypothesize that, in a manner analogous to host/pathogen evolution, sperm proteins are selected for increased diversity through extensive sequence divergence and recurrent duplication driven by conflict mechanisms.

The abalone lysin-VERL interaction is a classic model for understanding the molecular mechanisms mediating egg coat dissolution. During abalone fertilization, sperm lysin dissolves the vitelline envelope of the egg in a species-specific manner by binding VERL ZP-N domains. Using surface plasmon resonance, a quantitative biophysical method, we have measured the binding affinities between heterospecific and conspecific lysin-VERL domain pairs. In our study, we discovered that species-specific binding between lysin and VERL ZP-N Repeat 1 (VR1) was consistent with the experimentally measured species-specific VE dissolution function of lysin. ZP2, a mammalian egg coat protein, shows similarity in evolutionary patterns, protein structure, and function to abalone VERL despite extensive divergence between the proteins. Experimentally guided molecular docking indicates that both abalone and mammalian ZP-N domains bind the same region of lysin, consistent with a shared mechanism between mammalian and abalone sperm-egg interactions.

My thesis explores how the diversification of fertilization genes in sequence and gene content across even closely related species can impact mechanisms of fertilization. Further, my thesis

emphasizes the value of using diverse model systems for investigating mechanisms of fertilization.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I am grateful for my training in the Department of Genome Sciences at the University of Washington. I appreciate everyone who has contributed to my training and supported me throughout my graduate studies. During my graduate training I was supported by a National Science Foundation Graduate Research Fellowship. In addition, I am thankful to members of the genus *Haliotis*, without whom I could not have had this inspiring research experience.

Ironically, despite abalone being a classic model for understanding human fertility, humans have hunted and damaged the habitats of the pinto abalone (*Haliotis kamtschatkana*) to the point of being functionally extinct in Washington state. The population densities of abalone in Washington are too low to support successful broadcast spawning. Around the world different species of abalone are becoming endangered through habitat destruction, pollution, disease, and overfishing. On the west coast of North America there are several abalone recovery projects (the pinto abalone recovery project: https://restorationfund.org/programs/pintoabalone/). I request that anyone reading this thesis consider donating to one of the abalone recovery projects, and to only consume abalone products (including jewelry) from farm-raised abalone.

I feel very lucky to have had Dr. Willie Swanson as my advisor. Willie is a thoughtful, empathetic, and kind mentor that prioritizes his trainees. I appreciate him giving me the space to grow as a scientist and believing in me. I am grateful to my lab mates in the Swanson Lab. In particular, I

appreciate Damien's mentorship and his willingness to have long, educational scientific discussions with me. I am grateful to Jan's technical assistance during so many experiments and his thoughtful advice. I enjoyed joining the lab with Alberto, one of the funniest people I know, and how we encouraged each other throughout our graduate student journeys.

During my PhD I had the opportunity to be the primary research mentor of two talented undergraduate trainees. Megan Glenski and Derya Gurbuz participated in the research discussed in Chapters 2 and 3 respectively. Thank you both for letting me be your mentor.

Performing SPR experiments was a profound technical challenge during my thesis. I am deeply appreciative of the support of the Molecular Analysis Facility at UW and particularly to Scott Soelberg, John Sumida, and Ellen Lavoie.

During the beginning of my graduate school journey, I switched thesis labs. I was lucky to be supported by an incredible committee that believed in me and made me feel supported during this transition. All of my committee members have given me kind advice and guidance. Thank you so much Harmit Malik, Daniel Promislow, Doug Fowler, and Jesse Bloom. Also, I am grateful to the Genome Sciences staff (particularly Brian Giebel) who helped me through the logistical challenges of graduate school.

I had the opportunity to work in a few research environments during my undergraduate training. Monica Albe at UC Berkeley helped me get my first genetics research experience at the Evolutionary Genetics Laboratory in the Museum of Vertebrate Zoology. At the MVZ I was

fortunate to be mentored by Dr. Carla Cicero and Elyse Freitas. I am thankful to Dr. Jose Aponte for mentoring me for a summer at NASA Goddard National Laboratory in Maryland. Further, I am grateful to Dr. Michael Shapira for allowing me to contribute to genetics research in his lab at UC Berkeley. In the Shapira Lab I was mentored by two incredible graduate students, Dr. Dena Block and Dr. Maureen Berg. Thank you both for your support and patience. I am also deeply grateful to my High School Teachers who helped guide me to the beginning of my science career.

Helping found and organize the Genomics Salon has been one of the most influential experiences of my graduate career. I am appreciative of all my fellow Salon Organizers, particularly Dr. Katherine Xue who proposed the idea and was a driving force. Further, I appreciated being a participant in the STEP-UP Teacher Training Program at UW, supervised by Ben Wiggins.

I am incredibly fortunate to have made so many wonderful friendships with kind, thoughtful, and inspiring people during graduate school. In particular, I want to thank Dr. Katherine Xue, Dr. Seungsoo Kim, Dr. Bryce Taylor, Claire Robinson, Alberto Rivera, and Dr. Alex Hu for their friendship.

I am grateful to the support of my parents, siblings, and grandma

Finally, I would like to thank Evan Cox. Thank you for your support and care through a formative period of my career and a global pandemic.

DEDICATION

To all the women in my family that helped pave my way. In particular, my great, great grandmother Annie Elizabeth Nicklin (née Taylor) 1888-1973.

Chapter 1. MOLECULAR MECHANISMS AND EVOLUTION OF FERTILIZATION PROTEINS

A version of this chapter has previously been published as:

Carlisle JA, Swanson WJ. 2020. Molecular mechanisms and evolution of fertilization proteins. J Exp Zool B Mol Dev Evol.

1.1 INTRODUCTION

Fertilization is essential for all sexually reproducing organisms, but the molecular mechanisms mediating this process remain largely unknown. Interacting sperm and egg molecules mediating fertilization have only been identified in a handful of animals. In none of these animal models have molecular partners mediating each step of the sperm-egg fertilization cascade been identified (Swanson & Vacquier, 2002). For the majority of these proteins, no obvious homologs can be identified between distant taxa. Perhaps this is not surprising; sequences of reproductive proteins tend to be highly divergent in sequence between even closely related species and often show signatures of positive Darwinian selection (Swanson & Vacquier, 2002; Swanson, Clark et al. 2001; Clark, et al. 2006; Clark, et al. 2005; Grayson, 2015; Swanson, Yang, et al. 2001). Rapid diversification paired with patterns of gene duplication and loss could thwart efforts to integrate molecular fertilization models between species. In addition, functional redundancy and lineage-specific co-option of proteins (e.g. signaling pathway proteins) into fertilization pathways could be prevalent and further complicate comparisons of fertilization mechanisms across taxa. Although identifying a singular molecular pathway to describe animal fertilization is highly unlikely, comparisons of diverse taxa across the tree of life provide trends for the mechanisms mediating gamete fusion and enable insights into the biochemistry of fertilization. For example, homology

between some fertilization proteins from diverse taxa and recurrent use of certain protein domains may indicate conserved molecular mechanisms for some steps of fertilization.

Here we provide a review of proteins that mediate gamete recognition in abalone, sea urchins, and mammals (Figure 1). Despite their evolutionary distance from mammals, abalone and sea urchins are invaluable models for studying mechanisms of gamete recognition (Vacquier, 2019; Vacquier, 2012; Panhuis, 2006; Evans, 2013). We will discuss how discoveries in these systems have led to new insights into human fertilization and into the evolution of reproductive genes across taxa (Swanson & Vacquier, 2002; Moy, 1996; Hughes, 1999; Swanson, 2011). Together with mammals, these marine invertebrate models contain the majority of currently known sperm-egg gamete recognition protein (GRP) pairs (Table 1).

Many of the proteins discussed in this review have been demonstrated to mediate species-specific gamete recognition (Vacquier, 1993; Swanson & Vacquier, 1997; Glabe & Vacquier, 1977; Glabe & Lennarz, 1979; Kamei & Glabe, 2003; Alves, 1997; Hermens, 2012; Bianchi & Wright, 2015). Species-specific function is believed to be a result of the rapid diversification of these proteins leading to species-specific GRP binding interactions. Species-specific GRP function could result in barriers to hybridization and thereby contribute to the formation of species boundaries. However, what drives the rapid divergence of GRPs and the significance of species-specific GRP function for the process of species formation are matters of open inquiry. In other words, species-specific function could be an indirect outcome of the rapid diversification of GRPs rather than a driver of their evolution. Our goal in describing species-specific GRP interactions in this review is to highlight important functional work done on these proteins, not to explain why these interactions form. In the last sections of the review we explore the various hypotheses as to what

drives the evolution of GRPs (sexual conflict, sexual selection, pathogen avoidance, and reinforcement) and inquire into how changes in GRP sequence can lead to the formation of species boundaries.

1.2 Mechanisms of Gamete Recognition

The majority of known GRP pairs mediating sperm-egg recognition in animals are represented by those found within abalone, sea urchins, and mammals (Table 1). Here, we provide a summary of notable GRPs from these species and describe known species-specific gamete recognition functions. The major steps of fertilization are broadly consistent across taxa (Figure 1). Chemoattractants released by the egg incite directed movement by the sperm towards the egg. Once the sperm has reached the egg, the sperm-egg fertilization cascade begins with the induction of the sperm acrosome reaction by the egg. The sperm acrosome is an organelle on the anterior head of the sperm containing digestive enzymes and important GRPs. The acrosome reaction causes the release of molecules from the sperm acrosome essential for sperm entry (Buffone, 2012). After passing the egg jelly (sea urchins and abalone) or corona radiata (mammals), sperm next must bind and dissolve a hole in the egg coat (Bhakta, 2019; Trimmer & Vacquier, 1986). The egg coat is a protective glycoprotein layer outside the egg's plasma membrane. The egg coat is referred to as the vitelline envelope in non-mammals and the zona pellucida in mammals. After the sperm passes this barrier, the acrosome-reacted sperm, now prepared for fusion, enters the perivitelline space and binds to the plasma membrane (oolemma) of the egg. After successful binding, fusion of the sperm and egg plasma membranes results in sperm nuclei deposition in the egg and zygote formation.

Molecular mechanisms of fertilization discovered in marine invertebrates have proved to be fundamental to scientists' understanding of mechanisms of species-specific gamete recognition and have led to the identification of genes and functional domains relevant to fertilization in mammals (Hughes, 1999; Turner & Hoekstra, 2008). Sea urchin bindin, mediating oolemma fusion, was the first gamete recognition protein discovered in animals (Turner & Hoekstra, 2008). Abalone's lysin and VERL proteins are the first known sperm-egg protein-protein interaction (PPI) pair mediating gamete recognition (Swanson & Vacquier, 1997). Early fertilization research in these marine invertebrate models was aided by the availability of large amounts of gametic material from these prolific external fertilizers. In both broadcast spawning animals like abalone and sea urchins and internal fertilizers like mammals and Drosophila, many GRPs have been shown to have species-specific binding functions (Swanson, Yang, et al., 2001; Findlay, 2014). It is unclear how species-specific GRP function contributes to the formation of species boundaries, however, understanding the molecular mechanisms mediating species-specific GRP function in diverse animal models could lead to new insights into the biochemistry of gamete recognition.

1.2.1 *Sperm Chemotaxis*

During the process of fertilization, sperm must find and reach the egg. Sperm attraction to the egg based on egg-derived chemical cues may be evolutionarily significant for maintaining species boundaries through preferential attraction of conspecific sperm by the egg (Riffell, 2004; Eisenbach & Giojalas, 2006; Kaupp, 2008). The first animal example of a sperm chemoattractant was sea urchin resact and sea urchins remain a preeminent model for studying sperm chemotaxis (Ward, 1985). Sperm chemoattraction models have also been investigated in the marine invertebrates abalone, starfish, and ascidians (Welborn & Manahan, 1995; Yoshida, 2002; Riffell, 2002; Minakata, 2003). However, studying this process in humans has been complicated.

Mammalian sperm do not show as prominent of a chemoattractant response to eggs compared to marine invertebrates.

The sea urchin egg is surrounded by an extracellular matrix that releases Sperm Activating Polypeptides (SAPs) that attract sperm to the egg through directed movement of the sperm up a SAP concentration gradient (Suzuki, 1995). Many of these SAPs specifically stimulate respiration in sperm of the same order or suborder, but not other orders (Hardy, 1994). Many different SAPs have been characterized in various sea urchin species, but two notable examples are resact and speract (Hasbrough & Garbers, 1981). Both resact and speract modulate sperm movement by altering sperm intracellular [Ca2+] and cellular signalling pathways, including cyclic AMP and cyclic GMP, and activating a sperm guanyly cyclase (Suzuki, 1995; Hardy, 1994; Hussain, 2016; Dangott, 1989; Bentley, Shimomura, 1986; Bentley, Tubb, 1986). Resact is a small 14 amino acid polypeptide that attracts the migration of sperm in *Arbacia punctualata* (Ward, 1985; Ramirez-Gomez, 2020). Pre-treatment of sperm with resact inhibits their ability to have a chemotatic response, indicating that sperm chemotaxis can be a significant step in fertilization and important for reproductive fitness (Ward, 1985). Sea urchin speract is a decapeptide released by the egg known to increase sperm motility and cause chemotaxis to the egg in the species *Stronglycentrotus purpuratus* and *Lytechinus pictus* (Hussain, 2016; Ramirez-Gomez, 2020).

Abalone sperm also exhibit species-specific sperm chemotaxis, however the molecules mediating abalone chemotaxis are very different. The sperm of both *Haliotis rufescens* and *H. fulgens* show preferential chemoattraction to eggs of their own species (Riffell, 2004; Riffell, 2002). In *H. rufescens*, effective chemoattraction is largely mediated by a dose-dependent, stereoisomer specific attraction to L-Tryptophan released by the egg (Riffell, 2004; Riffell, 2002).

L-tryptophan induces both activation and increased mobilization in abalone sperm eggs (Riffell, 2002).

In mammals, identification of sperm chemoattractants has been challenging. Although ovarian follicular fluid and medium conditioned by the cumulus cells surrounding the egg have both been shown to drive sperm chemotaxis, the phenotype is weak compared to that seen in marine invertebrates and thereby more challenging to characterize (~10% chemotatic response observed in mice) (Eisenbach, 1999; Villanueva-Diaz, 1995). Progesterone, the main steroidal component emitted by the cumulus cells that surround the egg, is the only known egg molecule shown to initiate sperm chemotaxis in mammals. Progesterone released by the egg creates a stable concentration gradient along the cumulus oophorus that sperm can utilize to orient towards the egg (Teves, 2009). Progesterone has been shown to modulate the activity of sperm CatSper (Cation channels of Sperm), which are voltage-gated Ca2+ channels (Strunker, 2011).

Although olfactory receptors (ORs) are known to be present in human sperm, their egg binding partners and their potential role in sperm chemotaxis remain a mystery. It has been posited that sperm ORs may alter [Ca2+], potentially resulting in the acrosome reaction, capacitation, or hyperactive motility (Sinding, 2013; Ottoviano, 2013). Support has been found for the role of one human sperm OR, hOR17-4, in mediating chemotaxis through induction of the Ca2+ pathway by bourgenal, an aromatic aldehyde (Spehr, 2003). However, bourgeonal is not a naturally occurring molecule in humans, indicating that a biochemically similar molecule may initiate the process (Spehr, 2003). It is unknown whether this molecule would be released by the egg or at another location in the mammalian female reproductive tract.

1.2.2 *Induction of the Acrosome Reaction*

Induction of the sperm acrosome reaction by the egg is essential for releasing digestive enzymes and GRPs essential for sperm entry. In sea urchins and abalone, eggs are covered in a transparent jelly coat that protects the egg (Keller & Vacquier, 1994; Kresge, 2001; Mozingo, 1995). In mammals, eggs are surrounded by the corona radiata, a layer of cells of the cumulus oophorus adjacent to the zona pellucida (Bhakta, 2019). Induction of the acrosome reaction and the release of acrosomal proteins are essential for sperm passing these structures (Bhakta, 2019). However, the proteins mediating the acrosome reaction as well as the timing of this event remain controversial in some models. In mammals there is evidence for the induction of the acrosome reaction occurring upon sperm reaching the egg coat (Litscher & Wassarman, 1996; Buffone, 2009). However, there is also evidence that sperm undergo the acrosome reaction before reaching the cumulus mass and that acrosome reacted mice sperm are still capable of bypassing mouse egg coats (Inoue, 2011; Hino, 2016). This discrepancy may be explained by the functional redundancy of proteins mediating this step of fertilization, with proteins capable of initiating the acrosome reaction found within the egg and elsewhere in the female reproductive tract.

In sea urchins, when sperm comes in contact with the egg, the acrosome reaction is induced (Keller & Vacquier, 1994). The sea urchin's sperm receptor for egg jelly (suREJ-1) initiates the acrosome reaction after coming in contact with a particular egg jelly fucose sulfate polymer (FSP) (Moy, 1996; Vacquier & Moy, 1997). FSP in sea urchins has species-specific structures, and suREJ-1 contains sites under positive selection (Viela-Silva, 2002; Mah, 2005). The diversification of FSP and its receptor could be indicative of species-specific binding between these molecules acting as a barrier to hybridization (Viela-Silva, 2002). Homologs of *suREJ-1* are found in mammals

(Hughes, 1999). Mammalian *PKDREJ* has testes-specific expression and is under positive selection in humans and rodents (Hughes, 1999; Hamm, 207; Sutton, 2006; Vicens, 2015). Two mouse strains with *PKDREJ* knockouts show that male mice had no detectable decrease in litter size, however, more subtle fertilization phenotypes associated with male fertility may still be present (Miyata, 2016). Mice homozygous for a targeted mutation in *PKDREJ* show delays in induction of the acrosome reaction by zona pellucida recognition and a reduction in male fertility compared to wild type animals in sequential mating trials (Sutton, 2006; Sutton, 2008). Together these results indicate that while *PKDREJ* is not essential for successful fertilization, it still plays a significant role in fertilization. The location of PKDREJ receptor in mammalian eggs is still unknown, however delays in the induction of the acrosome reaction in mice indicate that there could be a shared gamete recognition mechanism between su-REJ in sea urchins and PKDREJ in mammals.

In mammals the location of the acrosome reaction and the proteins involved remain controversial. In addition to PKDREJ, progesterone from the corona radiata and ZP3 from the egg coat are also thought by some to mediate induction of the acrosome reaction (Buffone, 2009; Sabeur, 1996). In mice, solubilized zona and isolated ZP3 from the zona both initiate the acrosome reaction in vitro (Litscher & Wassarman, 1996). In addition, in vitro studies show that ZP3 functions as a sperm receptor in mice. In some studies, ZP3 sperm receptor activity is lost when its O-linked oligosaccharides have been oxidized, indicating a sperm-egg protein-carbohydrate interaction as is the case for suREJ-1 and FSP (Florman & Wassarman, 1985; Bleil & Wassarman, 1988). However, other research indicates that mouse embryos with mutant ZP3 lacking the O-linked glycosylation are still able to be fertilized, indicating a sperm-egg protein-protein interaction (Chalabi, 2006; Boja, 2003; Gahlay, 2010). Some research suggests that induction of the acrosome

reaction occurs before sperm reaches the egg coat, thereby negating the importance of ZP3 for this process (Bhakta, 2019). In mice and humans, acrosome reacted sperm are still capable of binding the zona and studies have revealed that the majority of sperm have undergone the acrosome reaction before reaching the cumulus mass (Bhakta, 2019; Hino, 2016; Muro, 2016).

A putative sperm receptor mediating the sperm-egg coat ZP3 interaction was identified in mice as ZP3r (Bleil, 1990). In mice, ZP3r, formally sp56, is shown to bind eggs and specifically bind ZP3 via photoaffinity crosslinking (Bleil, 1999; Buffone, 2008). While initially believed to be a peripheral membrane protein located on the outer surface of the sperm head plasma membrane, ZP3r was later shown to be a sperm acrosomal protein that becomes transiently exposed on the sperm surface post-capacitation (Muro, 2012). If ZP3r is a binding partner of ZP3, this result indicates that this protein-protein interaction mediates a fertilization step following the acrosome reaction. Hypothetically the ZP3-ZP3r interaction could instead mediate sperm binding to the egg coat surface post-sperm maturation. However, knockouts of *ZP3r* in mice do not result in an appreciable decrease in mouse fertility (Muro, 2016). This negative result indicates that either an interaction between ZP3 and ZP3r does not mediate fertilization or that functional redundancy of this fertilization step, potentially with PKDREJ, makes detecting a knockout phenotype challenging. Further functional investigation of both ZP3 and ZP3r may be necessary to elucidate their roles in fertilization.

ZP3r is a member of the complement 4-binding family. *ZP3r*'s putative human homolog, *C4BPA*, exists within a cluster of paralogs on chromosome 1. Genetic linkage disequilibrium between *ZP3* and Tag SNPs spanning the region of the genome containing *C4BPA* is present in humans, suggestive of coevolution between interacting binding partners (Rohlfs, 2010).

However, human *C4BPA* is most highly expressed in the liver, a characteristic inconsistent with a role in fertilization (Carithers, 2015). No studies have directly investigated whether C4BPA binds ZP3 in humans.

1.2.3 *Egg Coat Dissolution*

The next step in the fertilization cascade is the dissolution of the egg coat by the released sperm acrosomal proteins (Killingbeck & Swanson, 2018). The egg coat is an elevated glycoprotein layer outside of the plasma membrane. Egg coats are referred to as the zona pellucida in mammals and as the vitelline envelope (VE) in invertebrates (Mozingo, 1995; Killingbeck & Swanson, 2018). Egg coats have significant structural differences between taxa. In sea urchins, the VE is a thin layer of glycoproteins directly surrounding the plasma membrane of the egg (Glabe & Vacquier, 1977). In abalone the VE is an elevated layer of glycoproteins surrounding the egg. However, electron microscopy has detected a secondary VE-like structure in abalone directly surrounding the egg plasma membrane (Mozingo, 1995; Killingbeck & Swanson, 2018). In humans, the zona pellucida exists as an elevated glycoprotein layer that is morphologically and biochemically similar to abalone's vitelline envelope (Hasbrough & Garbers, 1981). Despite these morphological differences, glycoproteins found in egg coats across many distant taxa share similar protein domains (Swanson, 2011; Vacquier & Moy, 1977). These similarities may be suggestive of maintained mechanisms of egg coat binding and dissolution between species.

Egg coat glycoproteins in many diverse taxa contain zona pellucida (ZP) modules (Monne & Jovine, 2011). Some of these ZP-containing glycoproteins have been shown to mediate sperm-egg recognition (Killingbeck & Swanson, 2018). The ZP module consists of two immunoglobulin-like

domains, ZP-N and ZP-C, and is vital for forming a strong egg coat by contributing to protein polymerization (Monne & Jovine, 2011; Wilburn & Swanson, 2017). ZP proteins often have multiple sites of N- and O-linked glycosylation (Killingbeck & Swanson, 2018). The role of these glycans in binding sperm is debated in the literature. For example, there is evidence for and against the role of ZP3 O-linked glycans in inducing the acrosome reaction (Monne & Jovine, 2011). In some egg coat proteins, there are additional tandem ZP-N domains at the N-terminus of the molecule (Figure 2) (Wilburn & Swanson, 2016; Galindo, 2002). The repeat nature of binding sites in GRPs appears to be a recurring theme across taxa as this is also seen in sea urchin EBR1 (Killingbeck & Swanson, 2018). Species-specific binding of sperm to these N-terminal ZP-N domains has been shown to mediate gamete recognition in both abalone and mammals (Avella, 2014; Raj, 2017).

The abalone GRP pair of sperm lysin and egg coat VERL (Vitelline Envelope Receptor for Lysin) is the only GRP pair known to mediate egg coat dissolution (Swanson & Vacquier, 1997). In red abalone (*Haliotis rufescens*), VERL contains an extreme example of the phenomenon of tandem ZP-N domains in an egg coat protein. VERL has 22 tandem ZP-N domains upstream of its ZP module (Figure 2) (Galindo, 2002; Swanson & Vacquier, 1998). Abalone sperm lysin is a ~110 amino acid alpha helical protein which can dissolve abalone egg coats and binds VERL (Swanson & Vacquier, 1997; Aagaard, 2013). Lysin's egg coat dissolution function and VERL binding ability is species-specific with lysin able to more rapidly dissolve egg coats from its own species as compared to egg coats from another species (Swanson & Vacquier, 1997; Raj, 2017). It should be noted that species-specificity in GRP function is rarely all or none. Hybrid fertilization is still possible between species, but the efficiency is reduced (Lyon & Vacquier, 1999). This reduced

efficiency in heterospecific sperm-egg recognition can contribute along with other ecological and behavioral mechanisms to reduce the frequency of hybridization events. Lysin's functional specificity is believed to be driven by species-specific binding of lysin to the tandem ZP-N domains of its primary receptor VERL (Raj, 2017; Aagaard, 2013).

In mammals, successful binding of sperm to the egg coat is dependent on the zona pellucida protein ZP2. Eggs from wild type mice are unable to bind human sperm, however, transgenic mice expressing human ZP2 gain the ability to bind human sperm and lose their ability to bind mice sperm (Avella, 2014). For ZP2, genetic ablation of the glycan site in transgenic mice does not interfere with sperm-zona pellucida binding, indicating that recognition driven by the N-terminus of ZP2 does not depend on glycosylation (Tokuhiro, 2018).

Despite abalone and mammals being evolutionarily distant from each other, N-terminal ZP-N domains in ZP proteins are important to mediating sperm binding to the egg coat in both groups (Figure 2) (Avella, 2014; Raj, 2017). ZP2 contains 3 N-terminal ZP-N domains, with the first ZP-N domain controlling species-specificity of sperm-zona binding (Avella, 2014). In abalone, an N-terminal ZP-N is also demonstrated to have species-specific sperm protein binding function. In addition, both egg coat proteins consist of a tandem array of ZP-N domains outside the ZP module, 22 in abalone and 3 in mammals (Figure 2) (Wilburn & Swanson, 2016; Swanson & Vacquier, 1988). The origins of ZP-N domains as mediators of sexual reproduction may precede animals. Protein threading has revealed the presence of a ZP-N domain in the agglutination protein SAG1 in yeast (Swanson, 2011; Glabe & Vacquier, 1977). These observations are suggestive of some similarity in molecular mechanisms mediating gamete recognition across the distantly related taxa. However, it should be noted that the existence of egg coat proteins containing ZP modules have

not been discovered in all taxa, including Echinoderms, possibly indicating that ancient homology of these egg coat genes across all taxa does not exist. Instead, repeated co-option of ZP domains from non-fertilization proteins across the tree of life may have led to convergent evolution of egg coats with ZP module-containing proteins. Further investigation is needed to characterize egg coat proteins in diverse taxas to gather support for either hypothesis.

1.2.4 *Plasma membrane fusion*

In order for a zygote to form, the sperm and egg must successfully bind at the plasma membrane surface for membrane fusion to be initiated. An important PPI mediating both of these steps is Echinoderms' sperm bindin and its egg receptor EBR1 (Vacquier & Moy, 1977). Bindin coats the acrosomal process and performs two functions, egg binding and sperm-egg membrane fusion (Moy & Vacquier, 1979). Unlike in abalone, sea urchin VEs are in close contact with the plasma membrane (Glabe & Vacquier, 1977). Anchoring of the sperm to the VE by bindin enables successful plasma membrane fusion (Glabe & Vacquier, 1977). Species-specific variation in the N- and C- terminal regions of bindin is hypothesized to mediate the protein's species-specific function, potentially by specifically anchoring sperm to the vitelline envelope of its own species before fusion (Glabe & Vacquier, 1977; Glabe & Lennarz, 1979; Glabe & Vacquier, 1977; Metz, 1994).

The Egg Bindin Receptor 1 (EBR1) is the major binding partner of bindin and is found on the vitelline envelope (Kamei & Glabe, 2003). Genetic linkage disequilibrium between *EBR1* and *bindin* in *Strongylocentrotus purpuratus* indicates that there is selective pressure to maintain binding between these proteins despite both proteins having signatures of diversifying selection

(Stapper, 2015). The complete protein sequence of EBR1 is identified in the species *S. purpuratus* and *S. franciscanus* (Kamei & Glabe, 2003). In both species these large glycoproteins have similar N-terminal sequences, containing signal sequences, propeptides, reprolysin, and an ADAM domain (Figure 3) (Kamei & Glabe, 2003). Interestingly, ADAM proteins are also found in mammalian sperm where they are believed to have disintegrin and metalloprotease function and are critical for recognition of the egg by the sperm (Kamei & Glabe, 2003; Han, 2009). After the ADAM domain, EBRI in the two sea urchin species contains a series of tandem TSP (Thrombospondin-1) domains, followed by tandem EB repeat domains, which are composed of linked CUB (Complement C1s/C1r, Uegf, BMP1) and TSP domains (Figure 3) (Kamei & Glabe, 2003). Although the N-terminal halves of EBR1 are >80% identical between the species the C-terminus is highly divergent (Kamei & Glabe, 2003). In *S. franciscanus* sea urchins, there are 10 additional species-specific EB repeats, and in *S. purpuratus*, a sympatric species, there are 11 tandem hyalin repeats which are not found in *S. fransciscanus* (Figure 3) (Kamei & Glabe, 2003). This lack of C-terminal evolutionary resemblance between these two sea urchin species in an important egg GRP could result in conspecific bindin binding preference by EBR1 (Kamei & Glabe, 2003). Although composed of entirely different domains, abalone's egg coat protein VERL also is composed of tandem repetitive domains (Figure 2) (Galindo, 2002). This similarity may suggest shared evolutionary pressures or recognition mechanisms (Swanson, Vacquier, 1988).

Obi1, originally known as 350 kDa, is a putative bindin binding partner but its identification is controversial. From sea urchin eggs, researchers purified a 350 kDa glycoprotein that was shown to species-specifically inhibit fertilization and bind acrosome-reacted sperm (Ohliendieck, 1993). A protein fragment derived from 350 kDa was used to develop antibodies for screening an expression library of sea urchin ovary cDNA to identify the 350 kDa protein's DNA sequence

(Foltz, 1993). Besides indirect antibody cross-reactivity, there is no direct evidence (such as peptide sequencing) that the clone identified corresponds to the 350 kDa protein. Later examination of the cloned sequence revealed a number of cloning errors due to chimeric cDNAs (Just & Lennarz, 1997; Mauk, 1997). After correction, the new sequence was revealed to belong to the heat shock protein (hsp) family and is related to the mammalian cytoplasmic hsp70s (Vacquier, 2019; Just & Lennarz, 1997; Mauk, 1997; Vacquier, 2019). However, no further work has been done to confirm whether this sequence, gene name *Obi1*, truly represents the originally isolated 350 kDa protein. There is some indirect evidence supporting *Obi1* as an egg binding partner of bindin. Regions of Obi1 synthesized via bacterial expression show an ability to inhibit fertilization (Stears & Lennarz, 1997). Also, signatures of coevolution between *bindin* and *Obi1* in diverging populations of sea stars is characteristic of arms race dynamics driving the evolution of protein binding partners mediating fertilization (Hart, 2014). A predictive model has shown that the conserved B18 domain of bindin may be a substrate of the protein binding domains of hsp70s like Obi1 (Hart, 2014). However, it should be noted that computational predictions of protein-protein interactions are notoriously difficult, and often spurious. If Obi1 and EBR1 are both binding partners of bindin, it may be possible that they bind to different domains of the bindin protein and both egg protein-bindin interactions are necessary for successful zygote formation in sea urchins.

Sp18 is the putative mediator of gamete fusion in abalone. Sp18 is thought to mediate oolemma fusion in abalone due to its localization to the acrosome in acrosome-reacted sperm and it being a fusagenic molecule (Swanson & Vacquier, 1995). Sp18 and lysin are ancient paralogs which have undergone subfunctionalization, with sp18 mediating plasma membrane fusion and lysin mediating vitelline envelope dissolution (Kresge, 2001; Metz, 1998). These abalone acrosomal

protein paralogs demonstrate how existing fertilization protein structures may duplicate and diverge in function. Recurrent patterns of duplication and subfunctionalization could play an important role in generating new protein-protein interactions mediating gamete recognition in all animals. In addition, the conserved protein structures of sp18 and lysin may suggest a shared biochemical mechanism despite functional diversification (Kresge, 2001).

In mammals, two notable sperm-egg GRP pairs are identified as mediating oolemma binding before gamete fusion. IZUMO1 is a testes specific cell surface protein found on the plasma membrane of acrosome-reacted sperm (Jean, 2019). Knockouts of *Izumo1* result in sperm accumulation in the perivitelline space due to failure in oolema binding (Inoue, 2005). The binding partner of IZUMO1 is oolemmal-bound JUNO, previously known as FOLR4 (Bianchi, Doe, et al., 2014). JUNO is part of the folate receptor family, and while it has maintained structural similarity with FOLR1-3, it has lost its ability to bind folate (Grayson, 2015; Bianchi, Doe, et al., 2014; Bianchi & Wright, 2014). Mouse eggs without *Juno* are infertile due to an inability to bind sperm at the plasma membrane (Jean, 2019; Bianchi, Doe, et al., 2014; Chalbi, 2014). IZUMO1 and JUNO binding is species-specific between mice and humans (Bianchi & Wright, 2015).

Another important GRP pair mediating sperm-egg plasma membrane fusion in mammals is sperm acrosomal SLLP1 (also known as Spaca3) and egg metalloproteinase SAS1B (also known as Ovastacin and ASTL), found on the microvillar oolemma membrane (Hermens, 2012). SAS1B exists in humans as 6 alternate splicing variants and expression is ovary specific, decreasing post-successful fertilization (Hermens, 2012; Sachdev, 2011). SAS1B plays a role in the postfertilization block to polyspermy by cleaving ZP2 (Burkart, 2012). Knockouts of *SAS1B* show a 34% reduction in egg fertility, indicating that, while important for gamete fusion, it is not

required for successful fertilization (Hermens, 2012; Sachdev, 2011). *SLLP1* mutants show no effect on male litter size, however, more subtle male fertility phenotypes may still be present (Miyata, 2016). If functional redundancy of sperm-egg protein pairs mediating plasma membrane fusion or other steps in fertilization exist, it is possible that individual knockouts of any of these pairs may result in partial or incomplete loss of fertility.

Although their binding partners have not been identified, other mammalian sperm and egg proteins have been demonstrated to be important for gamete fusion. SPACA6 is a testes-specific protein whose loss of expression causes mouse infertility and accumulation of sperm in the perivitelline space, similar to the *IZUMO1* knockout phenotype (Lorenzetti, 2014; Barbaux, 2020). Interestingly, both IZUMO1 and SPACA6 are part of the immunoglobulin superfamily, potentially suggestive of ancient orthology and potential similarities in biochemical mechanisms (Lorenzetti, 2014; Nishimura, 2016). Another notable protein on the egg side is the tetraspanin CD9. Knockouts of *CD9* result in highly reduced fertility due to failure of sperm to bind to the oolemma (Chalbi, 2014). CD81 is another tetraspanin which may be involved in gamete fusion (Ohnami, 2012). Knockouts of *CD81* in mice lead to accumulation of sperm in the perivitelline space, however, the mouse eggs show only a modest decrease in fertility (Rubinstein, 2006). It is hypothesized that both CD9 and CD81 contribute to gamete fusion not by directly interacting with molecules on the sperm, but by recruiting other proteins and fusagens to mediate gamete fusion (Bhakta, 2019).

1.3 Evolution of Gamete Recognition Genes

Successful zygote formation depends on sperm-egg recognition mediated by species-specific molecular interactions. Despite the importance of fertilization for the propagation of a species, a

pervasive characteristic of reproductive proteins mediating gamete recognition across taxa is their rapid diversification, on par with proteins mediating immune response and transposon regulation (Mclaughlin & Malik, 2017). Most of the genes discussed in this review (abalone *lysin, VERL*, and *sp18*, sea urchin *bindin, suREJ-1, Obi1* and *EBR1*, and mammalian *PKDREJ, ZP2, ZP3, Izumo1*, and *Juno*) have been shown to be under positive selection (Table 1) (Grayson, 2015; Swanson, Yang, et al., 2001; Mah, 2005; Hamm, 2007; Hart, 2014; Metz, 1998; Swanson, 2001; Morgan, 2010; Swanson, 2003; Galindo, 2003; Zigler, 2005; Zigler & Lessios, 2003; Turner & Hoekstra, 2006). However, other studies have failed to identify positive selection in some of these same genes (Morgan & Hart, 2019). Methodological differences and choice of sequences to include can affect detection of positive selection. In addition GRPs may be under selection in some lineages but not others (Zigler & Lessios, 2003; Zigler & Lessios, 2004). Interestingly, some interacting pairs of proteins mediating sperm-egg recognition have been shown to coevolve. These physically unlinked sperm-egg GRPs can show signatures of coevolution through genetic linkage disequilibrium or correlation of lineage-specific rates of evolution (Grayson, 2015; Rohlfs, 2010; Stapper, 2015; Hart, 2014; Klaw, 2014; Clark, 2009; Hart, 2018). This phenomenon is perhaps unsurprising. Since fertilization relies on these sperm-egg protein-protein interactions, any change in either the sperm or egg protein that alters binding, must affect the function of the other partner and thereby promote coevolution. Yet, it is also striking that selection is strong enough to leave these coevolutionary footprints, suggesting that these coevolutionary signatures could be used as powerful predictors of PPIs (Findlay, 2014).

Polyspermy avoidance, sexual conflict, sexual selection, pathogen avoidance, and reinforcement have all been proposed as evolutionary mechanisms that could drive the rapid diversification of

gamete recognition proteins (Swanson & Vacquier, 2002; Evans, 2013; Vacquier & Swanson, 2011). These processes could alone or in combination drive the adaptive coevolution between sperm-egg GRP pairs and the relative contributions of each process could vary by population characteristics. Distinguishing the potential role of each of these drivers is complicated due to overlapping predictions of the outcomes each would have on the evolution of reproductive proteins (Evans, 2013; Betters & Levitan, 2018). Some research indicates that conditions such as sperm concentration paired with eggs' polyspermy susceptibility and sensitivity to sperm availability may dictate whether sexual selection or sexual conflict may occur (Styan & Butler, 2000; Franke, 2002; Levitan, 2004; Levitan, 2018). It is of interest to decipher the selective pressure driving observed divergence of gamete recognition genes since these changes have been linked to profound effects on sperm-egg recognition function and could contribute to the formation of species boundaries. In addition, since signatures of sequence evolution can guide functional studies of protein-protein interactions and binding domains, it is of interest to understand the underlying evolutionary mechanisms guiding the proteins' evolution.

1.3.1 *Polyspermy Avoidance*

Polyspermy occurs when multiple sperm fuse with the egg, leading to embryo death. In order to prevent polyspermy, diverse mechanisms blocking secondary sperm fusion exist across taxa such as the fast electrical polyspermy block in sea urchins and the slow block cortical reaction in both sea urchins and mammals (Bhakta, 2019; Stewart-Savage & Bavister, 1988; Jaffe, 1976). It has been hypothesized that gamete recognition mechanisms could also play an important role in preventing polyspermy. For example, in mammals, polyspermy is blocked by the cortical reaction which causes elevation and modification of the zona pellucida to prevent entry of additional sperm (Sato, 1979). One modification to the zona caused by this reaction is the cleavage of ZP2 and ZP3

by cortical granule ovastacin (Burkart, 2012). Post-fertilization cleavage of these proteins, both of which are believed to be required for sperm binding, may create an important boundary to secondary fertilization events (Bianchi, Doe, et al., 2014; Burkart, 2012). In addition, the release of decoy receptors post-fertilization may also reduce polyspermy risk. In mammals, egg plasma membrane JUNO is shed into the perivitelline space after sperm-egg fusion (Bianchi, Doe, et al., 2014). This serves two functions. First, shed JUNO is no longer able to mediate sperm-egg plasma membrane fusion. Second, the shed egg GRP can bind and neutralize incoming acrosome-reacted sperm in the perivitelline space to prevent secondary sperm-egg fusion events. In addition to these post-fertilization modifications, rapidly evolving sperm-egg receptors may slow down the fertilization process by slowing down sperm recognition, allowing time for permanent blocks to be established.

Investigation of allelic *su-bindin* diversity in sea urchins has provided some evidence supporting the role of GRPs in preventing polyspermy events. In *S. fransciscanus* sea urchins, researchers have shown that higher embryo formation success is achieved at optimal sperm densities (Levitan, 2019). When sperm densities are too low, sperm limitation reduces the number of successful fertilization events. When sperm densities are too high, increased risk of polyspermy reduces the number of viable embryos (Levitan, 2019; Levitan & Ferrell, 2006). Populations of sea urchins may adapt to non-optimal sperm densities by altering the frequency of specific *su-bindin* alleles (Levitan, 2007). In populations with high sperm densities, surviving embryos were more often fertilized by the rare *su-bindin* alleles with slow binding rates. Meanwhile, in populations with low sperm density, surviving embryos were more often fertilized by sperm carrying common *su-bindin* alleles with high binding rates (Levitan & Ferrell, 2006). This theoretically indicates that the risk

of polyspermy events can be reduced by decreasing the efficiency of gamete recognition steps. In addition, sea urchin species with lower susceptibility to polyspermy have fewer *bindin* alleles than species with higher risks of polyspermy, further supporting the idea that polyspermy avoidance may drive GRP evolution (Levitan, 2007). By understanding how variation in GRP sequence can impact fertilization success in natural populations, we may gain insights into causes of human infertility and into the development of novel contraceptive strategies (Xue, 2016; Wang, 2009).

1.3.2 *Sexual Conflict*

An egg's need to prevent polyspermy is balanced with a sperm's need to outcompete other sperm to be the first to fuse with the egg. This scenario results in a difference in optimal mating rates between gametes and could lead to a phenomenon called sexual conflict (Swanson & Vacquier, 2002). Under this model, egg receptor proteins may alter their sequence to reduce incidence for polyspermy. In response, sperm competition would select for sperm carrying a recognition ligand that is a better fit to the altered egg receptor. This pattern of a sperm recognition protein adapting to bind a continuously changing egg coat receptor protein results in a co-evolutionary chase. Sexual conflict driven by polyspermy prevention may be the selective force underlying the rapid protein diversification, maintenance of allelic polymorphism, and signatures of positive selection often found in GRPs (Swanson & Vacquier, 2002). However, it should be noted that these evolutionary signatures could also be caused by sexual selection or pathogenic escape.

The evolution of su-bindin and EBR1 of the North American sea urchin species *Mesocentrotus fransciscanus* provides a compelling example for how sexual conflict can drive the evolution of sperm-egg GRPs (Levitan, 2019). Sea urchins are known to have a predictable relationship

between lifespan and age, with larger sea urchins being older. M. fransciscanus sea urchin life-spans may exceed 200 years (Ebert, 2008). Over the past 200 years, *M. fransciscanus* sea urchin population densities have increased approximately 100 fold. This expansion is the result of human predation driving sea otters, the main predator of sea urchins, to near extinction (Levitan, 2019). Such a drastic increase in sea urchin population densities is predicted to increase polyspermy rates. In a remarkable study, sea urchins with an age range spanning 200 years were genotyped for *bindin* and *EBR1* in order to investigate how this population may have applied strategies to reduce polyspermy risk mediated by GRPs (Levitan, 2019). The observed pattern showed that the historically common RG *bindin* allele and TSP8 A *EBR1* allele were highly compatible in fertilization. However, progressively smaller/younger sea urchins showed a rapid increase in the frequency of the historically rare TSP8 G *EBR1* allele followed more slowly by an increase in the less common GR *bindin* allele. Ultimately, the most recently conceived sea urchins were observed to have these newly common alleles at approximately the same frequencies as the historically common alleles. The GR *bindin* allele and TSP8 G *EBR1* allele are highly compatible with each other but mismatches with the historical RG *bindin* and TSP8 A *EBR1* group are less compatible. One hypothesis to explain this phenomenon is that the new TSP8 G egg allele arose to reduce the efficiency of *bindin-EBR1* binding and that the GR *bindin* allele compensated for this decrease in binding efficiency. This pattern of female egg receptor innovation to decrease polyspermy risk followed more slowly by an increase in frequency of compensatory sperm alleles to regain binding rates is strongly suggestive of GRP coevolution via sexual conflict.

1.3.3 *Sexual Selection*

When sexual selection occurs, female choice drives selection for improved offspring reproductive fitness via arbitrary cryptic female choice criteria. This differs from scenarios involving sexual

conflict where female choice is acting to prevent a reduction in female fitness (e.g. polyspermy risk) (Vacquier & Swanson, 2011; Holt & Fazeli, 2014; Kekalainen & Evans, 2018). Sperm are more abundant and less costly to create than eggs. This may cause sperm to be under high selective pressure to compete for fertilization events. Meanwhile with an abundance of sperm, eggs may be under relaxed selection for maintaining binding compatibility (Vacquier & Swanson, 2011). Under these conditions, sexual selection could cause the coevolution of sperm-egg GRPs. For example, new variation arising in the egg GRP via genetic drift could reduce sperm binding ability. As a result, sexual selection may act on sperm GRP variation present in the population to preferentially select sperm carrying GRPs with compensatory mutations for the new mutation in the egg GRP. Under models of GRP evolution due to sexual selection, signatures of positive selection are expected to be higher in the sperm protein, indicative of higher directional selection. Studies in marine invertebrates indicate that eggs are highly sensitive to sperm concentrations resulting in either sperm limitation or polyspermy, potentially limiting the conditions under which sexual selection may occur (Levitan, 2004; Levitan, 2018).

1.3.4 *Pathogen Avoidance*

As surface-exposed proteins, gamete recognition systems could potentially be harnessed by pathogens to invade gametes. This may be particularly true for externally fertilizing species since the fertilized eggs settle into a microbe-rich extracellular environment. If this is the case, then GRPs would need to constantly be changing to escape pathogen binding events. According to the Red Queen Hypothesis, a coevolutionary chase between the pathogenic predator and gametic prey could cause accelerated evolution of GRPs independent of these proteins' roles in fertilization. In this scenario, the observed rapid coevolution of sperm-egg GRP pairs to maintain binding to each other would be indirectly driven by host-pathogen coevolutionary conflict. This hypothesis for

explaining the evolution of GRPs is particularly appealing since genes involved in immunological defense also tend to be rapidly diversifying, presumably due to conflict dynamics. As in the case of sexual conflict caused by polyspermy avoidance, female receptor evolution caused by pathogen avoidance is driven by natural selection to prevent embryo loss. This scenario then creates selective pressure on the sperm GRP and in turn drives its evolution. In this scenario, pathogen avoidance applies indirect selection on the sperm protein's evolution.

1.3.5 *Reinforcement*

Many of the gamete recognition genes previously discussed have been demonstrated to mediate species-specific fertilization function. A possible driver for the rapid adaptive evolution of gamete recognition genes is selection to generate these post-mating, pre-zygotic barriers to hybridization (Swanson & Vacquier, 2002; Vacquier & Swanson, 2011). By reinforcing species boundaries, species-specific gamete recognition systems can help prevent the creation of less fit offspring or hybrid embryo death caused by genomic incompatibilities. Reinforcement does not describe a conflict scenario between the sperm-egg GRP pairs but instead refers to natural selection acting against hybridization events. However, the presence of species-specific binding between GRP pairs does not necessarily indicate reinforcement drove its creation. Species-specific binding can also arise as a natural consequence of the independent evolutionary trajectories of reproductive proteins occurring within individual species. Reinforcement occurs only post-speciation and can contribute to the strengthening of species boundaries.

Reinforcement would be expected to drive the evolution of GRPs in sympatric species but not allopatric species. Sea urchins once again prove to be a useful model to examine this hypothesis. Allopatric species of sea urchin from multiple genuses have been shown to have low rates of *su-*

bindin divergence compared to closely related species and no signatures of positive selection (Zigler & Lessios, 2003; Zigler & Lessios, 2004). This is the converse of what is normally found in *su-bindin* ofi sympatric *su-bindin* species (Palumbi & Lessios, 2005). Allopatric species are not at risk of hybridization events and therefore would not need to evolve species-specific GRP interactions. These results provide indirect evidence that reinforcement could be a significant driver of GRP evolution in some species.

1.4 Evolution of Species-Specific GRP Binding

By combining evolutionary insight with knowledge of protein structure, it may be possible to investigate the biochemical origins of the species-specific functions of gamete recognition proteins. The rapid sequence divergence of GRPs, particularly in their binding domains, may lead to species-specific binding interactions between sperm and egg proteins and create boundaries to hybridization. Abalone's coevolving lysin and VERL proteins provide an intriguing example of how the coevolution of GRPs may create a species-specific gamete recognition mechanism (Clark, 2009). In both proteins, sites under positive selection are enriched at the binding interface (Raj, 2017; Wilburn, 2018). In VERL, sites under positive selection in the first N-terminal ZP-N domain are clustered around the fourth and fifth Beta-sheet, the putative lysin binding site for this domain (Raj, 2017). In lysin, 13 sites under positive selection are clustered within the N- and C-terminus and create a "nexus" of positively selected sites within lysin's VERL binding interface (Wilburn, 2018). Chimeric lysin experiments have confirmed the importance of the N- and C- terminus for species-specific VE dissolution function. Together these results indicate that lysin's N- and C- terminal sequence divergence may mediate its functional specificity (Lyon & Vacquier, 1999). By combining experimental, structural, and evolutionary research on this sperm-egg GRP pair, it seems likely that positive selection in the

binding interface of lysin and VERL leads to lysin's species-specific egg coat dissolution function. Species-specific binding interactions between sperm-egg GRP pairs in other organisms may also be a natural byproduct of the coevolution of interacting GRPs. However, the formation of species-specific interactions of GRPs does not necessarily indicate that creating boundaries to hybridization events is the driver of GRP coevolution. Further analysis of GRP structure, function, and evolutionary history in other systems could provide new insights into the biochemistry of sperm-egg binding and the creation of molecular boundaries to hybridization (Findlay, 2014).

1.5 Conclusion

In this review we describe gamete recognition proteins from abalone, sea urchins, and mammals and explore the evolutionary pressures mediating their rapid diversification and coevolution. Although fertilization is an essential process for most animals, mechanisms of fertilization broadly remains a mystery. A common theme of GRPs found within the external fertilizers abalone and sea urchins and within internally fertilizing mammals is that many GRPs are undergoing positive selection. Although the evolutionary mechanisms mediating this pervasive characteristic of GRPs are difficult to determine, the diversification of GRPs between closely related species can lead to the formation of post-mating, pre-zygotic boundaries to hybridization. The rapid diversification of these proteins can hinder detection of homologs even in closely related species. Yet, despite this diversification, homologous protein structures are present in diverse taxa, e.g. abalone and mammalian egg coats contain proteins with ZP-N domains. Future investigations into fertilization mechanisms in more taxa across the tree of life may fill in important phylogenetic gaps to determine the origins of important GRPs and identify novel molecular mechanisms relevant to existing animal fertilization models.

1.6 ACKNOWLEDGMENTS

This work was supported by NIH grant HD076862 to W.J.S and a National Science Foundation Graduate Research Fellowship to J.A.C. Special thanks to Dr. Damien B. Wilburn, Dr. Matthew B. Taylor, Dr. Katherine Xue, and Alberto Rivera for discussions and comments.

1.7 FIGURES AND TABLES

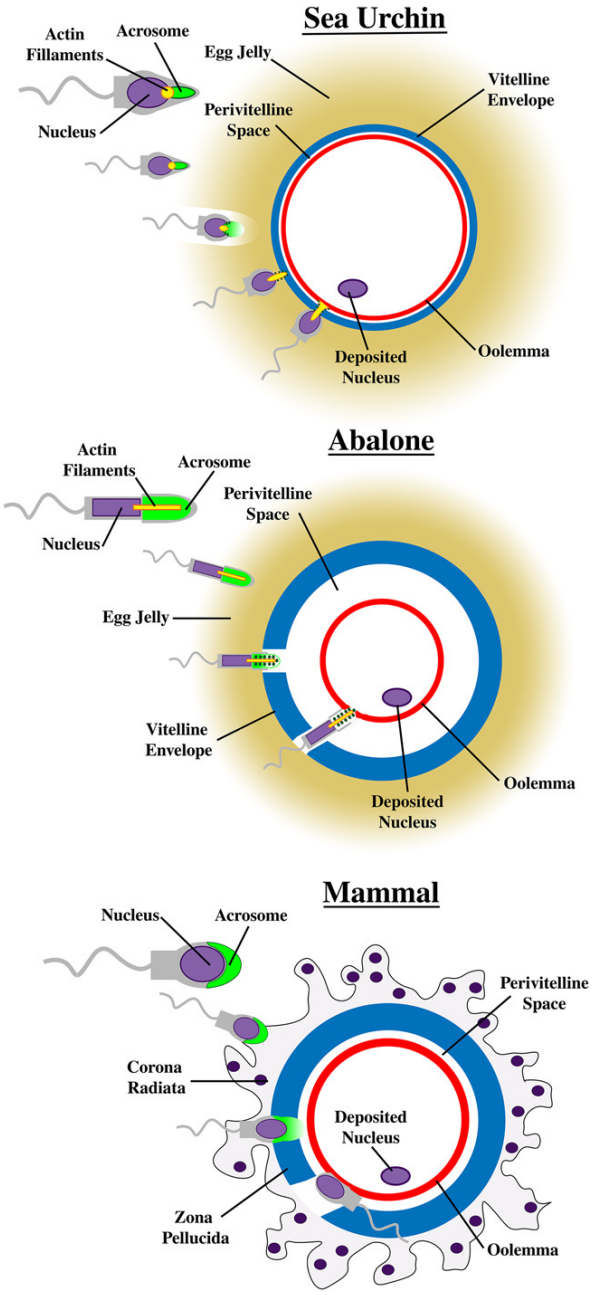


Figure 1.1. Overview of fertilization in sea urchins, abalone, and mammals.

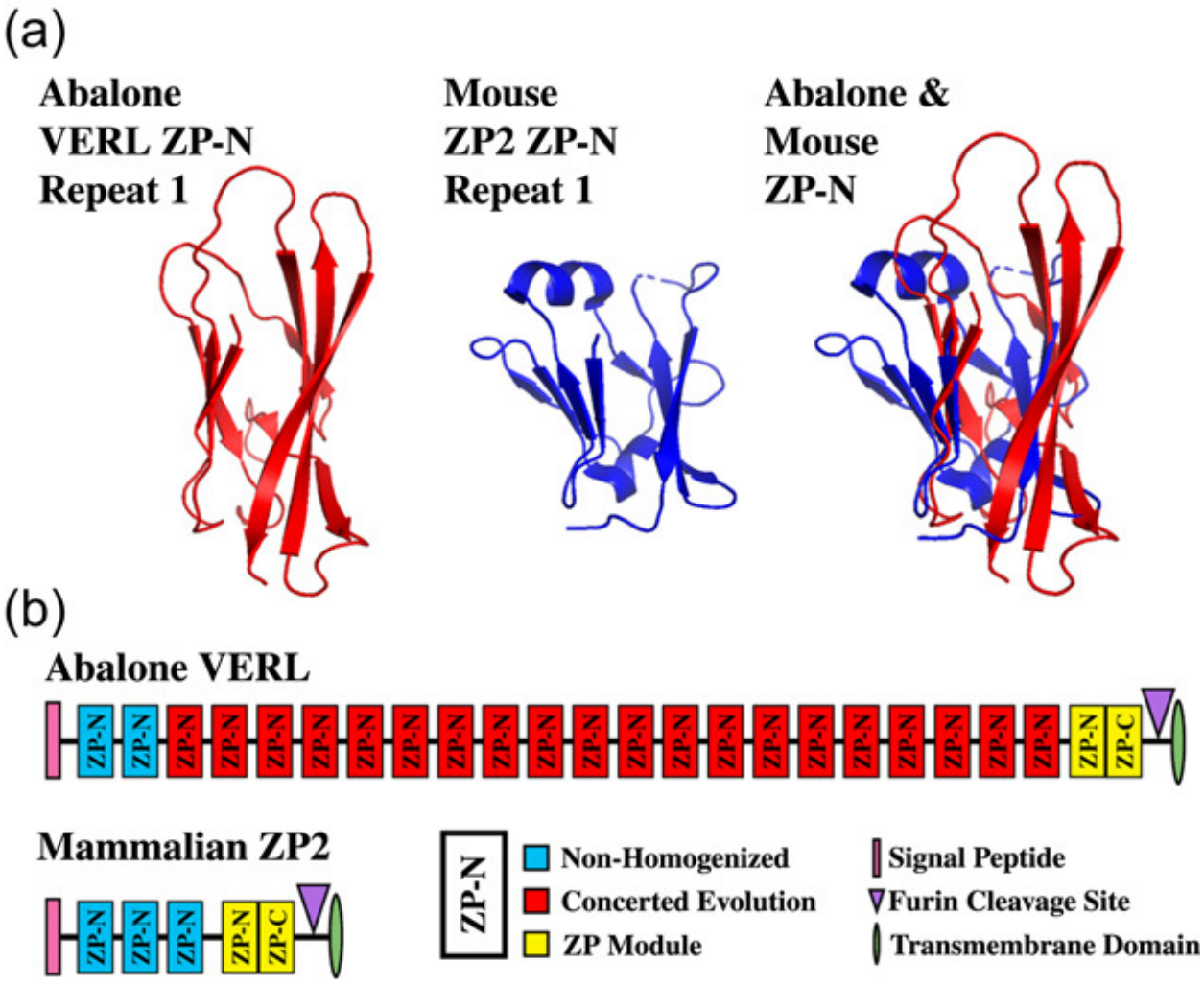


Figure 1.2. Protein and gene structures of abalone and mammal ZP proteins.

(a) Crystal structures of the N-terminal ZP-N domain of abalone VERL and human ZP2 proteins alone and overlaid. (b) Gene structures of abalone VERL and mammalian ZP2. Both egg coat ZP proteins contain tandem repeats of ZP-N domains with a C-terminal ZP module. VERL, vitelline envelope receptor for lysin.

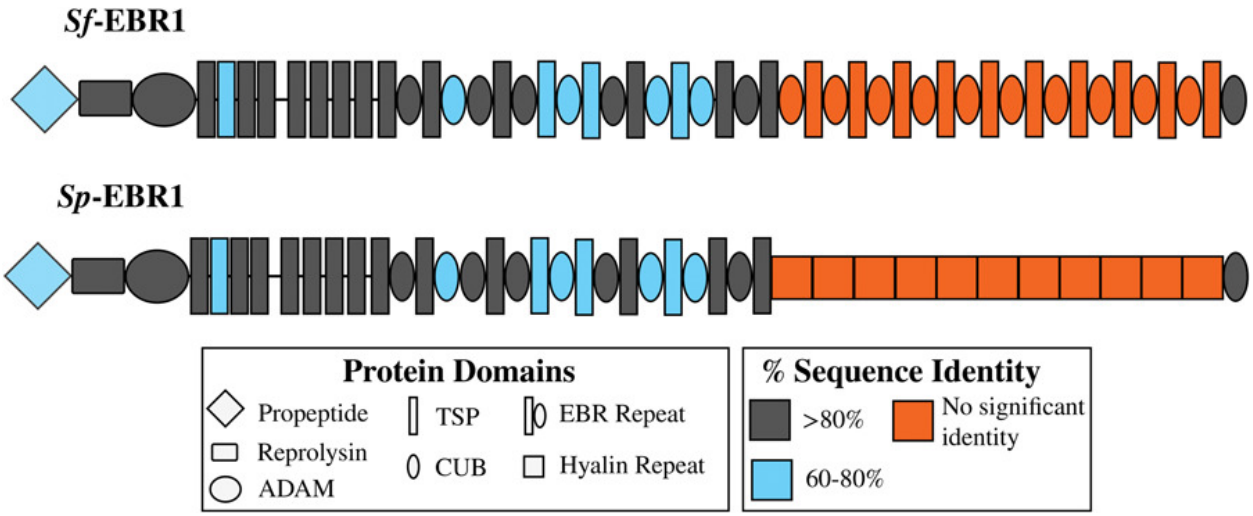


Figure 1.3. Comparison of EBR1, the egg-binding partner of sperm bindin, between Strongylocentrotus franciscanus and S. purpuratus sea urchin species.

The C-terminal region of EBR1 after the ninth EB repeat is not similar between the two species in either sequence identity or protein domains. This species-specific EBR1 sequence could be important for mediating species-specific sperm binding by the egg in these sea urchins. EBR1, egg bindin receptor 1.

	Animal	Sperm	Egg	Positive selection?
Acrosome reaction	Sea urchin	su-REJ	FSP	Yes (su-REJ only)
	Abalone	?	?	?
	Mammal	PKDREJ	?	Yes
	Mammal	ZP3r/sp56	ZP3	Yes (ZP3 only)
Egg coat	Sea urchin	?	?	?
	Abalone	Lysin	VERL	Yes
	Mammal	?	ZP2	Yes
Oolemma adhesion/ fusion	Sea urchin	Bindin	EBR1	Yes
	Abalone	sp18	?	Yes
	Mammal	Izumo1	Juno	Yes
	Mammal	SLLP1	SAS1B	?

Table 1.2. Significant GRP models from sea urchins, abalone, and mammals.

Note: This table is not comprehensive, redundant mechanisms mediating steps of fertilization are likely to exit. No animal model has genes identified for every step of fertilization. Positive Darwinian selection is a common feature of GRPs.
Abbreviations: GRP, gamete recognition protein; VERL, vitelline envelope receptor for lysin.

Chapter 2. RECURRENT DUPLICATION AND DIVERSIFICATION OF ACROSOMAL FERTILIZATION PROTEINS IN ABALONE

A version of this chapter is available at:

Carlisle JA, Glenski, M.A., Swanson, W.J. 2021. Recurrent Duplication and Diversification of Acrosomal Fertilization Proteins in Abalone. BioRxiv.

2.1 INTRODUCTION

Despite their essential role in many organisms, genes functioning in fertilization or sexual reproduction often rapidly diverge between closely related species including mammals, birds, fish, and invertebrates (Swanson and Vacquier 2002; Swanson, et al. 2003; Carlisle and Swanson 2020). Some pairs of interacting sperm and egg gamete recognition proteins have been shown to be rapidly co-evolving, pointing to sexual conflict or sexual selection driving the rapid evolution of fertilization genes (Kamei and Glabe 2003; Clark, et al. 2009; Bianchi, et al. 2014; Grayson 2015). This rapid diversification of sperm and egg gamete recognition proteins at a sequence level can result in species-specific fertilization (Zigler, et al. 2005; Avella, et al. 2014; Raj, et al. 2017). Investigations into the evolution of reproductive genes paired with characterization of species-specific function can provide unique insights into infertility and reproductive isolation (Lehmann 2018). In addition to sequence evolution, gene duplication can also contribute to the molecular diversification of reproductive protein families. *Drosophila* seminal fluid proteins often undergo duplication and diversification, with many of these duplications being species-specific (Wagstaff and Begun 2005; Findlay, et al. 2008; Almeida and Desalle 2009; Sirot, et al. 2014; Doty, et al. 2016; Wilburn 2017). Mammalian genes mediating fertilization including *CatSper*, *Izumo*, and the *ZP* gene families also demonstrate interesting patterns of duplication (Cai and Clapham 2008;

Aagaard, et al. 2010; Grayson and Civetta 2012; Cooper and Phadnis 2017). Here, we investigate how both sequence diversification and duplication together contribute to the molecular diversification of fertilization proteins.

The marine gastropod abalone (genus *Haliotis*) is a classic model system for studying the function of gamete recognition proteins and their evolution. Abalone sperm have an extremely large acrosome containing two gamete recognition proteins, lysin and sp18 (Lewis, et al. 1980). Lysin is the sperm borne protein that helps dissolve the egg's vitelline envelope (VE), whereas sp18 is thought to mediate sperm-egg plasma membrane fusion (Wilburn, et al. 2018; Carlisle and Swanson 2020). The abalone egg VE is an elevated glycoproteinaceous layer homologous to the mammalian egg zona pellucida (ZP) (Carlisle and Swanson 2020). The abalone VE and mammalian ZP are biochemically and structurally similar (Mozingo, et al. 1995) and both contain proteins with ZP-N domains (Avella, et al. 2014; Carlisle and Swanson 2020). Binding between lysin and the ZP-N domains of the vitelline envelope receptor for lysin (VERL) leads to the non-enzymatic dissolution of the VE (Swanson and Vacquier 1997; Aagaard, et al. 2013; Raj, et al. 2017). Three-dimensional structures of lysin and VERL have been investigated using NMR and crystallography and their ability to bind each other have been quantified (Kresge, et al. 2000, 2001; Aagaard, et al. 2013; Raj, et al. 2017; Wilburn, et al. 2018). The highly fusogenic protein sp18 is the putative mediator of sperm-egg plasma membrane fusion in abalone (Swanson and Vacquier 1995; Kresge, et al. 2001). The structure of sp18 has been determined via crystallography, but its binding receptor is unknown (Kresge, et al. 2001).

Investigations into the evolution of abalone reproductive proteins have provided valuable insights within the field of reproductive biology. *Lysin*, *sp18,* and *VERL* have each been shown to evolve

under positive selection by analysis of the ratio of rates of nonsynonymous substitutions to rates of synonymous substitutions (dN/dS > 1) (Lee, et al. 1995; Swanson and Vacquier 1995; Galindo, et al. 2003). Further, population genetic analysis indicates that *lysin* and its binding partner *VERL* coevolve with each other (Clark, et al. 2009). Previous studies of abalone egg and sperm gamete recognition proteins hint at the importance of duplication events for their evolution. Despite their divergent functions in reproduction and low sequence similarity, sp18 and lysin are paralogs with similar three-dimensional protein structures (Kresge, et al. 2001). The common ancestor of lysin and sp18 is hypothesized to have mediated both VE dissolution and sperm egg fusion, then parsed these combined ancestral functions post-duplication to lysin and sp18, respectively.

Additional evidence suggests that duplication may be contributing to the evolution of abalone sperm fertilization proteins on a more recent timescale. The European abalone *H. tuberculata* has a species-specific duplication of lysin (Clark, et al. 2007). On the egg side there is also evidence of extensive gene duplication. In addition to VERL, abalone VEs contain ~30 homologous VEZP proteins containing ZP-N domains (Aagaard, et al. 2006; Aagaard, et al. 2010). Many of these VEZPs may be structural components of the VE that play no role in gamete recognition (Killingbeck and Swanson 2018). However, one protein (VEZP-14) is a close paralog of VERL that has undergone positive selection and is capable of binding lysin (Aagaard, et al. 2013). Abalone VEZPs have only been described in two species (*H. rufescens*, and *H. fulgens*) from the North American clade. It is unknown whether there is variation in gene content across more distantly related abalone species (Aagaard, et al. 2006).

In this study, we investigated the contributions of duplication and sequence diversification to the evolution of proteins mediating fertilization across the genus *Haliotis*. Using new transcriptomic

data from testes and ovaries and published genome assemblies, we discovered novel duplications of lysin and sp18. Some of these paralogs are ancestral to abalone speciation whereas others are clade-specific. Furthermore, we discovered signatures of positive selection in many of the paralogs and identified patterns of positively selected sites indicative of specialization. Our detailed evolutionary genomic analyses reveal how recurrent patterns of duplication together with diversification led to the evolution of abalone gamete recognition proteins and their variation between species. Repeated duplications within the protein family containing lysin and sp18 parallels the duplication and diversification of reproductive protein families in other lineages, such as mammalian Izumo and Catsper families.

2.2 Materials and Methods

2.2.1 *PacBio Library Preparation and Sequencing*

To identify potential transcripts present in abalone gonadal tissue, methods were adapted from the PacBio Iso-seq protocol to create cDNA libraries. Ovary and testes transcriptome libraries were prepared for PacBio sequencing. RNA was extracted from *H. tuberculata* ovary and testes samples by cesium chloride density gradient centrifugation (MacDonald, et al. 1987). RNA samples were enriched for mRNA by using the Oligotext mRNA Mini Kit from Qiagen. Purified mRNA was used as the template for single stranded cDNA synthesis using the Clontech SMARTer cDNA Synthesis Kit. The cDNA was amplified by PCR using the AccuPrime High-Fidelity Taq system (Invitrogen, Carlsbad, CA, USA). Double stranded synthesis conditions were optimized with the following PCR program: °C for 2 minutes, followed by 20 cycles of 94°C for 30 s, 55 °C for 30 s, and 68 °C for 10 minutes. We used unique identifying barcoded PCR primers to amplify testes (barcode: CTGCGTGCTCTACGAC) and ovary (barcode: TCAGACGATGCGTCAT) cDNA. Because of size bias during PacBio sequencing, double stranded cDNA was fractionated using

Ampure XP Beads (Beckman Coulter Life Sciences) into two fractions (Ratio of 4:1 of 0.45x : 0.6x size selection). Testis and ovary cDNA were sequenced using the PacBio RSII and was performed by the Washington State University Genomics Core.

2.2.2 *Identification of Acrosomal Protein Paralogs*

Sequences of *sp18* and *lysin* from the genus *Haliotis* were retrieved from NCBI Genbank sequence repository (accession numbers for lysin: L26270-79, L26281-83, L35180-81, L36589, M34388-89, M59968-72, M98874-75, HM582239; accession numbers for *sp18*: L36552-54, L36589-90, MN102340-42). These sp18 and lysin sequences were used as the initial query sequences when identifying paralogs in abalone transcriptomes and genomes. Queries of the *H. rufescens* Illumina-based testis transcriptome (Palmer, et al. 2013) and the *H. tuberculata* PacBio testes transcriptome were conducted with tblastn with an e-value cutoff of 1e-10. Significant matches from the testes transcriptomes were searched against the NCBI sequence repository (July 2020) using tblastn in order to confirm homology to lysin or sp18 (McGinnis and Madden 2004). New sequences were uploaded to Genbank under accession numbers OK491874-OK491877.

Regions of publicly available abalone genomes containing novel acrosomal protein duplications of sp18 and lysin were identified by using tblastn with a e-value cutoff of 1e-10 (Nam, et al. 2017; Botwright, et al. 2019; Gan, et al. 2019; Masonbrink, et al. 2019). Samtools faidx was used to extract the region of scaffolds containing the tblastn hits and 20,000 base pairs upstream and downstream of the hit. We predicted the exonic sequences of the *sp18* and l*ysin* paralogs from these extracted regions using the Protein2Genome command of the program Exonerate version 2.2.0 (Slater and Birney 2005). The top scoring prediction from Exonerate was used to define the paralog's exons. We used the same lysin and sp18 sequences from the tblastn search as query

sequences. For all full-length sequences, the SignalP-5.0 prediction server was used to predict presence of functional signal peptides (Almagro Armenteros, et al. 2019). Presence of signal peptides were predicted with probabilities >0.9; the signal peptide cleavage site was predicted with a probability >0.5.

2.2.3 *Phylogenetic Analysis*

The phylogenetic inference tool RAxML-NG was used to construct all phylogenetic trees with the LG substitution matrix (Le and Gascuel 2008; Kozlov, et al. 2019). RaxML-NG conducts maximum likelihood based phylogenetic inference and provides branch support using non-parametric bootstrapping (Kozlov, et al. 2019). The best scoring topology of 20 starting trees (10 random and 10 parsimony-based) was chosen. RaxML-NG was used to perform non-parametric bootstrapping with 1000 re-samplings that were used to re-infer a tree for each bootstrap replicate MSA. Finally, we mapped the bootstrap scores on the best-scoring starting tree. The Transfer Bootstrap Expectation (TBE) was used as a branch support metric (Lemoine, et al. 2018).

DNA multiple sequence alignments (MSA) for phylogenies of *lysin* and *sp18* and their respective paralogs were constructed. First the protein sequences of the genes were aligned using PROMALS3D (Pei, Kim, et al. 2008; Pei, Tang, et al. 2008). PROMALS3D uses three-dimensional protein structures to inform protein alignments, and representative PDBS for lysin (5UTG) and sp18 (1GAK) were specified (Kresge, et al. 2000; Wilburn, et al. 2018). The protein MSAs were used to create DNA alignments of the same genes using the Pal2Nal sever (Suyama, et al. 2006). For phylogenetic analysis of *H. rufescens* and *H. tuberculata* VEZP sequences, the protein sequences of the C-terminal ZP modules of each of the proteins were aligned using PROMALS3D (Pei, Tang, et al. 2008).

2.2.4 *Syntenic Comparison Between Haliotis rufescens and H. rubra*

The published genome of *Haliotis rufescens* is annotated with ORFs identified via transcriptomic sequencing (Masonbrink, et al. 2019). We collected the sequences of 2-3 large annotated ORFs surrounding lysin, sp18, and their newly described paralogs within the *H. rufescens* genome. We used these sequences as BLAST queries against the *H. rubra* genome (Gan, et al. 2019). The top hits for the *H. rufescens* ORFs were annotated onto the *H. rubra* genome and used to establish synteny between *H. rubra* scaffold 62 and the *H. rufescens* scaffolds 48 and 101. A reciprocal blast of the regions identified as orthologous ORFs in *H. rubra* were queried against the *H. rufescens* genome to verify orthology.

2.2.5 *Detecting Selection and Positively Selected Sites*

Values of dN/dS for genes were estimated using the codeml program of PAML 4.8 (Yang 2007). We compared models of selection using a likelihood ratio test (LRT) between neutral models and models with positive selection. Specifically, we compared M1 v. M2, M7 v. M8, and M8a v. M8. Likelihood ratio tests were performed where the likelihood ratio (LRT) statistic was twice the negative difference in likelihoods between nested models. For M1a v M2a or M7 v Model 8 the LRT was compared to the $\chi 2$ distribution with 2 degrees of freedom (Yang 2007). For the M8a v. M8 comparison, twice the negative difference in likelihoods between the nested models being compared, the LRT statistic, is approximated by the 50-50 mixture distribution of 0 and $\chi 2$ with 1 degree of freedom (Swanson, et al. 2003). To identify specific sites in proteins evolving under positive selection, we used Naive Empirical Bayes (Pr(omega > 1) = 0.95). The sites under positive selection were identified using Model 8.

2.2.6 *Testing for Divergence in Regions Undergoing Positive Selection in Duplicate Sperm Proteins*

We designed three unbiased tests to determine if sites under positive selection in either sp18 or sp18-dup are differentially clustered between paralogs. First, we created a parametric test based on the Wald-Wolfowitz runs test (Magel 1997) to determine whether positively selected sites in the paralogs sp18 and sp18-dup were non-randomly distributed in a protein alignment of both paralogs. The Wald-Wolfowitz runs test determines the randomness of a two-category data string by examining changes between categories by counting "runs." We designed a parametric version of the test to allow the inclusion of three categories. The categories were sites under positive selection in sp18, sites under selection in sp18-dup, and sites under selection in both. The order in which these sites under selection in the categories appeared in an alignment of *H. fulgens* sp18 and *H. sorenseni* sp18-dup became our data string. For the data string generated from our paralog alignment we counted how many times the identity of sites in the string changed plus one. A visualization of the pipeline for preparing this data string and counting "runs" is shown in Figure 1.

To make a parametric version of this runs test, we generated 1000 simulated data strings via bootstrapping based on the proportion of sites shown to be under positive selection in either paralog. For each of these simulated data strings we also calculated the number of runs. The number of data strings with a count of "runs" less than or equal to the count of "runs" found in the data string derived from the paralog protein alignment divided by the total number of simulated data strings gives the parametric probability that by random chance categories of sites would be more or equally clustered compared to the true clustering observed. We also performed a version of the test where sites under selection in both paralogs were eliminated from the analysis. When

these sites were eliminated, the parametric runs test retained statistical significance (p-value < 0.001).

For our second test we evaluated whether the distribution of sites under positive selection differed significantly between sp18 vs sp18-dup using the sp18 crystal structure (1GAK). In MATLAB Online version 9.9, we identified the plane of best fit between the c-alpha carbons (the first carbon attached to the functional group of an amino acid) of the sp18 crystal structure using linear regression (Figure 2A). This plane divided the sp18 crystal structure into two sides that we arbitrarily designated "left" and "right" (Figure 2B). We mapped the 43 sites in sp18 and the 33 sites in sp18-dup that are under positive selection onto the sp18 crystal structure. Given the number of amino acids sites in each side of the crystal structure, we estimated the expected number of positively selected sites from each paralog that would be expected to be located on either side as the number of amino acid sites on a side divided by the total number of amino acid sites in the molecule and multiplied by the number of positively selected sites in a sp18 paralog. We used a chi-square test to examine whether the real distribution of sites between categories rejected the null expectation. This test determined whether the distribution of sites under selection in either paralog was not distributed similarly between the sides of the protein. We also created a plane perpendicular to the plane of best fit to divide the sp18 crystal structure into the categories "top" and "bottom" (Figure 2C). We repeated an analysis for this new pair of categories that is identical to what was described previously. This analysis gave us a sense if sites under selection in either paralog were clustered non-randomly throughout the crystal structure in different ways.

For our last clustering test we determined whether sites under positive selection in a paralog were more likely to be close in proximity in three-dimensional space to another site under positive

selection in the same paralog rather than a site under positive selection in the other paralog. For each site that was under selection in a paralog, we identified the closest positively selected site in three-dimensional space that belonged to either paralog. We then calculated the expected number of times by chance the closest adjacent site for each site under positive selection would belong to the same gene rather than the other paralog. We used a chi-square test to determine whether observed sites under selection in one paralog were statistically more likely to be close to positively selected sites belonging to the same paralog than what would be expected by chance. This test has four categories of sites (for each paralog the nearest site could belong to the same paralog or not), and therefore three degrees of freedom were used to determine the p-value. Sites that were under selection in both paralogs were counted twice in this analysis since these sites were undergoing positive selection independently in both paralogs. When the closest adjacent site to a positively selected site in one paralog was undergoing positive selection in both paralogs, the adjacent site was treated as belonging to the same paralog.

2.2.7 *Identification of SP18 Peptides*

H. tuberculata testis tissue was homogenized in 1% sodium dodecyl sulfate with BME at 70 °C for 30 minutes. Testis samples were separated by SDS-PAGE using a Tris-Tricine buffering system with discontinuous 4% resolving/15% separating acrylamide gels. Samples were electrophoresed at 50V for 15 min followed by 100 V for 90 min. The gel was run with the BioRad Broad Range Ladder and stained with Coomassie Blue R-250 for 15 minutes. Using the ladder as reference, the lysin and sp18-containing region (~14-22 kDa) of the polyacrylamide gel was excised using a clean scalpel, with multiple rounds of perfusion with an ammonium bicarbonate solution followed by acetonitrile to extract detergents and salts. Trypsin proteolysis of immobilized proteins was by perfusion of a Trypsin solution (40 ug/ml stock Trypsin 1:10 in 50 mM ammonium

bicarbonate) and incubation at 37°C overnight. The supernatant from the digest was collected along with the supernatant from two rounds of hydration with ammonium bicarbonate and extraction with 50% acetonitrile. The collected supernatant containing the liberated peptides was concentrated to a dry pellet using a vacuum centrifuge then reconstituted in 0.1% FA for liquid chromatography tandem mass spectrometry (LC/MS-MS). Unique peptides for sp18 copy #1 were identified in the sample using the Crux toolkit comet command (Park, et al. 2008). The protein sequence database was composed of a six-frame translation of the H. tuberculata testis transcriptome.

2.2.8 *Identification of ZP Proteins*

An exhaustive BLAST search of the *H. tuberculata* ovary transcriptome identified all cDNA sequences with homology to *H. rufescens* VEZPs. A previous study used a similar approach to originally identify known VEZPs in *H. rufescens* indicating that this approach should be sufficient to identify novel VEZPs (Aagaard, et al. 2010). All cDNA sequences that matched VEZPs were filtered for duplicates using CD-HIT-EST with a threshold of 0.9 sequence identity (Huang, et al. 2010). The longest sequence from each cluster created by CD-HIT-EST was chosen as the cluster's representative sequence. All *H. tuberculata* sequences from this filtering process were translated and the C-terminal ZP modules were identified by identifying conserved cysteine residues. The ZP module protein sequences from both *H. tuberculata* and *H. rufescens* were aligned using PROMALS3D (Pei, Tang, et al. 2008). The MSA of these ZP modules from were used to construct a VEZP homolog protein phylogeny using the same RAXML-NG protocol described above for lysin and sp18 paralog phylogenies. New sequences were uploaded to Genbank under accession numbers OK491878- OK491909.

2.3 RESULTS

2.3.1 *Genomic Analysis Reveals Tandem Duplications of Ancestral Abalone Acrosomal Proteins*

By pairing phylogenetic and genomic analysis of abalone species belonging to the North American clade (*H. rufescens, H. sorenseni, H. discus*) and the Australian clade (*H. rubra, H. laevigata*), we identified ancestral duplications of both sp18 and lysin (Figure 3; sp18-dup and lysin-dup, respectively). We calculated maximum likelihood DNA phylogenies independently for lysin and sp18 with their paralogs and rooted the phylogenies by orthology (Figure 3A & 3B). Predicted intron/exon boundaries of the novel acrosomal protein paralogs were shared with lysin and sp18 (Metz, et al. 1998). No mutations causing pseudogenization were detected within the predicted CDS of either paralog. For the abalone species with published genomes, only one (*H. rufescens*) has a published testes transcriptome (Palmer, et al. 2013). Full-length sequences of lysin, sp18, and sp18-dup are expressed in the testes transcriptome of *H. rufescens*; however, lysin-dup was not detected.

Sequence analysis is consistent with the sp18-dup gene encoding a functional reproductive protein ancestral to *Haliotis*. Sp18-dup is predicted to have a signal peptide sequence and maintains a pair of cysteine residues involved in forming a structurally important disulfide-bond in sp18 (Kresge, et al. 2000, 2001). Sp18-dup has not been identified in previous analysis due to its high divergence from sp18 (27.5% sequence identity between *H. rufescens* sp18 paralogs) obscuring homology.

Lysin-dup was identified in all abalone genomes investigated but was not detected in the previously published short-read sequencing-derived testes transcriptome of *H. rufescens* (Palmer, et al. 2013). The absence of *lysin-dup* in the testes transcriptome could indicate insufficient read depth,

differences in tissue-expression, or potentially pseudogenization. Since the full-length sequence of *lysin-dup* was not identified within the *H. rufescen*s testis transcriptome, *lysin* sequences were used instead to identify *lysin-dup* exons within abalone genomes. However, divergence between *lysin* and *lysin-dup* likely prevented the identification of full-length coding sequence from abalone genomes. Only exons 2-4 could be identified (79% of query sequence) within *H. rufescens* and *H. rubra*. The intron/exon structure of lysin-dup was shared with lysin and sp18. The missing exons 1 and 5 of lysin-dup contain the signal peptide and the N- and C-termini of the molecule. Direct searches of the genomic region surrounding exons 2-4 did not reveal these exons. In lysin, the N- and C-terminus are under strong positive selection promoting extensive divergence that reduces the ability to identify these exons using homology-based approaches (Lee, et al. 1995; Lyon and Vacquier 1999).

In the Australian abalone genomes *lysin*, *lysin-dup*, *sp18*, and *sp18-dup* are all found within a single contig with 233 kb separating the paralog pair of *lysin* and *lysin-dup* from the paralog pair of *sp18* and *sp18-dup*. However, in the genome of the North American abalone species *H. rufescens*, the paralog pair of *lysin* and *lysin-dup* are on a separate scaffold from the paralog pair of *sp18* and *sp18-dup*. We compared the Australian contig containing the four acrosomal protein paralogs with the two *H. rufescens* contigs containing the lysin and sp18 paralog pairs respectively (Figure 3C). We found several ORFs surrounding each paralog pair in *H. rufescens* that were found in the same order as in *H. rubra*, indicating synteny between scaffolds. All four acrosomal proteins being located near each other in the same scaffold in the *H. rubra* genome suggests that tandem duplication led to recurrent duplications of this protein family (Reams and Roth 2015). The sp18 ORF is in a different orientation than the other paralogs, suggesting that transposition and inversion may have also contributed to duplications within this protein family (Reams and Roth 2015).

2.3.2 *Patterns of Divergence of Ancestral Acrosomal Protein Paralogs Indicate Subfunctionalization*

Lysin, *sp18*, and *sp18-dup* all contained positively selected sites (Table 1). *Lysin-dup* did not show signatures of positive selection, though this could be due to having insufficient sequences to provide the statistical power to conduct the test (Table 1) (Anisimova, et al. 2001). Clustering and distribution of amino acid sites undergoing positive selection can identify regions important to the function of rapidly evolving genes (Anisimova, et al. 2001). For example, many of the positively selected sites in lysin (11/23) are in a region of the molecule that binds its egg receptor VERL (Wilburn, et al. 2018). We investigated the distribution of sites undergoing positive selection in sp18 and sp18-dup. Similar regions of the molecule undergoing positive selection in both paralogs would suggest a shared biochemical mechanism while differences in distributions of positively selected sites would indicate divergence in biochemical mechanism.

By mapping sites under positive selection onto a protein alignment of sp18 and sp18-dup, we determined that sites under positive selection in either paralog are non-randomly distributed across the protein alignment and differentially clustered. We analyzed the selected sites in the primary sequence alignment with a parametric adaptation of the runs test (Wald-Wolfowitz test) (Figure 4A) (Magel 1997). This analysis showed that there were significant runs of sites undergoing positive selection in either paralog (p-value = 0.026), consistent with different regions evolving under positive selection among paralogs. Positive selection acting on different regions of the protein alignment is consistent with subfunctionalization of paralogs.

We investigated clustering of positively selected sites in three-dimensional space. By mapping sp18 and sp18-dup positively selected sites onto the sp18 structure, it is visually apparent that there are distinct clusters of sites under selection between paralogs (Figure 4B). Using the plane of best

fit through the crystal structure we divided the molecule into “left” and “right” sides agnostic to the location of positively selected sites. To define the “top” and “bottom” of the molecule we used a plane perpendicular to the plane of best fit. Sites under positive selection in sp18-dup were statistically more likely to be on the “right” side than on the “left” according to fisher’s exact test (p-value <0.05) (Figure 2). Sp18 sites, using the same test, showed no statistically significant difference from the null distribution. However, sites under positive selection in sp18 were enriched on the “top” of the molecule rather than the “bottom” (p-value <0.05) while the distribution of sp18-dup positively selected sites did not significantly differ (Figure 4). These two tests show that sites under selection in sp18 and sp18-dup are distributed differently across their three-dimensional structures.

We also developed a test to examine whether sites under selection in sp18 and sp18-dup were statistically more likely to be adjacent to a site under selection from the same paralog. Such a pattern of clustering would indicate a spatial relationship between positively selected sites belonging to a particular paralog. For each site under positive selection in sp18 or sp18-dup, we identified whether the closest positively selected site in three-dimensional space was significantly more likely to belong to the same paralog. We found that sites under selection in both paralogs were more likely to have the closest positively selected site belong to the same paralog rather than the other paralog according to a chi-squared test (p-value < 0.01). Together, the runs test analysis and three-dimensional analyses point to diversifying selection post-duplication of these proteins to promote functional diversification.

Because Lysin-dup was not detected to be under positive selection, we did not test for differences in sites under selection between lysin paralogs. However, we did evaluate how the lysin-dup

sequence diverged from lysin. Many of the sites shown to be undergoing positive selection in lysin differ in sequence from lysin-dup (13/14) when comparing the *H. rufescens* sequences (Figure 5B). Although this comparison is not significant ($p = 0.088$), this suggests similar sites driving the diversification in sequence of lysin between species and between lysin and its paralog lysin-dup.

2.3.3 *H. tuberculata Lysin and SP18 Duplications are Species-Specific*

Previous work described a lysin duplication unique to *H. tuberculata* (Clark, et al. 2007). The lysin paralogs were shown to be evolving under positive selection and to be maintained in the testis proteome (Clark, et al. 2007). Sites that vary between European lysin paralogs are largely located on the face of the molecules interacting with lysin receptor VERL (Figure 5A). To investigate the presence of additional sp18 and lysin paralogs in *H. tuberculata*, we constructed a long-read PacBio testis transcriptome. Performing tBLASTN searches of the *H. tuberculata* transcriptome for lysin and sp18 revealed the previously described species-specific duplication of *lysin* (*H. tuberculata* lysin copy #1 and copy #2) and a novel duplication of *sp18* (*H. tuberculata* sp18 copy #1 and copy #2).

Phylogenetic analysis indicates the *H. tuberculata* sp18 paralogs are the result of a recent duplication and not ancestral to *Haliotis*. Sequence information from other abalone species is needed to determine whether this duplication is species-specific to *H. tuberculata*; it appears to be specific to the abalone clade containing the European species. The signal sequences of the *H. tuberculata* sp18 paralogs are more similar to each other than to signal sequences from other species' sp18 paralogs. Signal sequences are not part of the mature protein and not subjected to the same evolutionary pressures driving rapid divergence, therefore these sequences show more conservation between closely related paralogs. This similarity in signal sequence between *H.*

tuberculata sp18 paralogs (15/19 sites are identical) further supports that these paralogs are the result of a non-ancestral duplication. Alternatively, they could have been subject to more recent gene conversion. Despite being a more recent duplication of sp18, the paralogs have a low sequence identity (39%), lower than that of the *H. tuberculata* lysin paralogs (83%). This is consistent with sp18 having a higher dN/dS and evolving more rapidly than lysin (Table 1). *H. tuberculata* sp18 paralogs maintain a pair of structurally important cysteine residues involved in forming a disulfide bond. Rapid sequence divergence, no premature stop codons, and both genes being expressed in the testis transcriptome are all indicators that both sp18 paralogs (referred to here as *H. tuberculata* copy #1 and copy #2) are likely to be functional.

The *H. tuberculata* sp18 copy #1 is more divergent from the ancestral sequence than copy #2, as indicated by its long branch in the sp18 phylogeny (Figure 3B). When comparing the sequence identity of *H. tuberculata* sp18 paralogs to H. rubra sp18 (an outgroup sequence), copy #1 shows a lower sequence identity (41%) than copy #2 (69%). This rapid sequence divergence of copy #1 without accruing mutations causing pseudogenization suggests strong positive selection. However pairwise dN/dS between *H. tuberculata sp18* paralogs could not be reliably estimated due to extensive divergence resulting in saturation of dS (multiple substitutions per site) (Swanson and Vacquier 1995). Maintenance of both paralogs in the testis proteome despite the observed sequence divergence would indicate that both paralogs encode non-redundant functions, presumably related to fertilization. We used data dependent acquisition mass spectrometry to search for peptides belonging to either paralog in the *H. tuberculata* testes proteome. Diagnostic peptides were detected for copy #1 but not copy #2. Despite being the more divergent sp18 sequence, copy #1 is maintained in the proteome. However, This result indicates that copy #1 is likely fulfilling an important fertilization function in H. tuberculata.

2.3.4 *Lack of Recent Duplications in Egg Coat Proteins*

We generated an ovary PacBio transcriptome for *H. tuberculata* to identify VEZP proteins. Using the 33 VEZP and ZP-domain sequences from the *H. rufescens* ovary transcriptome as the initial query sequences, exhaustive tBlastn searches of the ovary transcriptome were used to identify all cDNA sequences with sequence similarity to any *H. rufescens* VEZP. ZP module protein sequences were extracted from our *H. tuberculata* cDNA hits and the 33 *H. rufescens* VEZPs and then were aligned to construct a phylogeny (Figure 6). Clustering of *H. tuberculata* and *H. rufescens* ZP module sequences indicate that these distantly related abalone species have the same complement of ZP-proteins in their transcriptomes. In *H. tuberculata*'s ovary transcriptome, orthologs of 32 of the 33 *H. rufescens* ovary ZP-domain proteins were identified. No VEZPs, including VERL and its most closely related paralogs VEZP14 and VEZP9 were duplicated. The only missing sequence belonged to ZPC, a gene whose cDNA sequence contains a premature stop codon in *H. rufescens* and for which no peptides were detected in the *H. rufescens* VE proteome (Aagaard, et al. 2010). Therefore, ZPC is likely pseudogenized in *H. rufescens* andits expression is no longer maintained in European abalone. Remarkably, no new ZP-module-containing proteins were identified in *H. tuberculata* despite the species having multiple clade-specific duplications of acrosomal proteins. These results suggest that the clade-specific maintenance of duplicated sperm acrosomal proteins found in the European abalone *H. tuberculata* are unlikely to be the result of tit-for-tat coevolution with duplicated egg proteins.

2.4 DISCUSSION

Despite decades of research examining the evolution of abalone fertilization genes, only recently have genomic resources been available that enable a broad investigation into the evolution of the protein families to which lysin and VERL belong. Here, we explored the contributions of

duplication and sequence divergence to the evolution of abalone fertilization genes across the genus *Haliotis*. For our investigation, we generated ovary and testes transcriptomes from the European abalone *H. tuberculata* and utilized recently published North American and Australian abalone genomes and a North American abalone testes transcriptome (Palmer, McDowall et al. 2013, Nam, Kwak et al. 2017, Botwright, Zhao et al. 2019, Gan, Tan et al. 2019, Masonbrink, Purcell et al. 2019). We discovered evidence of novel duplications of both lysin and sp18 ancestral to abalone, indicating that abalone lysin and sp18 are members of an ancestral abalone protein family with four members. The newly discovered *sp18* paralog *(sp18-dup*) was shown to be undergoing positive selection, like *lysin* and *sp18,* and expressed in the testes of North American abalone. Further, differences in clustering of positively selected sites in sp18 compared to sp18-dup is consistent with models of subfunctionalization. We investigated whether there are clade-specific duplications of abalone VEZPs or acrosomal proteins. In addition to a species-specific lysin duplication described in a previous publication (Clark, et al. 2007), the *H. tuberculata* testes transcriptome contains a clade-specific duplication of sp18 not found in Australian or North American abalone species. However, no duplications of VERL or other VEZPs were observed between North American or European abalone, indicating that VEZP gene content is conserved across the genus *Haliotis*. Together, this data demonstrates that recurrent duplication and diversification driven by positive selection drives the evolution of an acrosomal protein family involved in fertilization in *Haliotis*.

2.4.1 *Recurrent Duplication and Positive Selection of Acrosomal Proteins in Abalone*

In the *H. rubra* abalone genome, the paralogs *lysin, lysin-dup, sp18,* and *sp18-dup* are found on a single scaffold. This clustering within the genome indicates that ancestral tandem duplication events occurred leading to the creation of this acrosomal protein family (Reams and Roth 2015).

Further, three of the four ancestral paralogs were shown to be maintained in the testis transcriptome and to be evolving under positive selection, a common characteristic of reproductive proteins.

This evolutionary pattern of duplication paired with sequence diversification found in the abalone acrosomal protein family can be compared to protein families in other taxa which contain sperm proteins mediating fertilization. Notably, the mammalian *Izumo* gene family contains four ancestral paralogs whose members all show testes-specific tissue expression in humans (Grayson and Civetta 2012). Izumo1 is an essential gene for sperm-egg plasma membrane fusion in mammals that functions by binding the egg plasma membrane protein JUNO (Bianchi, et al. 2014). There is evidence that the other three Izumo paralogs may also possess important, although potentially varied, functions in fertility (Ellerman, et al. 2009). All four paralogs have been shown to be undergoing rapid sequence evolution in at least one mammalian lineage, for Izumo1, 2, and 3 this is driven by positive selection and for Izumo4 this appears to be driven by relaxed selection (Grayson and Civetta 2012; Grayson 2015). Given that both the abalone acrosomal protein family and the mammalian Izumo family both contain multiple paralogs showing testis-specific function, subfunctionalization may be a common driver of the evolution of fertilization and reproductive genes across taxa. Understanding how fertilization proteins emerge and evolve can be important for identifying and understanding mechanisms of fertilization across diverse taxa.

2.4.2 *Recurrent Subfunctionalization of Abalone Acrosomal Proteins*

Differences in optimal mating rates for sperm and eggs can drive antagonistic coevolution of reproductive proteins. Under this sexual conflict scenario, evolution of egg coat proteins interacting with sperm acrosomal proteins could lead to constrained evolution on the sperm side (Gavrilets and Waxman 2002). Duplication followed by diversification of sperm fertilization

proteins can be an important means of sperm escaping evolutionary constraints imposed by egg protein evolution. For two duplication events within the abalone acrosomal protein family there is evidence of subfunctionalization from either functional experiments (lysin vs. sp18, (Swanson and Vacquier 1997; Kresge, et al. 2001) or site-clustering analysis (sp18 vs. sp18-dup, current manuscript).

Plasma membrane fusion in fertilization or other contexts is traditionally thought to consist of two steps, binding and fusion (Bianchi and Wright 2020). In sea urchins, both steps are mediated by different regions of the same protein, su-bindin (Vacquier and Moy 1977; Ulrich, et al. 1998; Vacquier and Swanson 2011). However, in abalone these steps may have been partitioned between sp18 and sp18-dup via subfunctionalization. Abalone eggs have a thin layer directly overlaying the surface of the plasma membrane which morphologically resembles a duplication of the elevated VE (Mozingo, et al. 1995). Just as lysin binds the VE protein VERL, sp18 may bind a VEZP protein found within the thin layer overlaying the abalone egg plasma membrane. Indeed, in addition, to having a strong fusagenic function, sp18 has been demonstrated to bind to VEZP proteins, an unsurprising trait for a lysin paralog (Aagaard, et al. 2010). One possibility is that the subfunctionalization of sp18 and sp18-dup may have been driven by the separation of the steps of plasma membrane binding and fusion between paralogs.

While the paralogs lysin-dup and lysin do show high sequence divergence, only lysin shows evidence of positive selection. Therefore, the observed sequence divergence is likely driven by lysin's evolution post-duplication. Unlike the other acrosomal protein family paralogs discussed in this paper, lysin-dup is not detected in the testis transcriptome. However, there is an appealing hypothesis as to its potential function. In abalone egg coats there are two VEZP proteins capable

of binding lysin: VERL the major binding partner of lysin and VEZP-14 the most recent paralog of VERL (Aagaard, et al. 2013). It is possible that lysin-dup may be the binding partner of VEZP-14 and if true this could explain why lysin shows correlated evolution with VERL but not VEZP-14 (Aagaard, et al. 2013). Currently there is insufficient data to test for correlated rates of evolution between lysin-dup and VEZP-14. However, further molecular and biochemical characterization through binding kinetic analysis could test the hypothesis that lysin-dup and VEZP-14 interact.

2.4.3 *Species-Specific Duplications of Acrosomal Proteins in Abalone*

Previous work described a lysin duplication unique to *H. tuberculata* and maintained in the testis proteome (Clark, et al. 2007). In this study a clade-specific duplication of sp18 was discovered within the *H. tuberculata* transcriptome. Despite having two acrosomal protein duplications, European abalone's ovary transcriptome did not reveal any novel VEZP protein sequences indicative of a duplication event. While gene duplications are an ongoing contributor to the evolution of sperm fertilization genes in abalone, this may not be true for egg fertilization genes. Our data suggests thatduplications on the egg side are not likely to drive the duplication of abalone acrosomal proteins in *H. tuberculata*. This could be explained by different selective pressures on the sperm and the egg, such as sperm competition and polyspermy risk (Carlisle and Swanson 2020). Further, it does not seem that a process of gene birth and loss can explain the evolution of abalone's acrosomal protein family since all paralogs are maintained in the transcriptome and have accrued no pseudogenizing mutations. A hypothesis for the duplication and diversification of acrosomal protein paralogs in *H. tuberculata* is that paralogs are specialized for different binding sites of their egg receptor or different allelic variants of their receptor. For example, *H. rufescens* VERL has 22 tandem ZP-N domains with three unique amino acid sequences, *H. tuberculata* VERL may show similar differences in ZP-N sequences

and *H. tuberculata* lysin paralogs may be optimized for binding different ZP-N sequences (Galindo, et al. 2002). In addition, the abalone *H. tuberculata* VERL may be polymorphic, as seen for *H. corrugata* VERL, and lysin paralogs are optimized for VERL allelic variants (Clark, et al. 2009). This study observed that sites that vary between *H. tuberculata* lysin paralogs are largely located on the face of the molecules interacting with lysin receptor VERL (Figure 5A). Unlike the distribution of positively selected sites between sp18 and sp18-dup where sites are differentially clustered on the protein structure. This pattern of diversification may be suggestive of specialization of function, such as interacting with different VERL allelic variants or VERL ZP-N domains. Further characterization of VERL in H. tuberculata and population-level variation is necessary to explore these hypotheses.

2.5 CONCLUSION

This study characterizes duplication events of a sperm acrosomal protein family with functions directly associated with fertilization. Although lysin was one of the first fertilization proteins discovered and the first for which an egg binding partner was defined, its evolutionary origins are unknown. By placing duplication events of lysin and sp18 within their genomic context and identifying clade-specific duplication events, our study reveals the importance of duplication for the evolution of this protein family. We describe six acrosomal protein paralogs arising from both ancestral and clade-specific duplication events (Figure 7). Recurrent duplication events of sperm acrosomal proteins have occurred throughout the evolutionary history of abalone. For the two abalone species with transcriptomic data both have paralogs maintained in the testis transcriptome. Remarkably none of these genes have been pseudogenized and many are undergoing strong positive selection consistent with maintenance of their function in abalone reproduction. Further inquiry is required to investigate why these proteins are undergoing duplication, the functional

consequences of these duplication events, and whether other fertilization proteins in other species (as also seen for the mammalian Izumo family) are undergoing recurrent duplication events.

2.6 ACKNOWLEDGMENTS

This study was supported by NIH Grant HD076862 to Willie J. Swanson and a National Science Foundation Graduate Research Fellowship to Jolie A. Carlisle. Megan Glenski was supported by University of Washington School of Medicine-Gonzaga University Regional Health Partnership. Thank you to France Haliotis for help acquiring *H. tuberculata* samples, to Evan Cox, Dr. Daniel Promislow, Dr. Josh Schraiber, and Dr. Damien Wilburn for help with data analysis, and to Alberto Rivera, Dr. Jan Aagaard, and Dr. Bryce Taylor for useful discussions and comments. All research was performed on the traditional lands of the Duwamish Tribe. To learn more about the Duwamish Tribe and their continuing legacy, please visit https://www.duwamishtribe.org/.

2.7 FIGURES AND TABLES

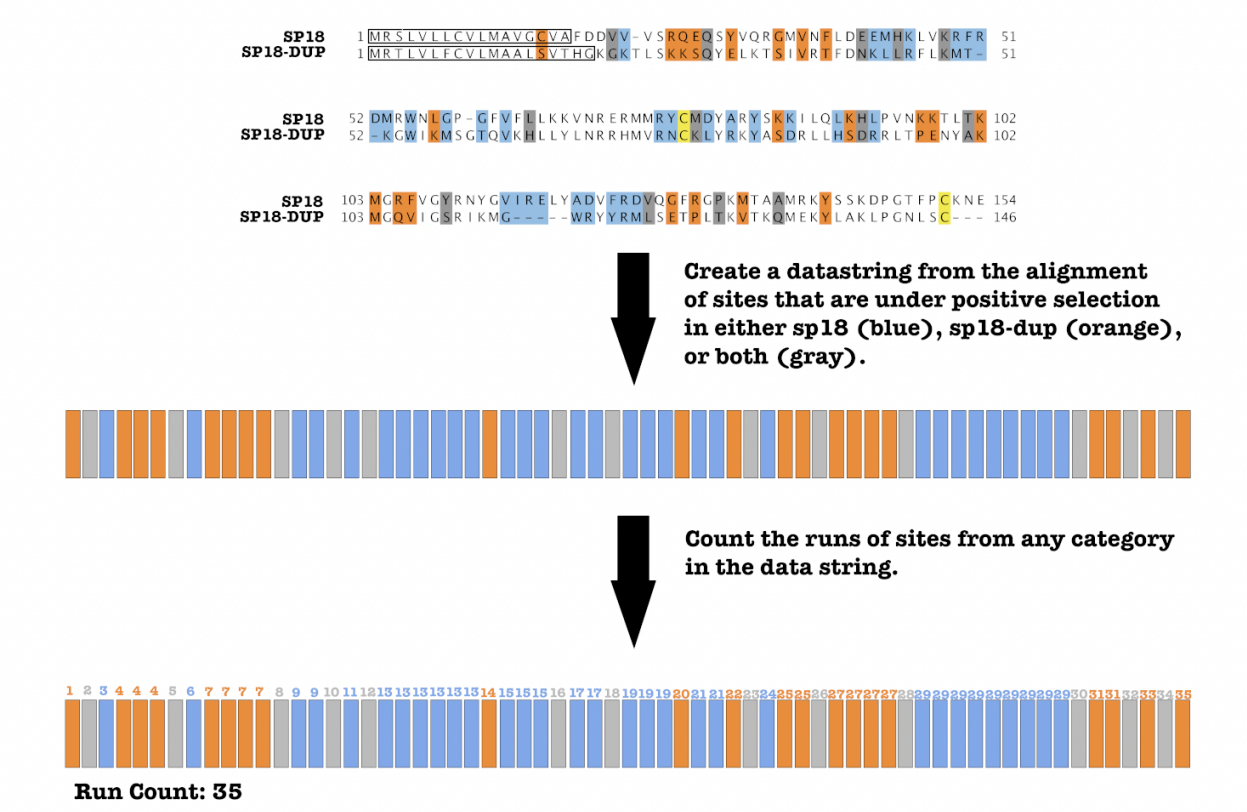


Figure 2.1. Calculations of Runs of Positively Selected Sites

The Wald-Wolfowitz runs test determines the randomness of a two-category data string by examining changes between categories by counting "runs." We designed a parametric version of the test to allow the inclusion of three categories. The categories were sites under positive selection in sp18, sites under selection in sp18-dup, and sites under selection in both. The order in which these sites under selection in the categories appeared in an alignment of *H. fulgens* sp18 and *H. sorenseni* sp18-dup became our data string. For the data string generated from our paralog alignment we counted how many times the identity of sites in the string changed plus one. A visualization of the pipeline for preparing this data string and counting "runs" is shown here. We also performed an alternate version of this analysis where the gray sites (both) were removed when calculation changes in categories.

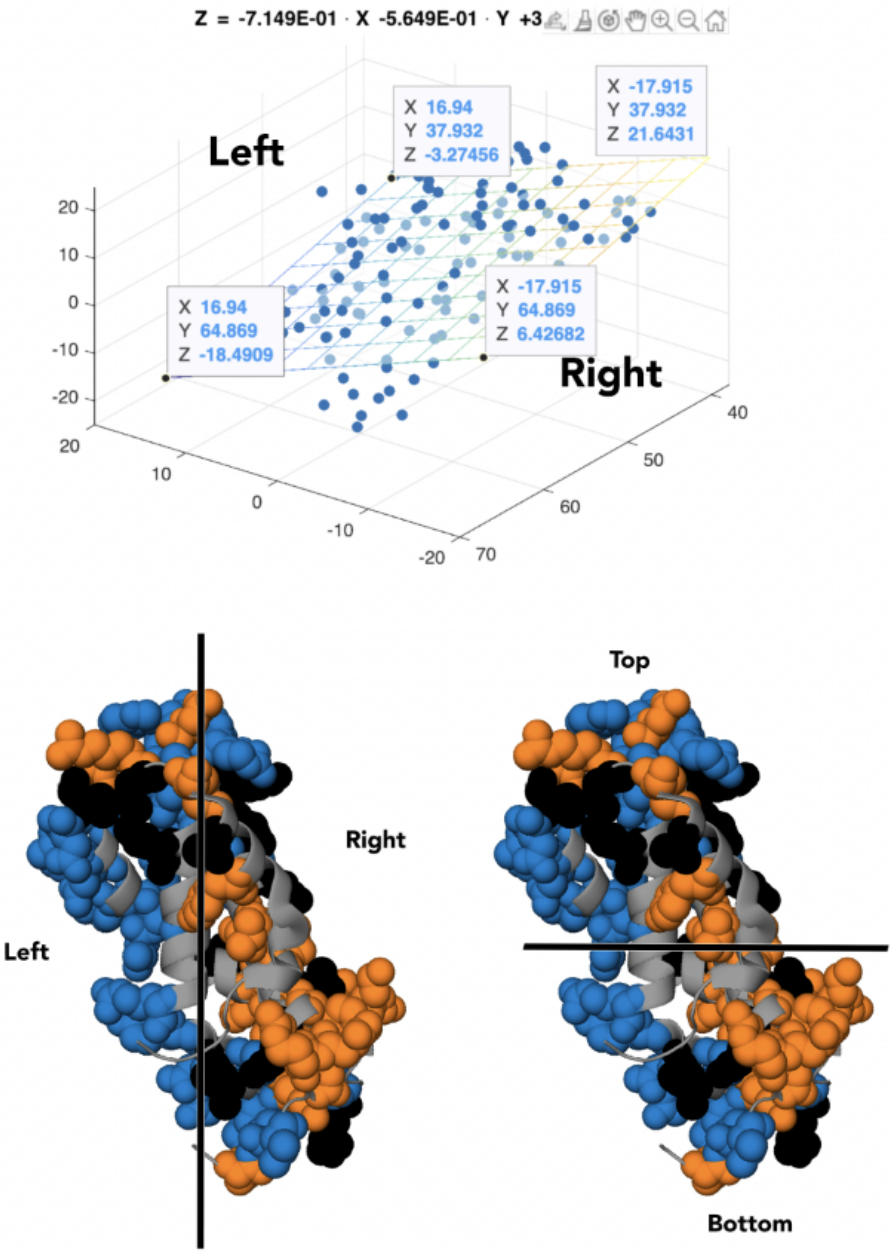


Figure 2.2. Planes separating sp18 crystal structure.

A.) Matlab was used to calculate the plane of best fit (Supplementary File 1) through the carbon alphas of the crystal structure (PDB 1GAK). **B.)** The plane of best fit divides the sp18 crystal structure into "left" and "right" sides. **C.)** A plane perpendicular to the plane of best fit divides the sp18 crystal structure into "top" and "bottom" sides.

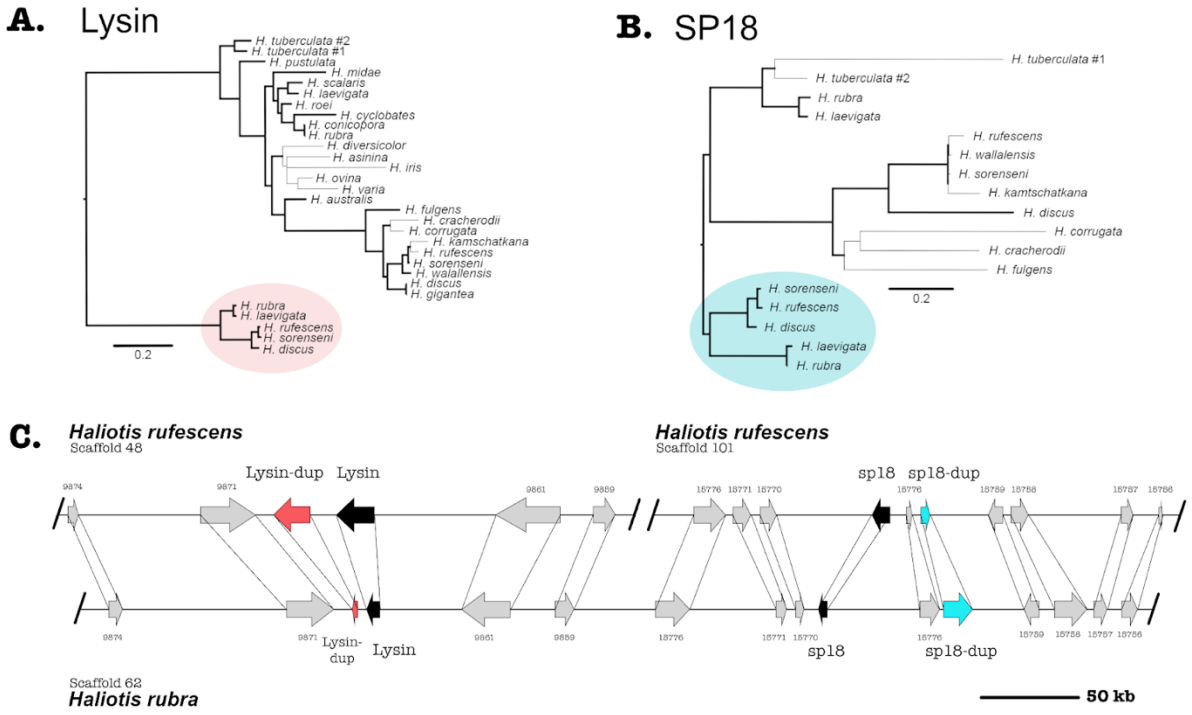


Figure 2.3. Syntenic and phylogenetic analysis indicate that four tandem acrosomal proteins are ancestral to all abalone.

A. Ancestral duplication led to the paralogs lysin and lysin-dup (Red). Bold lines indicate greater than 80% bootstrap branch support. **B.** Ancestral duplication led to the paralogs sp18 and sp18-dup (blue). A clade specific duplication of sp18 is present in the *H. tuberculata* testes transcriptome. Bold lines indicate greater than 60% bootstrap branch support. **C.** Lysin (black), sp18 (black) and their paralogs lysin-dup (red) and sp18-dup (blue) are found in the genomes of North American and Australian abalone. Syntenic analysis indicates that paralogs are located near each other in the abalone genome, although all four paralogs are only genetically linked in Australian abalone genome assemblies. In the North American abalone *H. rufescens* genome assembly, lysin and sp18 are found on separate scaffolds linked with their paralogs.

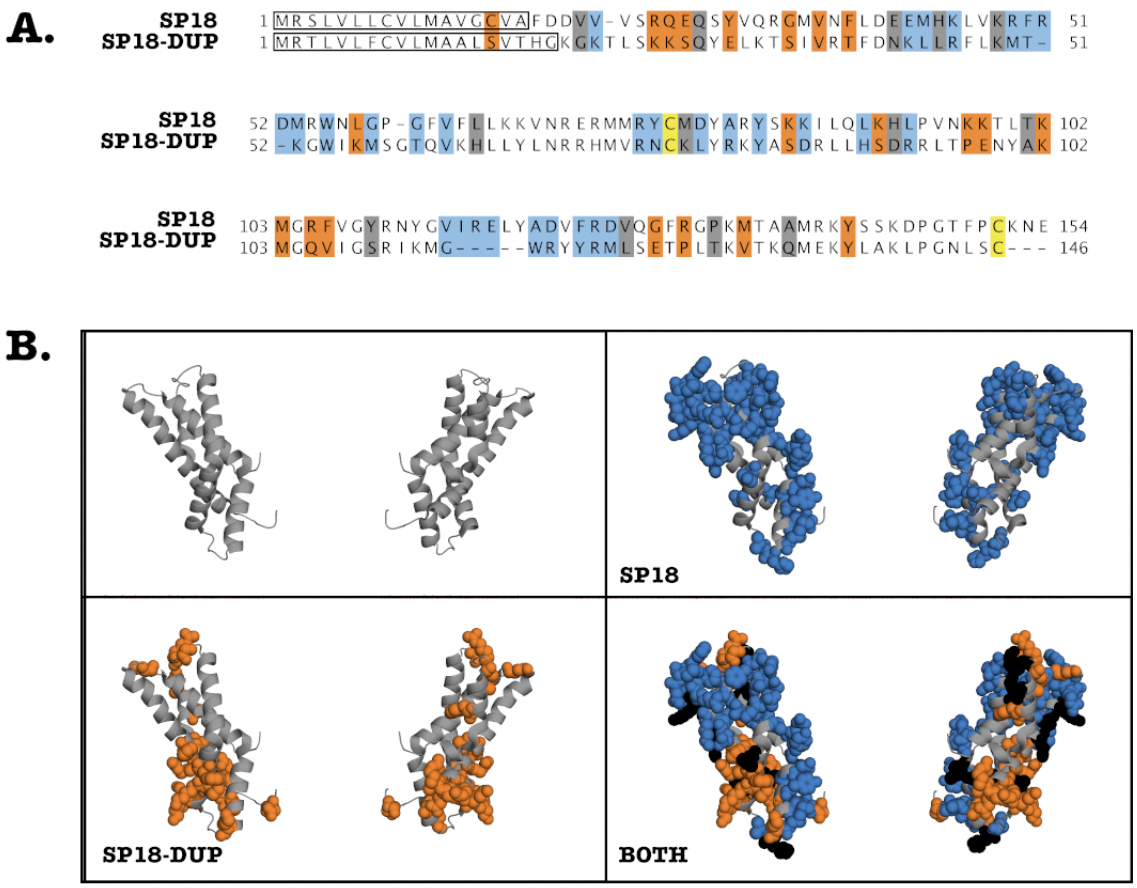


Figure 2.4. Clustering of positively selected sites is consistent with subfunctionalization of SP18 and SP18-dup.

A. Alignment of *H. fulgens* sp18 and *H. sorenseni* sp18-dup mature protein sequences. Sites under positive selection are highlighted. Blue (Sp18), Orange (Sp18-dup), and Black (both). A modified parametric runs test was used to determine that there were statistically significant runs in linear sequence space of sites under positive selection in sp18 or sp18-dup (p-value - 0.0XX). **B.** Sites under positive selection in sp18 or sp18-dup were mapped onto the crystal structure of sp18 from *H. fulgens* and appear clustered in 3-D space.

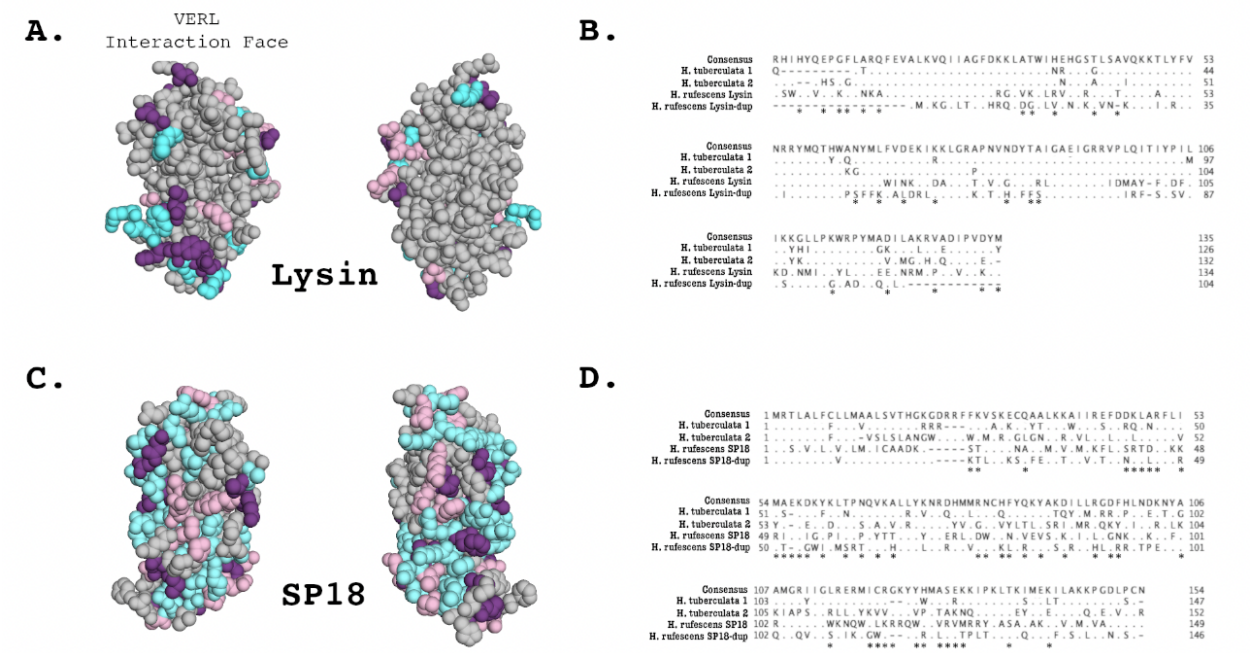


Figure 2.5. Duplication and divergence of lysin and sp18 paralogs in *H. tuberculata*.

A.) Sites that differentiate *H. tuberculata* lysin paralogs from *H. rufescens* lysin are mapped onto the *H. rufescens* NMR structure (PDB 5UTG). Blue – Sites that differ between *H. rufescens* lysin and *H. tuberculata* Copy #1, Pink – Sites that differ between *H. rufescens* lysin and *H. tuberculata* Copy #2, Purple – Sites that differ between *H. rufescens* lysin and both *H. tuberculata* lysin paralogs. Notably the face of the lysin molecule that interacts with VERL harbors most substitutions between *H. tuberculata* paralogs, suggestive of a shared interaction face for both *H. tuberculata* lysin paralogs. **B.)** Alignment of Lysin paralogs in *H. rufescens* and *H. tuberculata*. Both sets of paralogs arose from independent duplication events. Asterisks indicate sites that are shown to be under positive selection in *H. rufescens* lysin. **C.)** Sites that differ between *H. tuberculata* sp18 paralogs are mapped onto the crystal structure of sp18 *from H. fulgens* (PDB 1GAK). Blue – Sites that differ between *H. rufescens* sp18 and *H. tuberculata* Copy #1, Pink – Sites that differ between *H. rufescens* sp18 and H. tuberculata Copy #2, Purple – Sites that differ between *H. rufescens* sp18 and both H. tuberculata sp18 paralogs. **D.)** Alignment of sp18 paralogs in *H. rufescens* and *H. tuberculata*. Both sets of paralogs arose from independent duplication events. Asterisks indicate sites that are shown to be under positive selection in *H. rufescens* sp18.

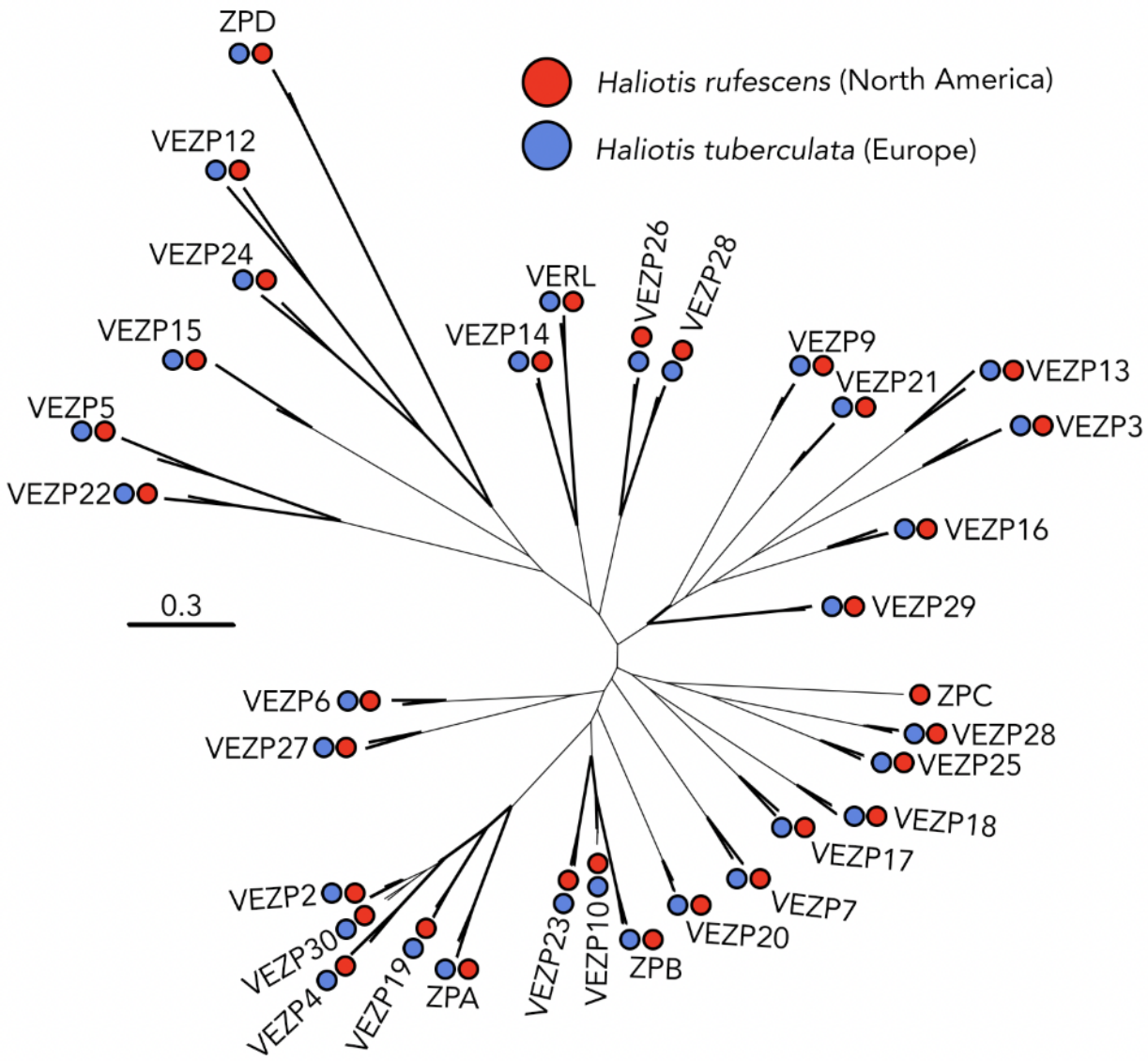


Figure 2.6. VEZP proteins are conserved across abalone species.

The distantly related abalone species *H. rufescens* (Red) and *H. tuberculata* (Blue) share the same vitelline envelope zona pellucida (VEZP) proteins within their ovary transcriptomes. Although there are species-specific duplications of lysin and sp18 *in H. tuberculata*, there is no evidence of species-specific duplications of VERL or other VEZP proteins. Bold lines indicate greater than 80% bootstrap branch support.

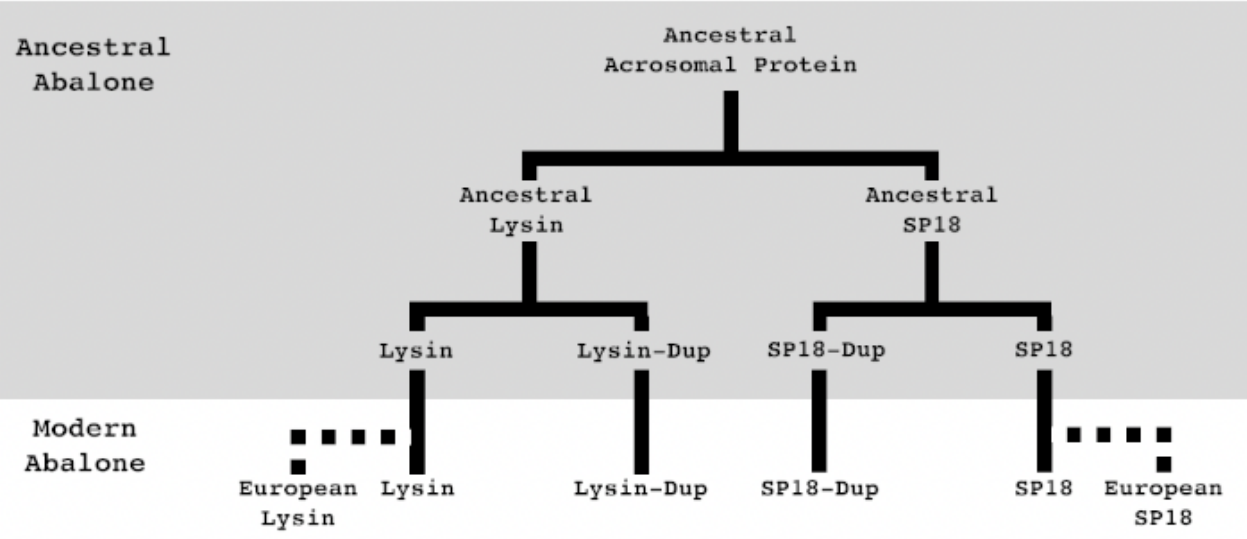


Figure 2.7. Summary of abalone acrosomal duplications.

Abalone acrosomal proteins have duplicated multiple times over the course of abalone evolution. Five duplication events can describe the currently identified protein family. Most duplications are ancestral to all abalone, but species-specific duplications in European abalone point to duplication paired with positive selection still playing an important role in the evolution of abalone fertilization pathways. Solid lines indicate orthologs present in all abalone, dotted lines indicate species-specific paralogs.

Gene	Model	-2Δl	d/d_s	% Positively Selected Sites
Sp18	M1a vs M2a	84.2**	4.4	48
	M7 vs M8	95.3**	4.2	48
	M8a v M8	84.3**	4.2	48
Sp18-dup	M1a vs M2a	4.1	-	-
	M7 vs M8	4.1	-	-
	M8a v M8	4.1*	1.8	49
Lysin-dup	M1a vs M2a	1.3	-	-
	M7 vs M8	1.5	-	-
	M8a v M8	1.3	-	-
Lysin	M1a vs M2a	156.1**	1.2	21
	M7 vs M8	156.7**	1.1	22
	M8a v M8	141.1**	1.1	22

Table 2.1. Acrosomal protein paralogs are evolving under positive selection**.**

Codon substitution models were used to analyze sequences of sp18, sp18-dup, lysin-dup, and lysin. Site models allowing for several neutral models (M1a, M7, and M8a) or selection models (M2a, M8, and M8a) allowing for variation among sites, were fit to the data using PAML. Sites undergoing positive selection were detected in sp18 and lysin for all model comparisons. A more powerful test (M8a v M8) detected positive selection in sp18-dup as well as sp18 and lysin. Estimates of the likelihood ratio statistic (**-2Δl**), d/d_s, and the percentage of sites that are under positive selection are given. Significant tests are highlighted in yellow. (*, significant at $P < 0.05$; **, significant at $P < 0.005$.)

Chapter 3. RECURRENT INDEPENDENT PSEUDOGENIZATION EVENTS OF THE SPERM FERTILIZATION GENE ZP3R IN APES AND MONKEYS

Carlisle JA, Gurbuz, D.H., Swanson, W.J. 2021. Recurrent Independent Pseudogenization Events of the Sperm Fertilization Gene ZP3r in Apes and Monkeys. BioRxiv.

3.1 INTRODUCTION

Complex molecular interactions between the sperm and egg mediate fertilization (Swanson and Vacquier 2002; Carlisle and Swanson 2020). Although recent discoveries have described many important molecules mediating mammalian sperm-egg plasma membrane fusion, the molecular mediators of sperm-egg coat interactions remain ambiguous (Carlisle and Swanson 2020). The glycoproteinaceous egg molecules ZP2 and ZP3 have been shown to bind sperm in a species-specific manner, indicating that these molecules may be involved in sperm-egg interactions (Bleil and Wassarman 1980; Litscher, et al. 2009; Avella, et al. 2014; Carlisle and Swanson 2020). While there is no known sperm protein binding partner of ZP2, ZP3r (formally known as sp56) is described as the receptor of ZP3 in mice (Buffone, et al. 2008; Wassarman 2009). ZP3r is a sperm acrosomal protein that becomes transiently exposed on the sperm head post-capacitation in mice (Muro, et al. 2012). Isolated ZP3r inhibits sperm binding by binding mouse eggs in vitro and specifically binds ZP3 as shown by photoaffinity cross-linking (Bleil and Wassarman 1990; Buffone, et al. 2008). Despite these compelling results, mouse knockouts of *ZP3r* do not result in observable reductions in fertility; however this may be due to more sensitive assays being needed to observe *ZP3r*'s function (Adham 1998; Muro, et al. 2012; Okabe 2018). For example, the sperm

protein PKDREJ, while not causing infertility in male mice knockouts, does lead to a delay in the induction of the acrosome reaction by ZP recognition and a reduction in male fertility compared to wild type animals in sequential mating trials (Sutton, et al. 2006; Sutton, et al. 2008; Miyata, et al. 2016). Multiple proteins, including ZP3r, may contribute to sperm recognition of the egg coat, and have redundant functions.

Identification of the human ortholog of *ZP3r* has been controversial. Previous studies have misidentified human *ZP3r* as either *SELENBP1* or *C4BPA*, due to nomenclature confusion or difficulties in establishing orthology respectively (Morgan, et al. 2010, 2017; Morgan and Hart 2019). *ZP3r* is found in chromosome 1 amongst paralogous protein-coding genes that make up the RCA cluster (Hourcade, et al. 1989; Krushkal, et al. 2000). Many of these genes are diverging rapidly between species (Hart, et al. 2018). This sequence divergence of the paralogs can further complicate accurate ortholog identification. In this study we demonstrate using syntenic and phylogenetic analysis that *C4BPAP1* is the primate ortholog of *ZP3r*. We examine the evolution of *ZP3r* in primates and uncover a pattern of recurrent independent pseudogenizations of *ZP3r* in Great Apes and Monkeys. This work highlights the complexities of identifying orthologs between species, particularly when pseudogenizations have occurred, and indicates that redundant mechanisms of gamete recognition may lead to the loss of reproductive genes.

3.2 Results and Discussion

3.2.1 *C4BPAP1 is the Human Ortholog of ZP3r*

Although well characterized in mice, the identification of primate *ZP3r* has been contentious. In mice, *ZP3r* is located within the RCA cluster. An examination of this genomic region in humans reveals the paralogs C4BPA and C4BPAP1. C4BPA is a large glycoprotein that acts as an inhibitor

within the complement system (Okroj M. 2018). *C4BPAP1* shares a domain structure and sequence similarity to *C4BPA* but contains a premature stop codon in Exon 2 that is fixed in humans and suggests pseudogenization. Human *C4BPA* and *C4BPAP1* are composed of 11 exons which contain 8 CCP/sushi domains and a C-terminal transmembrane domain (Hofmeyer, et al. 2013). Both C4BPAP1 and C4BPA contain an additional CCP domain to mouse ZP3r which is missing CCP domain 7. A recent investigation identified *C4BPA* as the human ortholog of *ZP3r*, perhaps overlooking *C4BPAP1* since it is pseudogenized in humans (Morgan and Hart 2019). The study hypothesized that a duplication in rodents of *C4BP*, the rodent ortholog of human *C4BPA*, led to the evolution of rodent *ZP3r* (Morgan and Hart 2019). However, human *C4BPA* is known to function in immunity and its highest tissue expression is in the liver, inconsistent with a function as the sperm fertilization gene *ZP3r* (Carithers and Moore 2015; Okroj M. 2018). Meanwhile, although pseudogenized, *C4BPAP1* shows highest RNA expression in the testis, consistent with an ancestral function as a sperm fertilization gene (Carithers and Moore 2015). Genes that have been recently pseudogenized are often still expressed until their regulatory regions also degenerate by further mutation (Bekpen, et al. 2009).

Using phylogenetic analysis and syntenic mapping, we showed that *C4BPAP1* is the human ortholog of mouse *ZP3r* (Figure 1). A protein alignment of *Homo sapiens* and *Macaca mulatta C4BPAP1* and *C4BPA* and *Mus musculus* and *Rattus norvegicus ZP3r* and *C4BP* sequences were used to construct a maximum likelihood phylogeny. Primate *C4BPAP1* and Rodent *ZP3r* clustered distinctly from Primate *C4BPA* and Rodent *C4BP*, indicating that Primate *ZP3r* (*C4BPAP1*), not *C4BPA*, is the ortholog of rodent *ZP3r* (Figure 1B). Further, we used the best reciprocal blast hits of ZP3r/C4BPAP1 (stop codon removed), C4BPA, and neighboring RCA cluster gene transcripts between the mouse and human genome to establish syntenic relationships. Syntenic comparisons

between the region of the human and mouse RCA clusters containing *C4BPAP1* and *ZP3r* respectively, support *C4BPAP1* as the human *ZP3r* ortholog. In humans, *C4BPAP1* is located between *C4BPA* and *CD55*, as is *ZP3r* in rodents (Figure 1A) (Kent, et al. 2002).

Previous research identified elevated linkage disequilibrium between the region of the human genome containing *ZP3r* and the region containing *ZP3* suggestive of coevolution between these loci (Rohlfs, et al. 2010). Since *ZP3r* is pseudogenized in humans, this was possibly a false positive result or a complex association. An alternative hypothesis would be that the human sperm receptor for *ZP3* is located nearby the pseudogenized human *ZP3r*. However, none of the annotated genes within the region shown to be in LD with *ZP3* show testes-specific expression (Carithers and Moore 2015).

3.2.2 *ZP3r Has Been Repeatedly and Rapidly Pseudogenized in Apes*

C4BPA and *ZP3r/C4BPAP1* are members of the RCA cluster, the genes in this locus are largely conserved across even distantly related species, with sequence variation between species being driven by positive selection, indels, and intragenic domain duplications and losses (Sanchez-Corral, et al. 1993; Heinen, et al. 2006; Wu, et al. 2012; Garcia-Fernandez, et al. 2021). However, some variation in RCA cluster gene content driven by clade-specific duplication or loss events have also been observed (Sanchez-Corral, et al. 1993; Pardo-Manuel de Villena 1995; Wu, et al. 2012). Notably, *C4BPB* is pseudogenized in mice and there is evidence of two additional pseudogenized duplications of *C4BPA* found in humans (*C4BPAP2* and *C4BPAP3*) (Pardo-Manuel de Villena 1995; Kent, et al. 2002). However, primate *ZP3r/C4BPAP1* is unique in independently acquiring pseudogenization events in most apes and several monkey species (Figure 2). Although there are examples in the literature of repeated pseudogenization events of genes

across species, it is rare for independent events to occur within a closely related clade (Bainova, et al. 2014; Velova, et al. 2018). Remarkably, since the common ancestor of all apes (~16-20 mya), at least four unique pseudogenization events of *ZP3r* have occurred (Figure 2) (Chatterjee, et al. 2009).

Parsimony analysis of the pseudogenized *ZP3r* sequences indicate 9 independent pseudogenization events have occurred in primates. Remarkably, many of these pseudogenizing mutations occurred independently in closely related species and are located in distinct codons (Figure 3). With the exception of Orangutan (*Pongo abelli*), *C4BPAP1* has been pseudogenized in all apes (Human, Chimpanzee, Bonobo, Gorilla, Gibbons) (Figure 2). This rapid, repeated pseudogenization appears to be an extreme example of gene loss in apes. Gorillas, Humans, and Gibbons all have premature stop codons within CCP domain 2, all in different codons (Figure 3). Orangutan's *ZP3r* does not have any pseudogenizing mutations, however, its second CCP domain is missing a conserved and potentially structurally important cysteine that is likely to disrupt the overall structure of the protein. In 10 New World monkey (NWM), 13 Old World monkey (OWM), and one Tarsier genome assemblies, we identified the full *ZP3r* locus. Out of the 10 NWM genomes examined, 4 contained pseudogenizing mutations unique to that NWM species (Figure 2). In OWMs, only one species, *Colobus angoloensis*, had a pseudogenizing mutation within *ZP3r* (Figure 2).

Recurrent, lineage-specific gene loss events between closely related species is suggestive of strong selection for gene loss. There is no obvious correlation between ZP3 sequence and glycosylation state and ZP3r loss in primates, therefore, it is still unclear what is driving the loss of ZP3r in primates. Phylogenetic analysis of all individual CCP domains found in human C4BPA and

C4BPAP1 and mouse C4BP and ZP3r indicate no evidence of concerted evolution between or within genes that could explain the repeated pseudogenization events (Figure 4). Further, a search of the human genome reveals no new duplications of ZP3r that could be fulfilling its receptor function. However, a more distantly related paralog with low sequence similarity could be performing ZP3r's function. Protein structure changes more slowly than protein sequence, therefore a paralog with low sequence identity may still retain similar function.

3.2.3 *ZP3r Evolves Under Positive Selection*

A recurrent feature of gamete recognition proteins are signatures of positive selection, potentially created through sexual selection or sexual conflict between the sperm and the egg (Carlisle and Swanson 2020). Genes mediating immune system functions are also frequently undergoing positive selection due to host-pathogen interactions driving arms race dynamics (Lazzaro 2012). So, it is unsurprising that both *ZP3* and *C4BPA* have both been shown in previous studies to be undergoing positive selection in rodents and primates (Swanson, et al. 2001; Swann, et al. 2007; Rohlfs, et al. 2010; Swann, et al. 2017; Hart, et al. 2018; Morgan and Hart 2019). In this study, we estimated values of dN/dS for *ZP3r, C4BPA,* and *ZP3* in rodents and primates using the codeml program of PAML 4.8 (Yang 1997, 2007). For primate *ZP3r*, we only analyzed full coding sequences, no pseudogenized primate sequences were included. We compared models of selection using a likelihood ratio test (LRT) between neutral models and models with positive selection. Specifically, we compared M1 v. M2, M7 v. M8, and M8a v. M8 (Yang, et al. 2000; Swanson, et al. 2003).

We detected positively selected sites in *ZP3r, C4BP*, and *ZP3* in rodents and *ZP3r* and *C4BPA* in primates, using the M8a v M8 comparison (Table 1). Signatures of positive selection in rodent and

primate *ZP3r* is suggestive of functionally important genetic innovation being selected for within both clades. Because interacting reproductive proteins must co-evolve to maintain reproductive compatibility, ZP3r's rapid evolution could be driven by the evolution of its putative binding partner ZP3. Although positively selected sites were not detected in primate ZP3 in this study, previous population genetic analysis has detected selection on ZP3 in humans (Rohlfs, et al. 2010; Hart, et al. 2018). Since ZP3r is undergoing positive selection in primates, ZP3r's repeated and independent pseudogenization in primates is likely not driven by relaxed selection.

3.3 Materials and Methods

3.3.1 *Construction of Phylogenetic Trees*

A protein alignment of C4BPAP1 and C4BPA from *Homo sapiens* and *Macaca mulatta* and ZP3r and C4BP from *Mus musculus* and *Rattus norvegicus* was constructed using Clustal Omega (Sievers and Higgins 2014). This protein alignment was used to construct a maximum likelihood phylogeny with bootstrapping. The phylogenetic inference tool RAxML-NG was used to construct the phylogenetic tree with the LG substitution matrix (Kozlov, et al. 2019). RaxML-NG conducts maximum likelihood based phylogenetic inference and provides branch support using non-parametric bootstrapping. The best scoring topology of 20 starting trees (10 random and 10 parsimony-based) was chosen. RaxML-NG was used to perform non-parametric bootstrapping with 1000 re-samplings that were used to re-infer a tree for each bootstrap replicate MSA.

3.3.2 *Construction of CCP Domain Phylogeny*

The primate CCP domain phylogeny was constructed with a similar method as described in the previous section. Full-length protein alignments of primate C4BPA and ZP3r sequences were

aligned with Clustal Omega (Sievers and Higgins 2014). Individual CCP domains were extracted from this alignment and realigned with Clustal Omega. RaxML-NG was used to construct a phylogeny as described above (Kozlov, et al. 2019).

3.3.3 *Detection of Positive Selection*

We estimated dN/dS values for primate and rodent ZP3r, C4BPA/C4BP, and ZP3 using the codeml program of PAML 4.8 (Swanson, et al. 2003; Yang 2007). We compared models of selection using a likelihood ratio test (LRT) between neutral models and models with positive selection. The model comparisons were made between M1 v. M2, M7 v. M8, and M8a v. M8. The likelihood ratio (LRT) statistic was calculated as twice the negative difference in likelihoods between nested models. For M1a v M2a or M7 v Model 8 the LRT was compared to the $\chi 2$ distribution with 2 degrees of freedom (Yang 2007). Twice the negative difference in likelihoods between the models M8a v. M8 comparison was compared. The LRT statistic was approximated by the 50-50 mixture distribution of 0 and x2 with one degree of freedom (Swanson, et al. 2003).

3.3.4 *Syntenic Comparison*

Using the annotated *Mus musculus* and *Homo sapiens* genomes available on the UCSC genome browser, syntenic comparison of the genomic regions containing ZP3r and C4BPA was performed (Kent, et al. 2002; Mouse Genome Sequencing, et al. 2002; Miga, et al. 2014). Annotated genes and pseudogenes in the human genome were examined for their tissue-specific expression patterns. A reciprocal blast of the regions identified as ZP3r in mice or humans was performed to verify orthology (McGinnis and Madden 2004).

3.4 CONCLUSION

Despite their functional importance, the molecular mediators of fertilization have been poorly described in mammals, particularly for identifying sperm proteins mediating egg coat recognition (Carlisle and Swanson 2020). Difficulty in finding sperm receptors to egg coat proteins may be driven by functional redundancy causing many fertilization genes to be nonessential contributors to gamete recognition. Typically, protein functional redundancy refers to paralogous proteins that are structurally similar, that maintain the same interaction partners, and whose loss can be compensated for by their paralog. However, proteins can also be functionally redundant without being paralogous or structurally similar. For example, the acrosomal sperm proteins Zona Pellucida Binding Protein (ZPBP/sp38) and acrosin are structurally unrelated proteins that competitively interact with the ZP in boars (Mori, et al. 1993; Lin, et al. 2007). Functional redundancy of genes mediating fertilization could lead to clade-specific gene loss events or changes in relative functional importance between species. Again, reflecting on acrosin, knockouts of acrosin in mice (*Mus musculus*) result in infertility; yet, in hamsters (*Mesocricetus auratus*) acrosin is essential for zona penetration (Baba, et al. 1994; Hirose, et al. 2020). Together, these results indicate that functional redundancy between ZPBP, acrosin, and potentially other unknown sperm proteins, allow the relative importance of acrosin to sperm bypassing the ZP to vary between species.

In this study, we demonstrate that the testes-expressed pseudogene C4BPAP1 is the human ortholog of rodent ZP3r using phylogenetic and syntenic analysis. While ZP3r is associated with ZP binding in mice, ZP3r shows repeated pseudogenization in primates (at least 9 times), most notably in apes. Recurrent independent pseudogenizations of a rapidly evolving protein are rarely discussed in the literature, and their existence is surprising. While usually rapid divergence is

focused on sequence diversification, changes in gene content caused by gene gains and loss events could also be a significant contributor to molecular diversity and tolerated due to functional redundancy (Carlisle 2021). ZP3r is a nonessential fertilization gene in mice, who may be one of many proteins interacting with ZP3 (Muro, et al. 2012; Miyata, et al. 2016; Okabe 2018). ZP3r's repeated loss in many primates, particularly apes, despite being subject to positive selection in other primate species, indicates that the relevant importance of ZP3r to fertilization differs across primates. This difference could be due to the emergence or increase in relative importance of a different fertilization gene mediating ZP3 binding in primates. Differences in relative functional importance between clades may also partially explain why reproductive proteins are rapidly evolving in some clades and not others (Carlisle and Swanson 2020). This study highlights the potential variability of molecular mechanisms of fertilization even within mammals and emphasizes the value of using diverse model systems for investigating mechanisms of fertilization.

3.5 ACKNOWLEDGMENTS

This study was supported by NIH Grant HD076862 to Willie J. Swanson and a National Science Foundation Graduate Research Fellowship to Jolie A. Carlisle. Thank you to Dr. Damien Wilburn, Dr. Jan Aagaard, and Alberto Rivera for useful feedback. All research was performed on the traditional lands of the Duwamish Tribe. To learn more about the Duwamish Tribe and their continuing legacy, please visit https://www.duwamishtribe.org/.

3.6 FIGURES AND TABLES

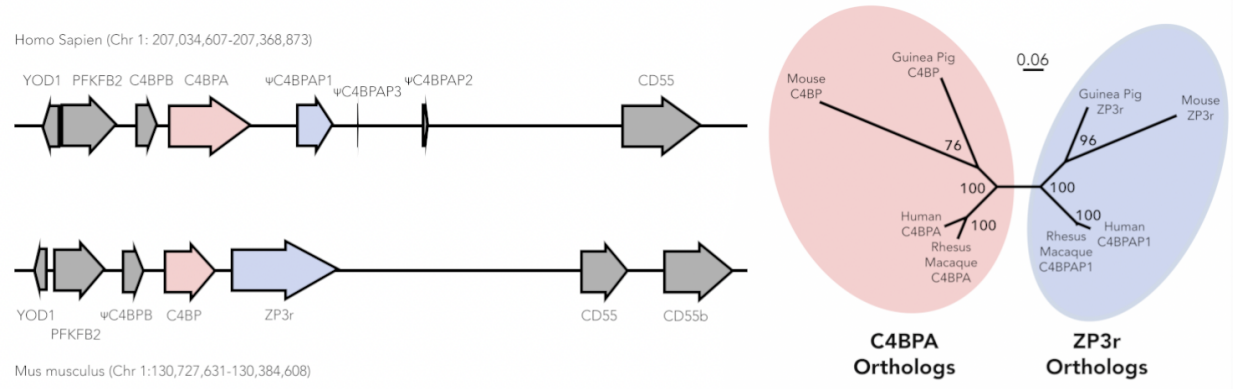


Figure 3.1. Syntenic and phylogenetic analysis indicates that C4BPAP1 is the human ortholog of mouse ZP3r.

A.) Syntenic comparison of the genomic region between Mus musculus and Homo sapiens reveals that C4BPAP1 is syntenic to mouse ZP3r. B) A protein alignment of C4BPAP1 and C4BPA from *Homo sapiens* and *Macaca mulatta* and ZP3r and C4BP from *Mus musculus* and *Rattus norvegicus* was constructed using Clustal Omega. This protein alignment was used to construct a maximum likelihood phylogeny with bootstrapping. In the phylogeny, Primate C4BPAP1 and Rodent ZP3r cluster separately from Primate C4BPA and Rodent C4BP, indicating that Primate ZP3r (C4BPAP1), not C4BPA, is the ortholog of rodent ZP3r. The phylogenetic inference tool RAxML-NG was used to construct the phylogenetic tree with the LG substitution matrix. RaxML-NG conducts maximum likelihood based phylogenetic inference and provides branch support using non-parametric bootstrapping. The best scoring topology of 20 starting trees (10 random and 10 parsimony-based) was chosen. RaxML-NG was used to perform non-parametric bootstrapping with 1000 re-samplings that were used to re-infer a tree for each bootstrap replicate MSA.

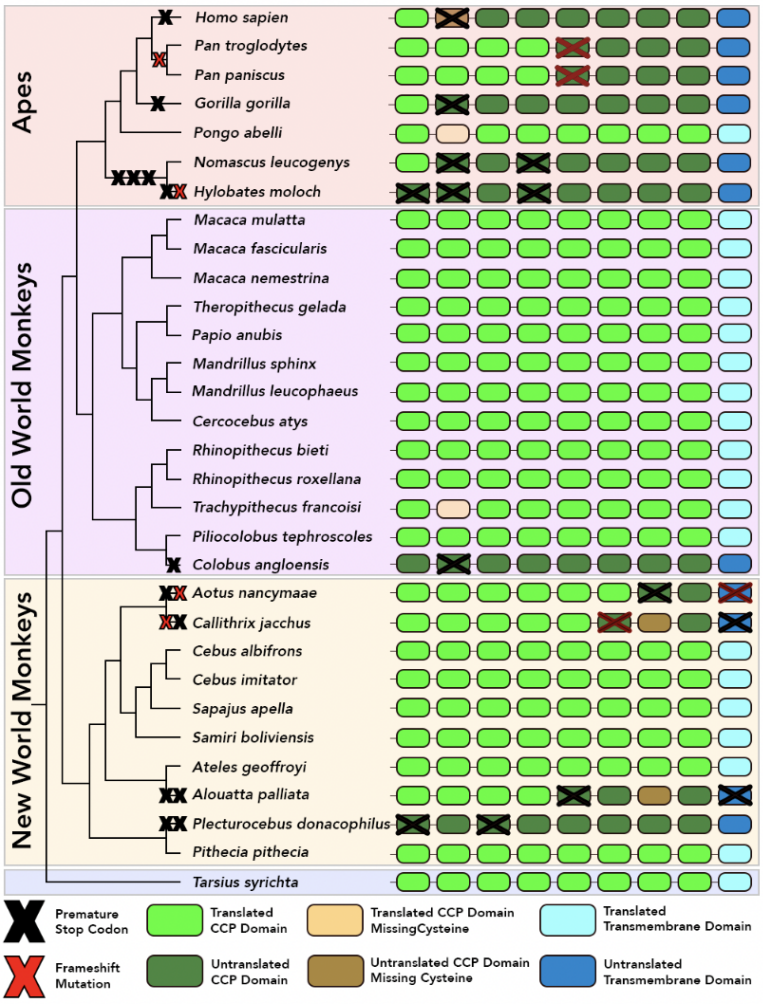


Figure 3.2. Recurrent and independent pseudogenization events of ZP3r in apes and monkeys.

We predicted the exonic sequences of ZP3r from primate genomes using the Protein2Genome command of the program Exonerate version 2.2.0 (Slater & Birney, 2005). The top scoring prediction from Exonerate was used to define the paralog's exons. We used human C4BPAP1, with the stop codon in CCP2 removed, as the protein query in Exonerate. Using HMMER, we identified CCP domains in the ZP3r sequences. On the right side is a not-to-scale cartoon of the CCP domains (Green ovals) and C-terminal transmembrane domain (Blue ovals) found in ZP3r in each species. The loss of a CCP domain caused by a missing structurally important cysteine are shown as Yellow ovals. Red crosses indicate an insertion/deletion mutation causing a frameshift mutation; Black crosses indicate a premature stop codon causing pseudogenization. Darker colored CCP domains indicate regions of ZP3r that would not be translated due to a pseudogenization event. Within some CCP domains, multiple mutations have occurred. See Figure 3 for a protein alignment of Ape ZP3r with pseudogenizing mutations marked. The cladogram on the left has pseudogenization events marked. For branches with multiple mutations the order in which these mutations occurred are unknown.

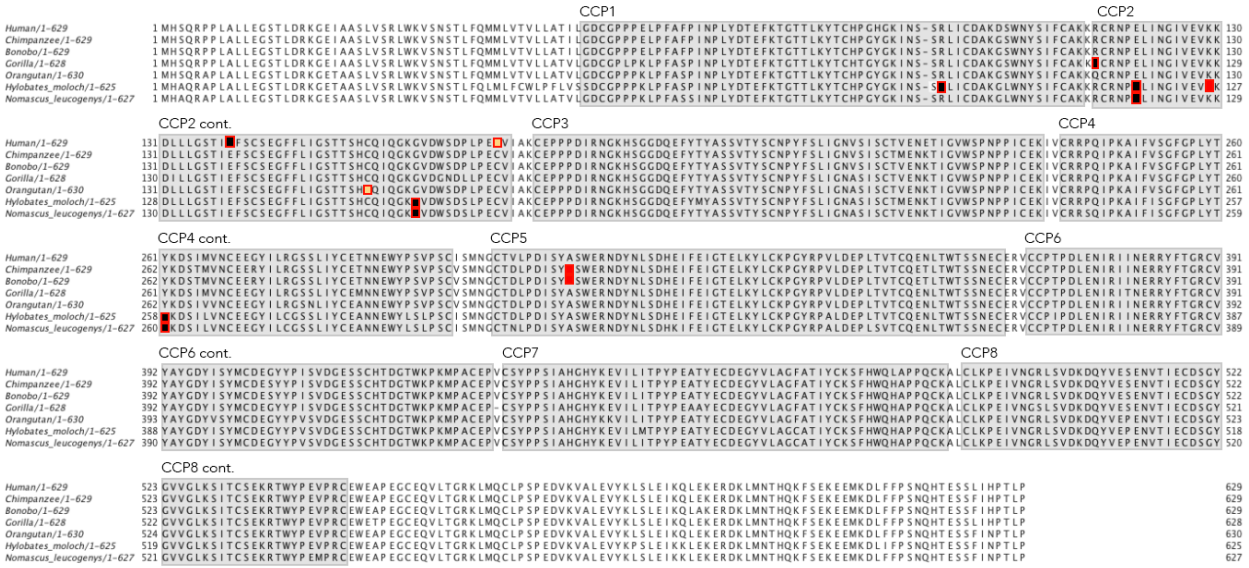


Figure 3.3. Protein Alignment of C4BPAP1/ZP3r sequences in Apes

Protein Alignment of C4BPAP1/ZP3r sequences in Apes. ZP3r in all apes, with the exception of Orangutans, have at least one mutation that results in premature termination of translation. Black squares indicate a premature stop codon; Red squares indicate an insertion/deletion mutation that results in premature termination of translation. Yellow squares indicate a cysteine that is structurally important to the CCP domain has been lost.

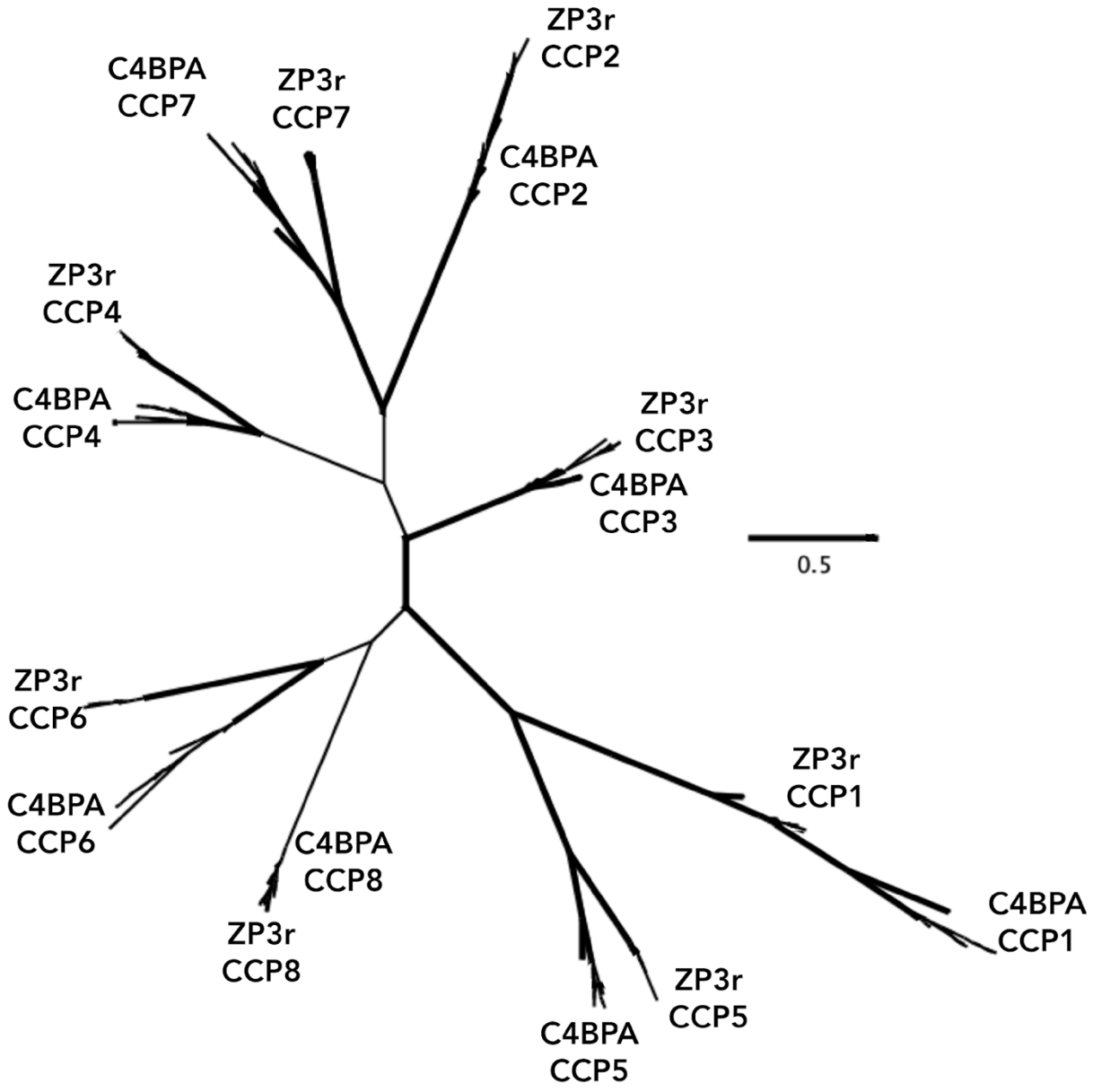


Figure 3.4. Protein phylogeny of primate C4BPA and ZP3r CCP domains reveals no evidence of concerted evolution.

Bootstrap support greater than 70 is shown by a bold line. The separation of CCP domains between either C4BPA and ZP3r can be difficult to see in this image. See the attached newick tree file (Supplementary File 1) to more closely examine the phylogenetic relationships between protein domains.

Gene	Clade	Model	-2Δl	dN/dS	% Positively Selected Sites
ZP3	Rodents	M8 v M8a	**14.63	3.66585	1.6
ZP3	Primates	M8 v M8a	1.75		
ZP3r	Rodents	M8 v M8a	**84.56	3.12174	7.5
ZP3r	Primates	M8 v M8a	*3.89	3.01771	3.4
C4BP	Rodents	M8 v M8a	**100.18	3.08952	9.7
C4BPA	Primates	M8 v M8a	**86.11	3.62581	12.6

Table 2.1. ZP3r and C4BPA contain positively selected sites in rodents and primates.

Codon substitution models were used to analyze sequences of ZP3, ZP3r/C4BPAP1, C4BP/C4BPA in rodents and primates. Site models allowing for several neutral models (M1a, M7, and M8a) or selection models (M2a, M8, and M8a) allowing for variation among sites, were fit to the data using PAML. In this table are the results from M8a v M8 comparison, for the results from other model comparisons see Supplementary Table 1. In rodents, sites under positive selection were detected in ZP3, ZP3r, and C4BP. In primates, M8a v M8 model comparison indicated sites under positive selection were detected in ZP3r and C4BPA, but not ZP3. Estimates of the likelihood ratio statistic (**-2Δl**), d_N/d_S, and the percentage of sites that are under positive selection are given. (*, significant at $P < 0.05$; **, significant at $P < 0.005$.)

Gene	Clade	Model	-2Δl	dN/dS	% Positively Selected Sites
ZP3	Rodents	M1a v M2a	**11.16	5.41816	1.1
ZP3	Rodents	M7 v M8	**17.61	3.66585	1.6
ZP3	Rodents	M8 v M8a	**14.63	3.66585	1.6
ZP3	Primates	M1a v M2a	1.65		
ZP3	Primates	M7 v M8	*9.97	1.32	14.4
ZP3	Primates	M8 v M8a	1.75		
ZP3r	Rodents	M1a v M2a	**101.47	4.19662	5.3
ZP3r	Rodents	M7 v M8	**109.43	3.12174	7.5
ZP3r	Rodents	M8 v M8a	**84.56	3.12174	7.5
ZP3r	Primates	M1a v M2a	4.38		
ZP3r	Primates	M7 v M8	*6.32	3.01771	3.4
ZP3r	Primates	M8 v M8a	*3.89	3.01771	3.4
C4BP	Rodents	M1a v M2a	**107.47	3.56697	8.1
C4BP	Rodents	M7 v M8	**115.69	3.08952	9.7
C4BP	Rodents	M8 v M8a	**100.18	3.08952	9.7
C4BPA	Primates	M1a v M2a	**86.48	3.72082	11.7
C4BPA	Primates	M7 v M8	**89.92	3.62581	12.6
C4BPA	Primates	M8 v M8a	**86.11	3.62581	12.6

Table 2.2. ZP3r, C4BPA, and ZP3 all contain positively selected sites.

Codon substitution models were used to analyze sequences of ZP3, ZP3r/C4BPAP1, C4BP/C4BPA in rodents and primates. Site models allowing for several neutral models (M1a, M7, and M8a) or selection models (M2a, M8, and M8a) allowing for variation among sites, were fit to the data using PAML. In this table are the results from M8a v M8 comparison, for the results from other model comparisons see Supplementary Table 1. In rodents, sites under positive selection were detected in ZP3, ZP3r, and C4BP. In primates, M8a v M8 model comparison indicated sites under positive selection were detected in ZP3r and C4BPA, but not ZP3. Estimates of the likelihood ratio statistic (**-2Δl**), d_N/d_S, and the percentage of sites that are under positive selection are given. (*, significant at $P < 0.05$; **, significant at $P < 0.005$.)

Chapter 4. BIOCHEMICAL DISSECTION OF SPECIES-SPECIFIC VITELLINE ENVELOPE DISSOLUTION IN ABALONE

Carlisle, J.A. Wilburn, D.B. Aagaard, J. Soelberg, S.D. Klevit, R.E. Swanson, W.J.

4.1 INTRODUCTION

A recurrent trait of proteins mediating gamete recognition is rapid sequence divergence (Swanson and Vacquier 1995, 2002; Swanson, et al. 2003; Carlisle and Swanson 2020). This rapid sequence diversification can have profound effects on protein function across species (Swanson and Vacquier 1997; Kamei and Glabe 2003; Avella, et al. 2014; Bianchi and Wright 2015). Sequence divergence of fertilization proteins between species can create boundaries to hybridization through species-specific gamete recognition function (Vacquier and Swanson 2011). Understanding the repercussions of rapid sequence divergence on the function of reproductive proteins is important to understand why reproductive proteins are rapidly evolving and a deeper understanding of the biochemical mechanisms mediating fertilization.

Despite being an essential process for sexually reproducing organisms, the proteins mediating fertilization are not well described in any animal species, including humans (Carlisle and Swanson 2020). Fertilization occurs as a cascade of sperm-egg molecular interactions. The sperm acrosomal reaction is induced after the sperm and egg zona pellucida come into contact. Released from the sperm acrosome are proteins that interact with zona pellucida molecules to mediate egg coat dissolution. After the sperm passes the egg coat, sperm and egg plasma membrane fusion occurs. Although in mammals, ZP2 and ZP3 have been identified as zona

pellucida proteins that can bind sperm in a species-specific manner, the identity of their human sperm binding partners are not known (Bleil and Wassarman 1980; Avella, et al. 2014). Animal models can be a powerful tool to gain insights into fertilization mechanisms. Despite the rapid diversification of fertilization proteins and absence of clear fertilization protein orthologs across distant taxa, there is evidence that the underlying biochemical mechanisms mediating fertilization may be conserved (Mozingo, et al. 1995; Swanson, et al. 2011).

The marine mollusc abalone (genus *Haliotis*) is a classic model for studying fertilization. Abalone eggs have an elevated glycoproteineaceous layer analogous to the mammalian zona pellucida (ZP) (Mozingo, et al. 1995). This raised extracellular layer, the Vitelline Envelope (VE), is morphologically similar to the mammalian ZP, as well as biochemically similar, with both isolated structures dissolving under similar conditions (Mozingo, et al. 1995). Both mammalian ZP and abalone VE are composed of glycoproteins with C-terminal ZP modules. A ZP module is composed of two immunoglobulin-like domains, ZP-N and ZP-C (Wilburn and Swanson 2016). The ZP module is important for contributing to the formation of a strong egg coat by contributing to protein polymerization (Killingbeck and Swanson 2018).

Many species have been shown to have egg coat ZP proteins containing N-terminal ZP-N domains, sometimes in tandem ZP-N arrays. In abalone, the VE protein VERL (VItelline Envelope Receptor for Lysin) is composed of an N-terminal tandem array of 23 ZP-N domains followed by a ZP module (Galindo, et al. 2002). In mammals, egg coat ZP2 has a tandem array of 3 ZP-N domains followed by a ZP module (Wilburn and Swanson 2016). Both ZP2 and VERL have been shown to have species-specific gamete recognition functions. In abalone, lysin

shows species-specific vitelline envelope dissolution function, and species-specific binding to VERL and the second ZP-N domain (Swanson and Vacquier 1997; Raj, et al. 2017). In transgenic mice where the N-terminal ZP-N domain of ZP2 is replaced with the orthologous human domain, mouse eggs gain an ability to bind human sperm, indicating that this N-terminal domain drives species-specific recognition (Avella, et al. 2014).

The similarities between abalone VERL and mammalian ZP2, as well as the similarities of their egg coats, point to similar biochemical mechanisms of fertilization in mammals and abalone. Since the sperm mediator of mammalian ZP dissolution is unknown, studying the interactions between abalone lysin and VERL ZP-Ns can provide valuable insights into conserved mechanisms of fertilization and how sequence divergence impacts species-specific gamete interactions.

In abalone, lysin and VERL have both been shown to be evolving under positive selection (Lee, et al. 1995; Galindo, et al. 2003). VERL in *H. rufescens* is composed of 23 tandem ZP-N domains and a C-terminal ZP module. The first two VERL ZP-N domains (VR1 and VR2) have unique amino acid sequences, but VERL repeats 3 (VR3) through 23 have near identical sequences due to concerted evolution (Swanson and Vacquier 1998). Lysin must bind to each VERL ZP-N domain in order to mediate VE dissolution. VERL's first two N-terminal ZP-N domains (VR1 and VR2) contain positively selected sites while VR3-23 is neutrally evolving. VERL VR1 contains four positively selected sites within the lysin binding interface; lysin also is evolving under positive selection with a cluster of positively selected sites clustered at the ZP-N domain binding interface (Wilburn, et al. 2018). In fact, lysin and VERL are shown to be rapidly

coevolving with each other, potentially driven by sexual conflict (Clark, et al. 2009). This rapid coevolution of lysin and VERL could lead to species-specific binding interactions.

In this study we compare conspecific and heterospecific lysin-VERL ZP-N binding affinities using purified natural lysin and *Pichia pastoris*-expressed VERL ZP-N domains from the abalone species *H. rufescens* and *H. fulgens* (common name Green abalone). Red and Green abalone are the most divergent within the North American clade, and therefore likely to be sufficiently diverged to detect differences in relative binding between species. Further both species are the most well-studied in fertilization research with crystal structures available for red and green abalone lysin, and red abalone VERL ZP-N domains (Kresge, et al. 2000b, a; Raj, et al. 2017). The sequence identity between pairs of green and red abalone ZP-N domains is 85% for VR1, 92% for VR2, and 94% for VR3-23. We hypothesize that the higher the sequence divergence, the more likely we are to observe a strong species-specific lysin-VERL ZP-N binding interaction. In mammals, the zona pellucida protein ZP2 has also been shown to be undergoing positive selection in humans and rodents. When the ZP2's ZP-N domains are compared across *Mus musculus* and *Homo sapiens*, a similar pattern is observed as seen in abalone, with sequence identity increasing as you move towards the C-terminus (sequence identity between mouse and human ZP2 repeat 1 47%, repeat 2 51%, repeat 3 66%). The most divergent domain is the N-terminal ZP-N domain that is shown to mediate species-specific sperm binding between mouse and humans (Avella, et al. 2014; Wilburn and Swanson 2016). Here we investigate whether there is evidence for conservation of interaction mechanisms between abalone VERL and mammalian ZP2.

4.2 Results and Discussion

4.2.1 *VERL VR1 Shows Species-Specific Lysin Binding Function*

Using SPR, red abalone lysin's binding affinity to red or green abalone VR1, 2, and 3 were estimated (Figure 1). Surface Plasmon Resonance is a commonly used technique for quantifying binding affinities. Red abalone lysin bound at similar affinity to red and green abalone VR3 (2.9 nM and 3.6 nM respectively). This estimated binding affinity is approximately what was measured between red abalone lysin and full-length red abalone VERL, consistent with VERL being composed of 21 domains with the same sequence as VR3 (Swanson and Vacquier 1997; Galindo, et al. 2002). Red abalone lysin did show species-specific binding function to VR2 with Red abalone lysin binding nearly three times as strongly to conspecific VR2 than green abalone VR2 (5.7 nM vs 16 nM respectively). Although this almost threefold difference does indicate that lysin-VR2 interaction may partially mediate species-specific fertilization function, this difference is modest and red abalone lysin's interactions with both red and green abalone VR2 are tight. The binding affinity estimates of red abalone lysin binding to either red abalone VR2 or VR3 was consistent with previously published measurements (Raj, et al. 2017). The largest difference in relative binding function is observed between red abalone lysin and red or green abalone VR1. In the SPR analysis, Red abalone VR1 is shown to qualitatively bind to red abalone VR1, but to not detectably bind to Green abalone VR1. While qualitative, this data implies a robust species-specific lysin-VR1 interaction. We quantitatively verified this result using NMR chemical shift perturbation analysis.

NMR chemical shift perturbation binding analysis can be performed at higher protein concentrations than SPR, allowing quantification of lower binding affinity interactions (Figure

2). For the NMR analysis, lysin was isolated from both red and green abalone testes and red and green abalone VERL domains were expressed using *Pichia pastoris*. The strongest VR1 interaction we observed was between red abalone lysin to conspecific VR1. Red abalone lysin has a binding affinity of 1.5 uM to red abalone VERL VR1 according to NMR perturbation analysis. This result indicates that for red abalone lysin there is a greater than three orders of magnitude difference in binding affinity between red VERL VR3 and VR1. A previous study was unable to detect the binding interaction between lysin and VR1 since experiments were not performed at micromolar lysin concentrations (Raj, et al. 2017). By investigating binding affinities within the micromolar range this study was able to accurately quantify this interaction. Red abalone lysin binds much more weakly to Green abalone VR1 (6.4 uM) than conspecific VR1 (1.5 uM). However, green abalone lysin to heterospecific red abalone VR1 was the second strongest evaluated binding interaction, followed by Green abalone lysin binding to conspecific VR1.

At first glance, the VR1-lysin binding affinity measurements may seem paradoxical (Figure 2) since green abalone lysin binds more strongly to red abalone VR1 then its own VR1. However, if the focus is on the egg's selectivity, which is what is likely to determine fertilization success, it is apparent that VR1 more strongly binds to conspecific lysin. These findings are consistent with sexual selection on lysin by VERL, potentially this selective pressure contributes to the coevolution of lysin and VERL (Clark, et al. 2009).

4.2.2 *Species-Specific Lysin-VERL VR1 Binding Interactions Correlate with Species-Specific Lysin VE Dissolution Function*

Lysin's species-specific VE dissolution function has been experimentally characterized in many conspecific and heterospecific pairings. Abalone produce gametes in prolific quantities, and large quantities of sperm lysin and egg VEs can be easily isolated. VE dissolution by lysin is not enzymatic but stoichiometric, therefore lysin's dissolution function can be analyzed by the amount of lysin required to dissolve the VE sample as measured through spectroscopy.

Previous research examining red or green abalone-derived lysin's ability to dissolve VEs from either red or green abalone, indicates that red and green abalone VEs can be better dissolved by conspecific than heterospecific lysin (Swanson and Vacquier 1997; Kresge, et al. 2000b). Figure 2 features lysin VE dissolute experiments adapted with permission from Kresge et al. 2000. The order of VE dissolution efficacy between lysin and VE pairs mirrors the order of strength of lysin-VR1 binding interactions. In fact, the estimated binding affinities between lysin and VR1 pairs correlates significantly (pearson's correlation coefficient > 0.99) with the concentration of lysin resulting in 50% VE dissolution. This correlation is suggestive of the VR1-lysin interaction explaining the observed specificity of lysin's VE dissolution function.

4.2.3 *Mammalian ZP2 R1 and VERL R1 Bind to the Same Region of Abalone Lysin in a Similar Orientation*

Abalone VERL and mammalian ZP2 are proteins associated with species-specific sperm-egg coat interactions. Both egg coat proteins are composed of a tandem ZP-N array with domains associated with species-specific sperm interactions. Although the sperm binding partner of ZP2 is unknown, because of the similarities in structures of ZP2's first repeat (ZP2 R1) and VERL

VR1, we decided to investigate whether ZP2 R1 can bind abalone lysin in a similar manner as VERL VR1. We examined the interactions between ZP2 R1 and VERL VR1 to red abalone lysin in two directions: lysin's perturbation in the presence of the ZP-N domains (Figure 3) and experimental modeling of each ZP-N domain's docking orientation to lysin (Figure 4).

A previous study created an NMR structure of red abalone lysin where each backbone amine had been assigned. We collected a NMR spectrum of red abalone lysin by itself and then with either red abalone VERL VR1 or mouse ZP2 R1. We observed that similar residues of lysin were perturbed in the presence of either ZP-N domain, although the magnitude of the perturbation was lower for mouse ZP2 R1 (Figure 3). These results indicate that both red abalone VR1 and mouse ZP2 R1 are interacting with the same region of lysin, albeit mouse ZP2 R1 binds to lysin more weakly.

Using NMR paramagnetic relaxation enhancement, we performed experimentally guided molecular docking to determine that both abalone VERL VR1 and mouse ZP2 R1 domains bind the same region of lysin (Figure 4). The orientation of mouse ZP2 R1 interacting with lysin was similar to the orientation for abalone VR1 determined in this study as well as the orientation detected for lysin to abalone VR2 and VR3 via crystallography in a previous study (Raj, et al. 2017). Together, these NMR results are consistent with a shared mechanism between mammalian and abalone sperm-egg coat interactions despite having no sequence homology.

4.3 CONCLUSION

Lysin is essential for VE dissolution in abalone and species-specific lysin-VE interactions are sufficient to form a powerful boundary to hybridization. Red abalone eggs and green abalone

sperm have a low fertilization success rate (~10%) however, upon introduction of conspecific sperm lysin, fertilization success can be recovered to over 90% (Leighton 1999) (Figure 5). In this study we show through detailed biochemical characterization of heterospecific and conspecific lysin-VERL interactions that the interaction between lysin and VERL VR1 can largely explain the long-observed species-specific VE dissolution function of lysin. Further we observe that it is the selectivity of egg VERL, rather than sperm lysin, that is responsible for species-specific gamete recognition in abalone. This reflects the fact that a much greater number of sperm are produced in abalone (just as in other animal species) than eggs, in keeping with the much higher investment of resources in egg- rather than sperm-production.

In this study, we observe that the species-specific lysin binding function of VERL ZP-N domains correlates with ZP-N sequence divergence between species. A previous study that was unable to detect lysin-VERL VR1 binding, suggested that the rapid sequence diversification of VERL VR1 was driven by pathogen resistance, not coevolution with VERL (Raj, et al. 2017). However, sites under selection in lysin and VERL VR1 are clustered within the binding interfaces and NMR docking analysis between lysin and VERL VR1 indicate that lysin is interacting with VERL VR1 in the same orientation as what was also observed by Raj et al. 2017 for VERL VR2 and VR3 interacting with lysin. Therefore, the rapid diversification of VERL VR1 is likely the result of sexual selection driving the coevolution of lysin and VERL.

A surprising result from this study is that the strength of lysin to VERL ZP-N binding interactions are inversely correlated with species-specificity. Therefore, the rapid coevolution of lysin and VERL ZP-Ns likely results in relatively weak but crucial binding interactions for

mediating species-specific gamete interactions. This is similar to what has been observed for transcription factor binding sites, where specific but weak affinity interactions enable target gene specificity (Crocker, et al. 2016; Brodsky, et al. 2021). While lysin's interactions with each ZP-N domain of VERL contribute to VE dissolution function, it is the lysin to N-terminal VERL ZP-N domain interaction that appears to mediate species-specific dissolution phenotypes.

Despite sharing no similarity at the sequence level, ZP2 ZP-N R1 and VERL VR1 both bind to lysin in a similar way and have similar protein structures, pointing to deep conservation of gamete recognition mechanisms in diverse taxonomic groups. Further investigation into mammalian gamete recognition genes may reveal a sperm protein with a similar structure to abalone lysin.

4.4 Materials and Methods

4.4.1 *Purification of Lysin from Live Abalone*

Haliotis rufescens abalone were obtained from The Abalone Farm, Inc in Cayucos, CA, USA. Lysin from red abalone was isolated and purified by CM cellulose chromatography (Vacquier & Lee, 1993). *Haliotis fulgens* lysin had been isolated from wild abalone and provided by Dr. Victor Vacquier at the Scripps Institution of Oceanography in San Diego, California. *H. rufescens* and *H. fulgens* concentrated lysin samples were buffer exchanged into 0.01 M HEPES pH 7.4, 0.5 M NaCl, 3 mM EDTA, 0.005% v/v Surfactant P20, using an equilibrated 5 mL Q Sepharose column (Sigma-Aldrich).

4.4.2 *Expression and Purification of VERL*

Recombinant VERL repeat 1, 2, and 3 was cloned into pPICZαA with an N-terminal ATCUN tag. Protein plasmids were propagated in BL21 *E. coli* (New England Biolabs), linearized by PCR, and transformed into SuperMan5 pep4-Δ1, sub2-Δ1, aox1-Δ1 P. pastoris cells (Biogrammatics) by electroporation. Expression conditions were optimized based on methods from Rodriguez and Rama Krishna (2001, J. Biochem, 130, 19-22). Expression clones were initially cultured overnight at 29ºC, 250 rpm in 5 mL YPD (1% w/v yeast extract, 2% w/v peptone, 2% glucose) with 100 µg/mL Zeocin to ensure activation of the expression construct, and initial cultures aliquoted into 300 mL YPD to achieve OD600 ~ 6 after an additional 24 hrs of culturing at 29ºC, 250 rpm. Cells were then harvested by centrifugation at 3000 x g for 10 minutes, briefly resuspended in 10mL BMM (0.5% methanol, 1.34% YNB, 0.00004% biotin, 100 mM potassium phosphate, pH 6) to remove residual glucose, centrifugation repeated, and cells resuspended in 75mL BMM to induce protein expression. Cultures were incubated at 29ºC, 250 rpm for 48 hours with an additional 0.5% methanol added after 24 hours. Cultures were centrifuged at 3000 x g for 10 minutes to remove cells, and the culture supernatant was filtered using 0.45 µm polyethylsulfone filters with glassfiber prefilters to remove residual cells. The supernatant was then concentrated and buffer exchanged using 10-kDa centrifugal ultrafilters (Millipore) in preparation for ion exchange chromatography. For VR1, supernatants were exchanged into 50 mM NaCl/20mM sodium acetate, pH 5, which was also used to equilibrate a packed 5 mL SP-Sepharose column (Sigma-Aldrich). Following application of VR1 samples, the column was washed with > 6 column volumes (CVs) of 50 mM NaCl/20mM sodium acetate, pH 5 and VR1 eluted with 3 CVs of 300 mM NaCl/20mM sodium acetate, pH 5. Elutions were concentrated and buffer exchanged into 20 mM Tris, pH 8, and applied to a 5 mL Q Sepharose

column (Sigma-Aldrich) equilibrated in 20 mM Tris, pH 8 to remove contaminant proteins with purified VR1 retained in the flowthrough. For VR2 and VR3, supernatants were exchanged into 20 mM Tris, pH 8, applied to a 5 mL Q Sepharose column (Sigma-Aldrich) equilibrated in 20 mM Tris, pH 8, washed with > 6 CVs 20 mM Tris, pH 8, and proteins eluted with 3 CVs of 60 mM NaCl/20 mM Tris, pH 8.

4.4.3 *Binding Affinity Estimation Via Surface Plasmon Resonance*

Two Series S CM5 SPR chips had VERL repeats 1, 2, and 3 from either *H. rufescens* or *H. fulgens* were immobilized to channels on the chip surface with amine coupling. One channel was left with no ligand and used to detect background binding. Instructions for creating the chip can be found in the GE Healthcare Life Science Biacore T200 Handbook. For the SPR experiments a running buffer of 0.01 M HEPES pH 7.4, 0.5 M NaCl, 3 mM EDTA, 0.005% v/v Surfactant P20 was used. Red Lysin concentrations ranging from 2 nM to 500 nM were used to estimate binding affinity to either red or green abalone VERL ZP-N domains. Each run the analyte was flowed over the surface at 30 ul/minute for four minutes followed by 2 minutes of dissociation, and finally two rounds of surface regeneration with glycine ph 2.0. Data from SPR analysis were fit using a two binding site version of the hill equation where separate Kds were estimated for the surface background and each immobilized ZP-N ligand. The data was fit with a nonlinear least squares regression using a custom python script.

4.4.4 *Binding Affinity Estimation Via NMR Equilibrium Constants*

Binding affinities between lysin and VR1 were measured by 1H NMR titrations for all combinations of proteins between red and green abalone. NMR spectra were recorded using a 500 Mhz Bruker Avance III NMR spectrometer for lysin purified from abalone testis (90 μM)

with concentrations of VR1 ranging from 0 to 117 μM in 200 mM NaCl/10 mM Tris, pH 7.4/7% D2O. As the lysin W34 HE1 resonance is downfield compared to all other protons within either lysin or VR1 from either species, its signal is easily distinguishable throughout the titration and its change in linewidth proportional to the percent of lysin bound by VR1, with Kd estimated by nonlinear least squares regression using the script function curve_fit.

4.4.5 *Lysin NMR Pertubation and NMR Docking*

Lysin-ZPN interactions for four different diverse ZPN domains (red abalone VR1, green abalone VR1, mouse ZP2N1, and human ZP2N1) were studied by NMR-based experiments comparing the 15N transverse relaxation rates (R2) under multiple conditions with delays of 8.48, 16.96, 25.44, 33.92, 42.40, 50.88, and 59.36 ms. Monomeric red abalone lysin (mutation F104A) with 15N labeling was expressed and purified as described in Wilburn et al. (2018) and standardized to 140 μM in 200 mM NaCl/10 mM Tris, pH 7.4/7% D2O, and transverse relaxation rates measured for lysin alone and with 42 μM (0.3 molar equivalents) of each ZPN domain. After spectra were recorded for each lysin-ZPN combination, 38 μM CuSO4 was added to each tube and R2 measurements recorded to infer paramagnetic relaxation enhancement effects on lysin from Cu2+ binding to the ATCUN motif present on each ZPN N-terminus. Changes in 15N R2 between free and ZPN-bound lysin as well as ZPN-bound lysin with and without Cu2+ were estimated by nonlinear least squares regression using the scipy function curve_fit. Docking of lysin to red abalone VR1 or mouse ZP2N1 were performed using restrained molecular dynamics simulations with Xplor-NIH v2.53. Starting structures were constructed using PDB models 5utg (lysin F104A solution structure), 5ii4 (red abalone VR1 crystal structure), and 5ii6 (mouse ZP2N1 crystal structure), and protein docking performed by simulated annealing from 4000 to 25 K using restraints derived from changes in R2 upon ZPN binding and PRE measurements.

4.5 Acknowledgments

This study was supported by NIH Grant HD076862 to Willie J. Swanson, a National Science Foundation Graduate Research Fellowship to Jolie A. Carlisle, and K99 Fellowship to Damien B. Wilburn. Thank you to Harmit Malik, Doug Fowler, and Daniel Promislow for useful comments. Further thanks to the MAF facilities at UW and the technical support of John Sumida and Ellen Lavoie. All research was performed on the traditional lands of the Duwamish Tribe. To learn more about the Duwamish Tribe and their continuing legacy, please visit https://www.duwamishtribe.org/.

4.6 FIGURES

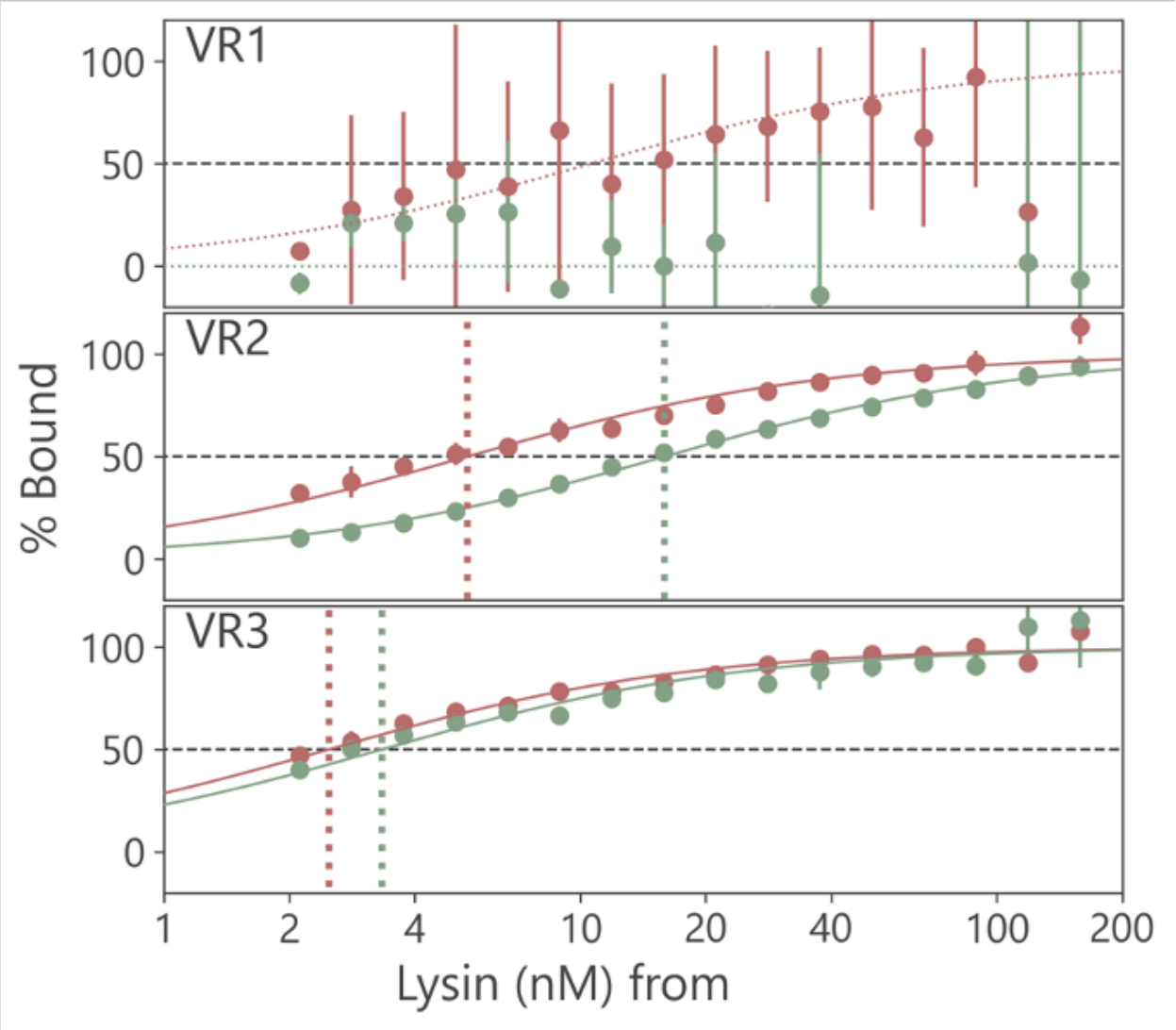


Figure 4.1. Surface Plasmon Resonance Reveals an Inverse Correlation Between Species-Specificity and Binding Strength

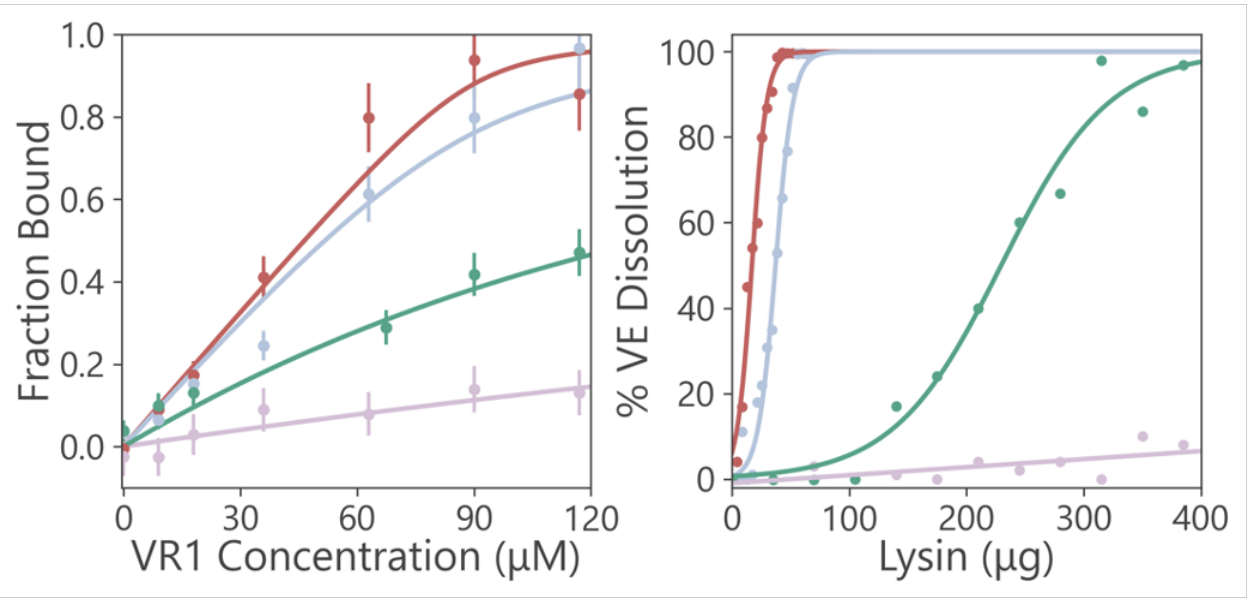


Figure 4.2. Lysin-VERL VR1 Binding affinities correlate with Lysin VE Dissolution Function.

A. NMR chemical shift perturbation analysis reveals binding affinities between conspecific and heterospecific pairs of abalone lysin and VERL R1.

B. Lysin more efficiently dissolves isolated conspecific VE than heterospecific VE. Figure adapted with permission from Kresge, Vacquier, & Stout, 2000.

Red Line – Red Lysin vs Red VERL VR1/Red VE
Blue Line – Green Lysin vs Red VERL VR1/Red VE
Green Line – Green Lysin vs Green VERL VR1/GreenVE
Pink Line – Red Lysin vs Green VERL VR1/Green VE

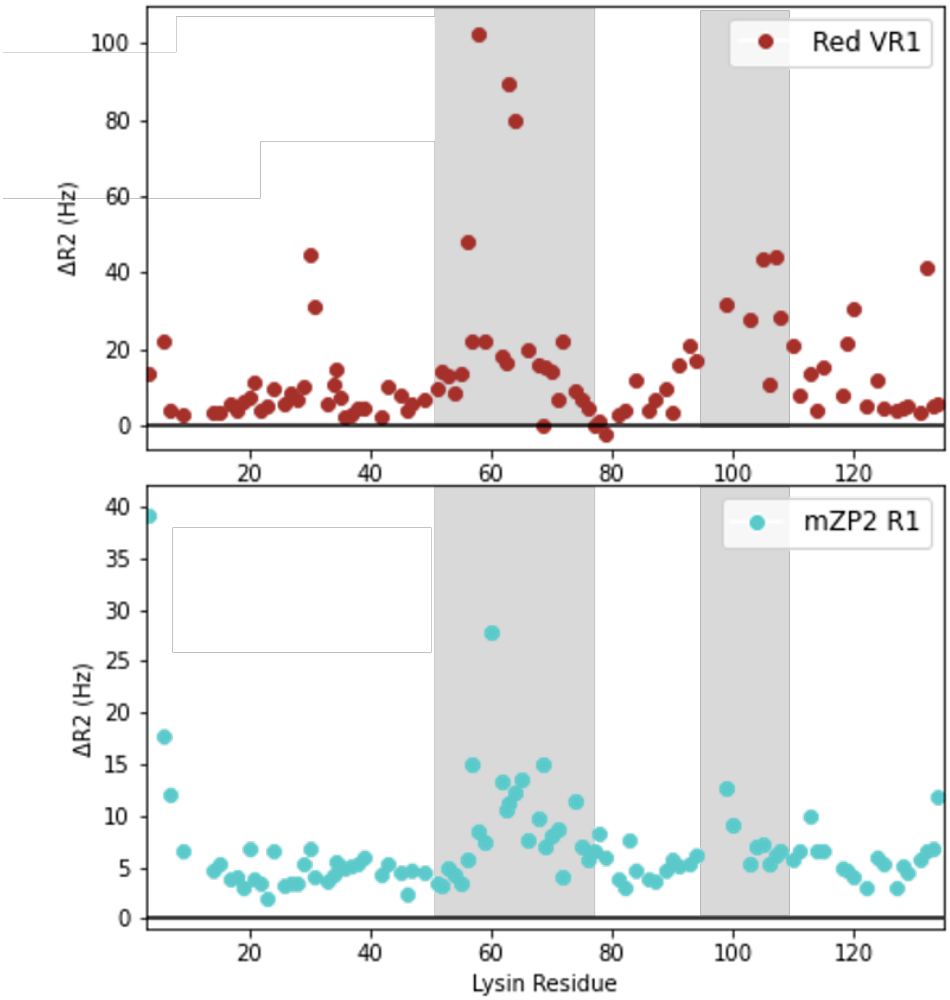
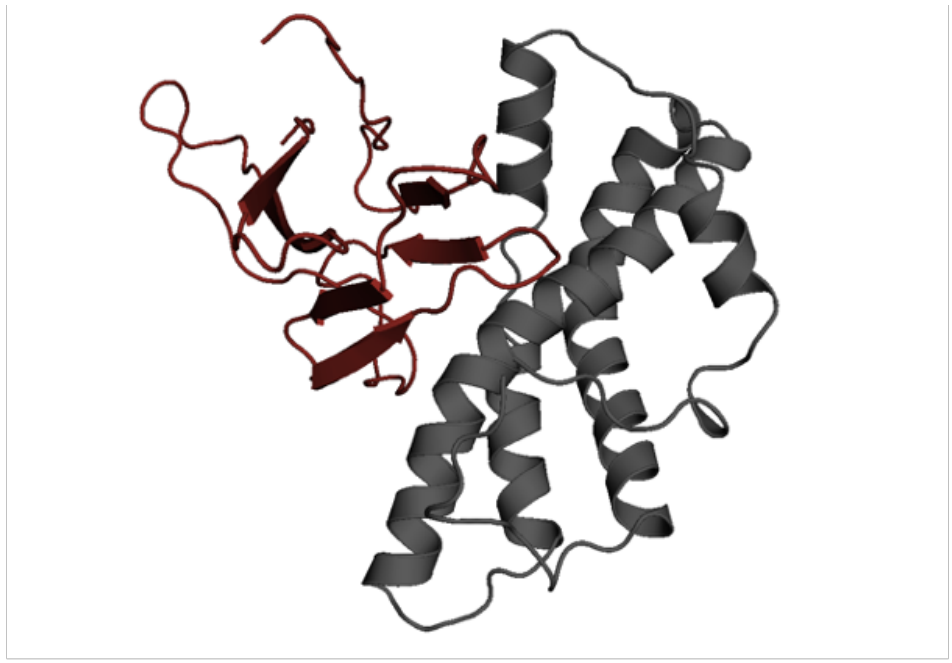


Figure 4.3. Similar regions of *H. rufescens* lysin are perturbed in the presence of *H. rufescens* abalone VERL R1 and *M. musculus* ZP2 R1.



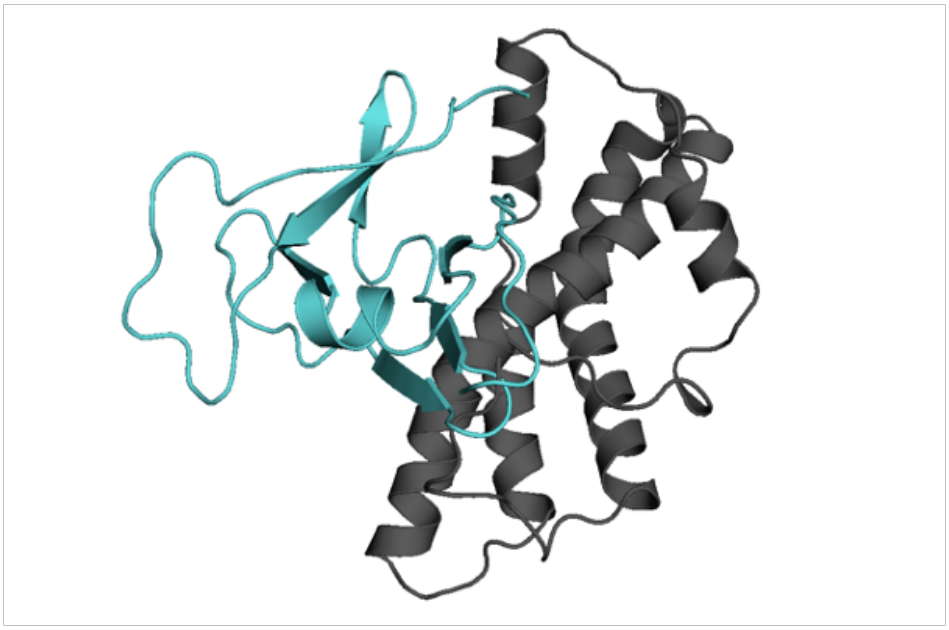


Figure 4.4. Mouse ZP2 R1 and Abalone VERL VR1 both interact with abalone lysin similarly.

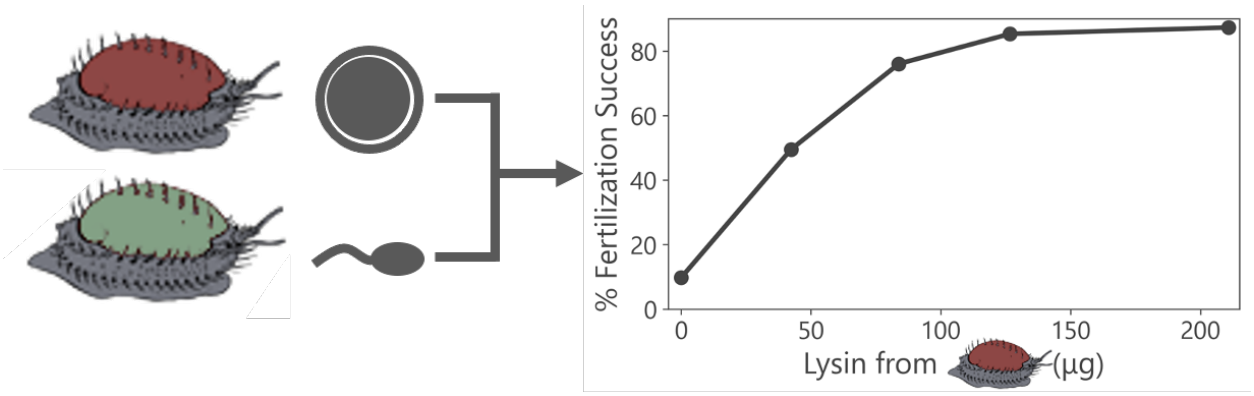


Figure 4.5. Hybrid Fertilization is Rescued by Conspecific Lysin

Hybrid fertilization success between *H. rufescens* eggs and *H. fulgens* sperm is rescued with the addition of *H. rufescens* lysin. Figure adapted with permission from Leighton, 1999.

Chapter 5. CONCLUSION AND FUTURE DIRECTIONS

The three research chapters of my thesis each focused on how different evolutionary mechanisms generate the molecular diversity of fertilization genes and the possible implications of this diversification on biochemical mechanisms of fertilization. Here, I summarize the main findings of each research chapter and directions for future research.

5.1 GENE DUPLICATION

Following duplication, paralogs may parse ancestral gene function (subfunctionalization) or acquire new roles (neofunctionalization). The marine gastropod mollusk abalone has two acrosomal proteins (lysin and sp18) which are ancient gene duplicates with unique gamete recognition functions that presumably arose through subfunctionalization. Through detailed genomic and bioinformatic analyses I have shown how duplication events followed by sequence diversification has played an ongoing role in the evolution of abalone acrosomal proteins beyond the split between lysin and sp18. The common ancestor of abalone had four members of its acrosomal protein family (two lysin paralogs and two sp18 paralogs) in a tandem gene array that repeatedly experienced positive selection. Both sp18 and its paralog sp18-dup both contain positively selected sites located in different regions of the sp18 structure, consistent with a subfunctionalization model where selection acted upon distinct binding interfaces in each paralog.

My research also uncovered recent species-specific duplications of both lysin and sp18 in the European abalone *H. tuberculata*. Despite possessing clade-specific acrosomal protein paralogs,

there are no concomitant duplications of egg coat proteins in *H. tuberculata*, indicating that duplication of egg proteins per se is not responsible for retention of duplicated acrosomal proteins. I hypothesize that, in a manner analogous to host/pathogen evolution, sperm proteins are selected for increased diversity through extensive sequence divergence and recurrent duplication driven by conflict mechanisms.

During my thesis I was unable to get a full sequence of VERL from *H. tuberculata* or any other abalone species outside the North American clade. VERL was identified as a long repetitive sequence in *H. rufescens*, and is difficult to sequence using PacBio approaches due to premature sequence termination from the C-terminal end of the molecule. Although I found no evidence of a whole gene duplication of any abalone VEZPs, potentially duplications of abalone ZP-N domains within VERL or VERL allelic variation could be driving the duplication of *H. tuberculata* lysin. Using illumina sequencing of the ovary transcriptome, it should be possible to acquire the sequence of the repetitive ZP-N array of VERL (assuming VERL is structured similarly to *H. rufescens* in *H. tuberculata*). Using this sequence I could use the Rapid Amplification of cDNA Ends (RACE) approach to acquire 3' VERL sequence in H. tuberculata. With sequence from the 3' and 5' end of the VERL transcript, I would be able to use PCR to determine the length of the molecule. Further, using *H. tuberculata* abalone population samples, I can sequence the N-terminal ZP-N domains of *H. tuberculata* VERL and determine whether VERL is undergoing positive selection within the population or contains allelic variation. Chapter 4 of my thesis investigated the binding interactions between lysin and VERL ZP-N domains. It would be interesting to investigate the binding interactions between both copies of H. tuberculata lysin and the ZP-N domains of *H. tubeculata* VERL.

5.2 Gene Loss

In the second research chapter of my thesis, I used syntenic and phylogenetic analysis to identify the pseudogene *C4BPAP1* as the primate ortholog of the sperm fertilization gene ZP3r. Previous studies have misidentified human *ZP3r* due to nomenclature confusion or difficulties in establishing orthology. *ZP3r* is found in chromosome 1 amongst paralogous protein-coding genes that make up the RCA cluster. Many of these genes are diverging rapidly between species. This sequence divergence of the paralogs can complicate accurate ortholog identification. Although ZP3r is associated with ZP binding in mice, *ZP3r* shows repeated pseudogenization in primates (at least 9 times), most notably in apes. In primate species where *ZP3r* has not acquired pseudogenizing mutations, *ZP3r* is evolving under positive selection, indicating that relaxed selection is likely not leading to *ZP3r* loss in primates.

Difficulty in finding fertilization proteins may be driven by functional redundancy causing many fertilization genes to be nonessential contributors to gamete recognition. Typically, protein functional redundancy refers to paralogous proteins that are structurally similar, that maintain the same interaction partners, and whose loss can be compensated for by their paralog. However, proteins can also be functionally redundant without being paralogous or structurally similar if either unrelated protein can compensate for the loss of the other. Functional redundancy of genes mediating fertilization could lead to clade-specific gene loss events or changes in relative functional importance between species.

This chapter highlights the potential variability of molecular mechanisms of fertilization even within mammals and emphasizes the value of using diverse model systems for investigating

mechanisms of fertilization. Knock outs of potential fertilization genes should be performed in more model systems, like hamsters or rabbits, and the screens to assay fertilization phenotypes should be more nuanced than to screen for sterility.

5.3 Sequence Divergence

As one of only a handful of known interacting sperm-egg protein pairs, the abalone (genus *Haliotis*) lysin-VERL interaction is an important model for understanding the molecular mechanisms mediating fertilization. Abalone egg VERL is composed of a series of tandem ZP-N domains. During abalone fertilization, sperm lysin dissolves the vitelline envelope (VE) of the egg by binding the ZP-N domains of VERL. ZP-N domains are a common feature of egg coat proteins, and by understanding how lysin-VERL interactions mediate dissolution of the egg vitelline envelope (VE) in abalone we may be able to gain insight into the biochemical mechanisms mediating egg coat dissolution in mammals.

VE dissolution by lysin is species-specific, presumably due to the rapid coevolution of lysin and VERL driven by sexual conflict. Since lysin shows species-specific dissolution function, it is likely that sequence divergence of lysin and VERL between species creates differences in conspecific and heterospecific binding affinities. Using surface plasmon resonance (SPR), a quantitative biophysical method, I measured the binding affinities between heterospecific and conspecific lysin-VERL domain pairs. I discovered that species-specific binding between lysin and VERL VR1 was consistent with the experimentally measured species-specific VE dissolution function of lysin. It is remarkable to recapitulate species-specific fertilization between these distinct proteins.

ZP2, a mammalian egg coat protein, shows similarity in evolutionary patterns, protein structure, and function to abalone VERL despite extensive divergence between the proteins (ZP-N domains between mammals and abalone are not recognized by BLAST but share similar 3D structures). ZP2 has a tandem array of ZP-N domains and the first N-terminal ZP-N domain mediates species-specific sperm binding between mice and humans. Using NMR chemical shift perturbation analysis, we measured binding of *H. rufescens* lysin to the N-terminal ZP-N of ZP2 from humans and mice. Experimentally guided molecular docking indicates that both abalone and mammalian ZP-N domains bind the same region of lysin. These NMR results are consistent with a shared mechanism between mammalian and abalone sperm-egg interactions.

Future research could focus on identifying mammalian sperm proteins that share a structural similarity to abalone lysin (although likely no sequence identity) that may be performing a similar egg coat dissolution function in mammals.

5.1 Concluding Remarks

My thesis highlights how the diversification of fertilization genes in sequence and gene content across even closely related species can potentially impact mechanisms of fertilization and provide new insights into the complexity of the pathways of interaction between the sperm and egg. My thesis also emphasizes the value of using diverse model systems for investigating mechanisms of fertilization.

In my post-doctoral work I hope to better understand how gene duplication paired with sequence divergences contributes to the diversification of reproduction mechanisms between even closely related species of *Drosophila*. *Drosophila* is a remarkable system for investigating complex

phenotypes due to the availability of sequenced DGRP lines, high-quality genomes from closely related species, and streamlined methods for genetically modifying flies.

BIBLIOGRAPHY

Aagaard JE, Springer SA, Soelberg SD, Swanson WJ. 2013. Duplicate abalone egg coat proteins bind sperm lysin similarly, but evolve oppositely, consistent with molecular mimicry at fertilization. PLoS Genet 9:e1003287.

Aagaard JE, Vacquier VD, MacCoss MJ, Swanson WJ. 2010. 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 27:193-203.

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

Adham IMN, K. Engel, W. 1998. Spermatozoa lacking acrosin protein show delayed fertilization. Molecular Reproduction and Development 46:370-376.

Almagro Armenteros JJ, Tsirigos KD, Sonderby CK, Petersen TN, Winther O, Brunak S, von Heijne G, Nielsen H. 2019. SignalP 5.0 improves signal peptide predictions using deep neural networks. Nat Biotechnol 37:420-423.

Almeida FC, Desalle R. 2009. Orthology, function and evolution of accessory gland proteins in the Drosophila repleta group. Genetics 181:235-245.

Alves AP, Mulloy B, Diniz JA, Mourao PA. 1997. Sulfated polysaccharides from the egg jelly layer are species-specific inducers of acrosomal reaction in sperms of sea urchins. J Biol Chem. 272:6965-6971.

Anisimova M, Bielawski JP, Yang Z. 2001. Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. Mol Biol Evol 18:1585-1592.

Avella MA, Baibakov B, Dean J. 2014. A single domain of the ZP2 zona pellucida protein mediates gamete recognition in mice and humans. J Cell Biol 205:801-809.

Baba T, Azuma S, Kashiwabara S, Toyoda Y. 1994. Sperm from mice carrying a targeted mutation of the acrosin gene can penetrate the oocyte zona pellucida and effect fertilization. J Biol Chem 269:31845-31849.

Bainova H, Kralova T, Bryjova A, Albrecht T, Bryja J, Vinkler M. 2014. First evidence of independent pseudogenization of toll-like receptor 5 in passerine birds. Dev Comp Immunol 45:151-155.

Barbaux S, Ialy-Radio C, Chalbi M, Dybal E, Homps-Legrand M, Do Cruzeiro M, Vaiman D, Nishimura K, Han L, Bianchi E, Wright GJ, de Sanctis D, Jovine L. 2016. The structure of sperm Izumo1 reveals unexpected similarities with Plasmodium invasion proteins. Curr Biol 26:R661 662.

Bekpen C, Marques-Bonet T, Alkan C, Antonacci F, Leogrande MB, Ventura M, Kidd JM, Siswara P, Howard JC, Eichler EE. 2009. Death and resurrection of the human IRGM gene. PLoS Genet 5:e1000403.

Bentley JK, Shimomura H, Garbers DL. 1986. Retention of a functional resact receptor in isolated sperm plasma membranes.

Bentley JK, Tubb DJ, Garbers DL. 1986. Receptor-mediated activation of spermatozoan guanylate cyclase. J Biol Chem 261:14859-14862.

Betters M, Levitan DR. 2018. The Role of Male Variation in Fertilization Success in Determining

the Costs and Benefits of Polyandry in the Broadcast Spawning Urchin Lytechinus variegatus. Biol Bull 235:63-70.

Bhakta HH. Refai FH, Avella MA. 2019. The molecular mechanisms mediating mammalian fertilization. Development 146.

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

Bianchi E, Wright GJ. 2014. Izumo meets Juno: preventing polyspermy in fertilization. Cell Cycle

13:2019-2020.

Bianchi E, Wright GJ. 2015. Cross-species fertilization: the hamster egg receptor, Juno, binds the human sperm ligand, Izumo1. Philos Trans R Soc Lond B Biol Sci 370:20140101.

Bianchi E, Wright GJ. 2020. Find and fuse: Unsolved mysteries in sperm-egg recognition. PLoS Biol 18:e3000953.

Bleil JD, Wassarman PM. 1990. Identification of a ZP3-binding protein on acrosome-intact mouse sperm by photoaffinity crosslinking. Proc Natl Acad Sci U S A 87:5563-5567.

Bleil JD, Wassarman PM. 1980. Mammalian sperm-egg interaction: identification of a glycoprotein in mouse egg zonae pellucidae possessing receptor activity for sperm. Cell 20:873-882.

Bleil JD, Wassarman PM. 1988. Galactose at the nonreducing terminus of O-linked oligosaccharides of mouse egg zona pellucida glycoprotein ZP3 is essential for the glycoprotein's sperm receptor activity. Proc Natl Acad Sci U S A 85:6778-6782.

Boja ES, Hoodbhoy T, Fals HM, Dean J. 2003. Structural characterization of native mouse zona pellucida proteins using mass spectrometry. J Biol Chem 278:34189-34202.

Botwright NA, Zhao M, Wang T, McWilliam S, Colgrave ML, Hlinka O, Li S, Suwansa-Ard S, Subramanian S, McPherson L, et al. 2019. Greenlip Abalone (Haliotis laevigata) Genome and Protein Analysis Provides Insights into Maturation and Spawning. G3 (Bethesda) 9:3067-3078.

Brodsky S, Jana T, Barkai N. 2021. Order through disorder: The role of intrinsically disordered regions in transcription factor binding specificity. Curr Opin Struct Biol 71:110-115.

Buffone MG, Ijiri TW, Cao W, Merdiushev T, Aghajanian HK, Gerton GL. 2012. Heads or tails? Structural events and molecular mechanisms that promote mammalian sperm acrosomal exocytosis and motility. Mol Reprod Dev 79:4-18.

Buffone MG, Rodriguez-Miranda E, Storey BT, Gerton GL. 2009. Acrosomal exocytosis of mouse sperm progresses in a consistent direction in response to zona pellucida. J Cell Physiol 220:611-620.

Buffone MG, Zhuang T, Ord TS, Hui L, Moss SB, Gerton GL. 2008. Recombinant mouse sperm ZP3-binding protein (ZP3R/sp56) forms a high order oligomer that binds eggs and inhibits mouse fertilization in vitro. J Biol Chem 283:12438-12445.

Burkart AD, Xiong B, Baibakov B, Jimenez-Movilla M, Dean J. 2012. Ovastacin a cortical granule protease, cleaves ZP2 in the zona pellucida to prevent polyspermy. J Cell Biol 197:37 44.

Cai X, Clapham DE. 2008. Evolutionary genomics reveals lineage-specific gene loss and rapid evolution of a sperm-specific ion channel complex: CatSpers and CatSperbeta. PLoS One 3:e3569.

Carithers LJ, Moore HM. 2015. The Genotype-Tissue Expression (GTEx) Project. Biopreserv Biobank 13:307-308.

Carlisle JA, Swanson WJ. 2020. Molecular mechanisms and evolution of fertilization proteins. J Exp Zool B Mol Dev Evol.

Carlisle JA, Glenski, M.A., Swanson, W.J. 2021. Recurrent Duplication and Diversification of Acrosomal Fertilization Proteins in Abalone. BioRxiv.

Chalabi S, Panico M, Sutton-Smith M, Haslam SM, Patankar MS, Lattanzio FA, Morris HR, Chatterjee HJ, Ho SY, Barnes I, Groves C. 2009. Estimating the phylogeny and divergence times of primates using a supermatrix approach. BMC Evol Biol 9:259.

Chalbi M, Barraud-Lange V, Ravaux B, Howan K, Rodriguez N, Soule P, Ndzoudi A, Noucheix C, Rubinstein E, Wolf JP, Ziyyat A, Perez E, Pincet F, Gourier C. 2014. Binding of sperm protein

Izumo1 and its egg receptor Juno drives Cd9 accumulation in the intercellular contact area prior to fusion during mammalian fertilization. Development 141:3732-3739.

Clark GF, Dell A. 2006. Differential O-glycosylation of a conserved domain expressed in murine and human ZP3. Biochemistry 45:636-647.

Clark NL, Aagaard JE, Swanson WJ. 2006. Evolution of reproductive proteins from animals and plants. Reproduction 131: 11-22.

Clark NL, Findlay GD, Yi X, MacCoss MJ, Swanson WJ. 2007. Duplication and selection on abalone sperm lysin in an allopatric population. Mol Biol Evol 24:2081-2090.

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

Clark NL, Swanson WJ. 2005. Pervasive adaptive evolution in primate seminal proteins. PLoS Genet 1: 335-342.

Claw KG, George RD, Swanson WJ. 2014. Detecting coevolution in mammalian sperm-egg

fusion proteins. Mol Reprod Dev 81:531-538.

Cooper JC, Phadnis N. 2017. Parallel Evolution of Sperm Hyper-Activation Ca2+ Channels. Genome Biol Evol 9:1938-1949.

Crocker J, Noon EP, Stern DL. 2016. The Soft Touch: Low-Affinity Transcription Factor Binding Sites in Development and Evolution. Curr Top Dev Biol 117:455-469.

Dangott LJ, Jordan JE, Bellet RA, Garbers DL. 1989. Cloning of the mRNA for the protein that crosslinks to the egg peptide speract. Proc Natl Acad Sci U S A 86:2128-2132.

Doty KA, Wilburn DB, Bowen KE, Feldhoff PW, Feldhoff RC. 2016. Co-option and evolution of non-olfactory proteinaceous pheromones in a terrestrial lungless salamander. J Proteomics 135:101-111.

Ebert TA. 2008. Longevity and lack of senescence in the red sea urchin Strongylocentrotus franciscanus. Exp Gerontol 43:734-738.

Eisenbach M. 1999. Mammalian sperm chemotaxis and its association with capacitation. Dev Genet 25:87-94.

Eisenbach M, Giojalas LC. 2006. Sperm guidance in mammals – an unpaved road to the egg. Nat Rev Mol Cell Biol 7:276-285.

Ellerman DA, Pei J, Gupta S, Snell WJ, Myles D, Primakoff P. 2009. Izumo is part of a multiprotein family whose members form large complexes on mammalian sperm. Mol Reprod Dev 76:1188-1199.

Evans JP, Sherman CD. 2013. Sexual selection and the evolution of egg-sperm interactions in broadcast-spawning invertebrates. Biol Bull 224:166-183.

Findlay GD, Sitnik JL, Wang W, Aquadro CE, Clark NL, Wolfner MF. 2014. Evolutionary rate covariation identifies new members of a protein network required for Drosophila melanogaster female post-mating responses. PLoS Genet 10:e1004108.

Findlay GD, Yi X, Maccoss MJ, Swanson WJ. 2008. Proteomics reveals novel Drosophila seminal fluid proteins transferred at mating. PLoS Biol 6:e178.

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

Foltz KR, Partin JS, Lennarz WJ. 1993. Sea urchin egg receptor for sperm: sequence similarity of binding domain and hsp70. Science 259:1421-1425.

Franke ES, Babcock RC, Styan CA. 2002. Sexual conflict and polyspermy under sperm-limited conditions: in situ evidence from field simulations with the free-spawning marine echinoid Evechinus chloroticus. Am Nat 160:485-496.

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

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

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

Gan HM, Tan MH, Austin CM, Sherman CDH, Wong YT, Strugnell J, Gervis M, McPherson L, Miller AD. 2019. Best Foot Forward: Nanopore Long Reads, Hybrid Meta-Assembly, and Haplotig Purging Optimizes the First Genome Assembly for the Southern Hemisphere Blacklip Abalone (Haliotis rubra). Front Genet 10:889.

Garcia-Fernandez J, Vilches-Arroyo S, Olavarrieta L, Perez-Perez J, Rodriguez de Cordoba S. 2021. Detection of Genetic Rearrangements in the Regulators of Complement Activation RCA Cluster by High-Throughput Sequencing and MLPA. Methods Mol Biol 2227:159-178.

Gavrilets S, Waxman D. 2002. Sympatric speciation by sexual conflict. Proc Natl Acad Sci U S A 99:10533-10538.

Glabe CG, Vacquier VD. 1977. Isolation and characterization of the vitelline layer of sea urchin eggs. J Cell Biol. 75:410-421.

Glabe CG, Vacquier, VD. 1977. Species specific agglutination of eggs by bindin isolated from sea urchin sperm. Nature 267:836-838.

Glabe CG, Lennarz, WJ. 1979. Species-specific sperm adhesion in sea urchins. A quantitative investigation of bindin-mediated egg agglutination. J Cell Biol 83:595-604.

Grayson P. 2015. Izumo1 and Juno: the evolutionary origins and coevolution of essential sperm-egg binding partners. R Soc Open Sci 2:150296.

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

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

Han C, Choi E, Park I, Lee B, Jin S, Kim DH, Nishimura H, Cho C. 2009. Comprehensive analysis of reproductive ADAMs: relationship of ADAM4 and ADAM6 with an ADAM complex

required for fertilization in mice. Biol Reprod 80:1001-1008.

Hansbrough JR, Garbers JR. 1981. Speract. Purification and characterization of a peptide associated with eggs that activates spermatozoa. J Biol Chem 256:1447-1452.

Hardy MH, Tatsuo H, Garbers DL. 1994. Sea urchin sperm receptors for egg peptides. Seminars in Developmental Biology 5:217-224.

Hart MW, Stover DA, Guerra V, Mozaffari SV, Ober C, Mugal CF, Kaj I. 2018. Positive selection on human gamete-recognition genes. PeerJ 6:e4259.

Hart MW, Sunday JM, Popovic I, Learning KJ, Konrad CM. 2014. Incipient speciation of sea star

populations by adaptive gamete recognition coevolution. Evolution 68:1294:1305.

Heinen S, Sanchez-Corral P, Jackson MS, Strain L, Goodship JA, Kemp EJ, Skerka C, Jokiranta TS, Meyers K, Wagner E, et al. 2006. De novo gene conversion in the RCA gene cluster (1q32) causes mutations in complement factor H associated with atypical hemolytic uremic syndrome. Hum Mutat 27:292-293.

Hino T, Muro Y, Tamura-Nakano M, Okabe M, Tateno H, Yanagimachi R. 2016. The Behavior and Acrosomal Status of Mouse Spermatozoa In Vitro, and Within the Oviduct During Fertilization after Natural Mating. Biol Reprod 95:50.

Hirose M, Honda A, Fulka H, Tamura-Nakano M, Matoba S, Tomishima T, Mochida K, Hasegawa A, Nagashima K, Inoue K, et al. 2020. Acrosin is essential for sperm penetration through the zona pellucida in hamsters. Proc Natl Acad Sci U S A 117:2513-2518.

Hofmeyer T, Schmelz S, Degiacomi MT, Dal Peraro M, Daneschdar M, Scrima A, van den Heuvel J, Heinz DW, Kolmar H. 2013. Arranged sevenfold: structural insights into the C-terminal oligomerization domain of human C4b-binding protein. J Mol Biol 425:1302-1317.

Holt WV, Fazeli A. 2014. Do sperm possess a molecular passport? Mechanistic insights into

sperm selection in the female reproductive tract. Molecular Human Reproduction 21:491-501.

Hourcade D, Holers VM, Atkinson JP. 1989. The regulators of complement activation (RCA) gene cluster. Adv Immunol 45:381-416.

Huang Y, Niu B, Gao Y, Fu L, Li W. 2010. CD-HIT Suite: a web server for clustering and comparing biological sequences. Bioinformatics 26:680-682.

Hughes J, Ward, CJ, Aspinwall R, Butler R, Harris, PC. 1999. Identification of a Human Homologoue of the Sea Urchin Receptor for Egg jelly: A Polycystic Kidney Disease-Like Protein. Hum Mol Genet 8:543-549.

Hussain YH, Guasto JS, Zimmer RK, Stocker R, Riffell JA. 2016. Sperm chemotaxis promotes individual fertilization success in sea urchins. J Exp Biol 219:1458-1466.

Inoue N, Ikawa M, Isotani A, Okabe M. 2005. The immunoglobulin superfamily protein Izumo is

required for sperm to fuse with eggs. Nature 434:234-238.

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

Jaffe LA. 1976. Fast block to polyspermy in sea urchin eggs is electrically mediated. Nature 261:68-71.

Jean C, Haghighirad F, Zhu Y, Chalbi M, Ziyyat A, Rubinstein E, Gourier C, Yip P, Wolf JP, Lee

JE, Boucheix C, Barraud-Lange V. 2019. JUNO, the receptor of sperm IZUMO1, is expressed by the human oocyte and is essential for fertilization. Human Reprod 34:118-126.

Just ML, Lennarz WJ. 1997. Reexamination of the sequence of the sea urchin egg receptor for

sperm: implications with respect to its properties. Dev Biol 184:25-30.

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

Kanai S, Ono C, Takahashi Y, Kimura K, Shida T, Miyado K, Umezawa A. 2012. CD81 and CD9

work independently as extracellular components upon fusion of sperm and oocyte. Biol Open 1:640-647.

Kaupp UB, Kashikar ND, Weyand I. 2008. Mechanisms of sperm chemotaxis. Annu Rev Physiol 70:93-117.

Killingbeck EE, Swanson WJ. 2018. Egg Coat Proteins Across Metazoan Evolution. Curr Top Dev Biol 130:443-488.

Kekalainen J, Evans JP. 2018. Gamete-mediated mate choice: towards a more inclusive view of sexual selection. Proc Biol Sci 285.

Keller SH, Vacquier VD. 1994. The isolation of acrosome-reaction-inducing glycoproteins from sea urchin egg jelly. Dev Biol 162:304-312.

Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. 2002. The human genome browser at UCSC. Genome Res 12:996-1006.

Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. 2019. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. Bioinformatics 35:4453-4455.

Kresge N, Vacquier VD, Stout CD. 2001. The crystal structure of a fusagenic sperm protein reveals extreme surface properties. Biochemistry 40:5407-5413.

Kresge N, Vacquier VD, Stout CD. 2000a. 1.35 and 2.07 A resolution structures of the red abalone sperm lysin monomer and dimer reveal features involved in receptor binding. Acta Crystallogr D Biol Crystallogr 56:34-41.

Kresge N, Vacquier VD, Stout CD. 2000b. The high resolution crystal structure of green abalone sperm lysin: implications for species-specific binding of the egg receptor. J Mol Biol 296:1225-1234.

Krushkal J, Bat O, Gigli I. 2000. Evolutionary relationships among proteins encoded by the regulator of complement activation gene cluster. Mol Biol Evol 17:1718-1730.

Lazzaro BP, Clark, A.G. 2012. Rapid evolution of innate immune response genes. In. Rapidly Evolving Genes & Genetic Systems: PMC Exempt – Book Chapter.

Le SQ, Gascuel O. 2008. An improved general amino acid replacement matrix. Mol Biol Evol 25:1307-1320.

Lee YH, Ota T, Vacquier VD. 1995. Positive selection is a general phenomenon in the evolution of abalone sperm lysin. Mol Biol Evol 12:231-238.

Lehmann R. 2018. Matchmaking molecule for egg and sperm. Science 361:974-975.

Leighton DL. 1999. The biology and culture of california abalones.

Lemoine F, Domelevo Entfellner JB, Wilkinson E, Correia D, Davila Felipe M, De Oliveira T, Gascuel O. 2018. Renewing Felsenstein's phylogenetic bootstrap in the era of big data. Nature 556:452-456.

Levitan DR. 2004. Density-dependent sexual selection in external fertilizers: variances in male and female fertilization success along the continuum from sperm limitation to sexual conflict in the sea urchin Strongylocentrotus franciscanus. Am Nat 164:298-309.

Levitan DR. 2018. Do Sperm Really Compete and Do Eggs Ever Have a Choice? Adult

Distribution and Gamete Mixing Influence Sexual Selection, Sexual Conflict, and the Evolution of Gamete Recognition Proteins in the Sea. AM Nat 191:88-105.

Levitan DR, Buchwalter R, Hao Y. 2019. The evolution of gametic compatibility and compatibility

groups in the sea urchin Mesocentrotus franciscanus: An avenue for speciation in the sea. Evolution 73:1428-1442.

Levitan DR, Ferrell DL. 2006. Selection on gamete recognition proteins depends on sex, density, and genotype frequency. Science 312:267-269.

Levitan DR, Terhorst CP, Fogarty ND. 2007. The risk of polyspermy in three congeneric sea urchins and its implications for gametic incompatibility and reproductive isolation. Evolution 61:2007-2014.

Lewis CA, Leighton DL, Vacquier VD. 1980. Morphology of abalone spermatozoa before and after the acrosome reaction. J Ultrastruct Res 72:39-46.

Lin YN, Roy A, Yan W, Burns KH, Matzuk MM. 2007. Loss of zona pellucida binding proteins in the acrosomal matrix disrupts acrosome biogenesis and sperm morphogenesis. Mol Cell Biol 27:6794-6805.

Litscher ES, Wassarman PM. 1996. Characterization of mouse ZP3-derived glycopeptide, gp55, that exhibits sperm receptor and acrosome reaction-inducing activity in vitro. Biochemistry 35:3980-3985.

Litscher ES, Williams Z, Wassarman PM. 2009. Zona pellucida glycoprotein ZP3 and fertilization in mammals. Mol Reprod Dev 76:933-941.

Lorenzetti D, Poirier C, Zhao M, Overbeek PA, Harrison W, Bishop CE. 2014. A transgenic insertion on mouse chromosome 17 inactivates a novel immunoglobulin superfamily gene

potentially involved in sperm-egg fusion. Mamm Genome 25:141-148.

Lyon JD, Vacquier VD. 1999. Interspecies chimeric sperm lysins identify regions mediating species-specific recognition of the abalone egg vitelline envelope. Dev Biol 214:151-159.

MacDonald RJ, Swift GH, Przybyla AE, Chirgwin JM. 1987. Isolation of RNA using guanidinium salts. Methods Enzymol 152:219-227.

Magel RCW, S. H. 1997. Comparing the Powers of the Wald-Wolfowitz and Kolmogorov-Smirnov Tests. Biometrical Jornal 39:665.

Mah SA, Swanson WJ, Vacquier VD. 2005. Positive selection in the carbohydrate recognition domains of sea urchin sperm receptor for egg jelly (suREJ) proteins. Mol Biol Evol 22:533-541.

Mandal A, Sachdev S, Digilio L, Panneerdoss S, Suryavathi V, Pires E, Flickinger C, Herr J. 2011. SAS1B is an egg specific high affinity oolemmal binding partner for sperm specific acrosomal SLLP1 during fertilization. Biology of Reproduction 85.

Masonbrink RE, Purcell CM, Boles SE, Whitehead A, Hyde JR, Seetharam AS, Severin AJ. 2019. An Annotated Genome for Haliotis rufescens (Red Abalone) and Resequenced Green, Pink, Pinto, Black, and White Abalone Species. Genome Biol Evol 11:431-438.

Matsumoto M, Solzin J, Helbig A, Hagen V, Ueno S, Kawase O, Maruyama Y, Ogiso M, Godde M, Minakata H, Kaupp UB, Hoshi M, Weyand I. 2003. A sperm-activating peptide controls a cGMP-signaling pathway in starfish sperm. Dev Biol 260:314-324.

Mauk R, Jaworski D, Kamei N, Glabe CG. 1997. Identification of a 97-kDa heat shock protein from S. franciscanus ovaries with 94% amino acid identity to the S. purpuratus egg surface receptor for sperm. Dev Biol 184:31-37.

McGinnis S, Madden TL. 2004. BLAST: at the core of a powerful and diverse set of sequence analysis tools. Nucleic Acids Res 32:W20-25.

Metz EC, Kane RE, Yanagimachi H, Palumbi SR. 1994. Fertilization between closely related sea urchins is blocked by incompatibilities during sperm-egg attachment and early stages of fusion. Biol Bull 187:23-34.

Metz EC, Robles-Sikisaka R, Vacquier VD. 1998. Nonsynonymous substitution in abalone sperm fertilization genes exceeds substitution in introns and mitochondrial DNA. Proc Natl Acad Sci U S A 95:10676-10681.

Miga KH, Newton Y, Jain M, Altemose N, Willard HF, Kent WJ. 2014. Centromere reference models for human chromosomes X and Y satellite arrays. Genome Res 24:697-707.

Miyata H, Castaneda JM, Fujihara Y, Yu Z, Archambeault DR, Isotani A, Kiyozumi D, Kriseman ML, Mashiko D, Matsumura T, et al. 2016. Genome engineering uncovers 54 evolutionarily conserved and testis-enriched genes that are not required for male fertility in mice. Proc Natl Acad Sci U S A 113:7704-7710.

Monne M, Jovine L. 2011. A structural view of egg coat architecture and function in fertilization.

Biol Reprod 85:661-669.

Morgan CC, Hart MW. 2019. Molecular evolution of mammalian genes with epistatic interactions in fertilization. BMC Evol Biol 19:154.

Morgan CC, Loughran NB, Walsh TA, Harrison AJ, O'Connell MJ. 2017. Erratum to: Positive selection neighboring functionally essential sites and disease-implicated regions of mammalian reproductive proteins. BMC Evol Biol 17:170.

Morgan CC, Loughran NB, Walsh TA, Harrison AJ, O'Connell MJ. 2010. Positive selection neighboring functionally essential sites and disease-implicated regions of mammalian reproductive proteins. BMC Evol Biol 10:39.

Mori E, Baba T, Iwamatsu A, Mori T. 1993. Purification and characterization of a 38-kDa protein, sp38, with zona pellucida-binding property from porcine epididymal sperm. Biochem Biophys Res Commun 196:196-202.

Mouse Genome Sequencing C, Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P, Agarwala R, Ainscough R, Alexandersson M, et al. 2002. Initial sequencing and comparative analysis of the mouse genome. Nature 420:520-562.

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

Moy GW, Vacquier VD. 1979. Immunoperoxidase localization of bindin during the adhesion of sperm to sea urchin eggs. Curr Top Dev Biol. 187:23-34.

Mozingo NM, Vacquier VD, Chandler DE. 1995. Structural features of the abalone egg extracellular matrix and its role in gamete interaction during fertilization. Mol Reprod Dev 41:493-502.

Muro Y, Buffone MG, Okabe M, Gerton GL. 2012. Function of the acrosomal matrix: zona pellucida 3 receptor (ZP3R/sp56) is not essential for mouse fertilization. Biol Reprod 86:1-6.

Muro Y, Hasuwa H, Isotani A, Miyata H, Yamagata K, Ikawa M, Yanagimachi R, Okabe M. 2016. Behavior of Mouse Spermatozoa in the Female Reproductive Tract from Soon after Mating to the Beginning of Fertilization. Biol Reprod 94:80.

Nam BH, Kwak W, Kim YO, Kim DG, Kong HJ, Kim WJ, Kang JH, Park JY, An CM, Moon JY, et al. 2017. Genome sequence of pacific abalone (Haliotis discus hannai): the first draft genome in family Haliotidae. Gigascience 6:1-8.

Ohlendieck K, Dhume ST, Partin JS, Lennarz WJ. 1993. The sea urchin egg receptor for sperm:

isolation and characterization of the intact, biologically active receptor. J Cell Biol 122:887-895.

Ohnami N, Nakamura A, Miyado M, Sato M, Kawano N, Yoshida K, Harada Y, Takezawa Y, McLaughlin RN Jr., Malik HS. 2017. Genetic conflicts: the usual suspects and beyond. J Exp Biol 220:6-17.

Okabe M. 2018. Sperm-egg interaction and fertilization: past, present, and future. Biol Reprod 99:134-146.

Okroj M. BAM. 2018. C4b-binding protein. In: Barnumb S. ST, editor. The complement handbook. New York: Elsevier.

Ottaviano G, Zuccarello D, Menegazzo M, Perilli L, Marioni G, Frigo AC, Staffieri A, Foresta C. 2013. Human olfactory sensitivity for bourgeonal and male infertility: a preliminary investigation. Eur Arch Otorhinolaryngol 270:3079-3086.

Palmer MR, McDowall MH, Stewart L, Ouaddi A, MacCoss MJ, Swanson WJ. 2013. Mass spectrometry and next-generation sequencing reveal an abundant and rapidly evolving abalone sperm protein. Mol Reprod Dev 80:460-465.

Palumbi SR, Lessios HA. 2005. Evolutionary animation: how do molecular phylogenies compare to Mayr's reconstruction of speciation patterns in the sea? Proc Natl Acad Sci U S A 102:6566 6572.

Panhuis TM, Clark NL, Swanson, WJ. 2006. Rapid evolution of reproductive proteins in abalone and Drosophila. Philos Trans R Soc Lond B Biol Sci 361:261-268.

Pardo-Manuel de Villena F, Rodriguez,S. 1995. C4BPAL2: A second duplication of the C4BPA gene in the human RCA gene cluster. Immunogenetics 41.

Park CY, Klammer AA, Kall L, MacCoss MJ, Noble WS. 2008. Rapid and accurate peptide identification from tandem mass spectra. J Proteome Res 7:3022-3027.

Pei J, Kim BH, Grishin NV. 2008. PROMALS3D: a tool for multiple protein sequence and structure alignments. Nucleic Acids Res 36:2295-2300.

Pei J, Tang M, Grishin NV. 2008. PROMALS3D web server for accurate multiple protein sequence and structure alignments. Nucleic Acids Res 36:W30-34.

Raj I, Sadat Al Hosseini H, Dioguardi E, Nishimura K, Han L, Villa A, de Sanctis D, Jovine L. 2017. Structural Basis of Egg Coat-Sperm Recognition at Fertilization. Cell 169:1315-1326 e1317.

Ramirez-Gomez HV, Sabinina VJ, Perez MV, Beltran C, Carneiro J, Wood CD, Tuval I, Darszon A, Guerrero A. 2020. Sperm chemotaxis is driven by the slope of the chemoattractant concentration field. Elife 9.

Reams AB, Roth JR. 2015. Mechanisms of gene duplication and amplification. Cold Spring Harb Perspect Biol 7:a016592.

Riffell JA, Krug PJ, Zimmer RK. 2002. Fertilization in the sea: the chemical identity of an abalone sperm attractant. J Exp Biol 205:1439-1450.

Riffell JA, Krug PJ, Zimmer RK. 2004. The ecological and evolutionary consequences of sperm chemoaatraction. Proc Natl Acad Sci U S A 101:4501-4506.

Rohlfs RV, Swanson WJ, Weir BS. 2010. Detecting coevolution through allelic association between physically unlinked loci. Am J Hum Genet 86:674-685.

Rubinstein E, Ziyyat A, Prenant M, Wrobel E, Wolf J, Levy S, Le Naour F, Boucheix C. 2006. Reduced fertility of female mice lacking CD81. Dev Biol 290:351-358.

Sabeur K, Edwards DP, Meizel S. 1996. Human sperm plasma membrane progesterone receptor(s) and the acrosome reaction. Biol Reprod 54:993-1001.

Sachdev M, Mandal A, Mulders S, Digilio LC, Panneerdoss S, Suryavathi V, Pires E, Klotz KL, Hermens L, Herrero MB, Flickinger CJ, van Duin M, Herr JC. 2012. Oocyte specific oolemmal SAS1B involved in sperm binding through intra-acrosomal SLLP1 during fertilization. Dev Biol. 363:40-51.

Sanchez-Corral P, Pardo-Manuel de Villena F, Rey-Campos J, Rodriguez de Cordoba S. 1993. C4BPAL1, a member of the human regulator of complement activation (RCA) gene cluster that resulted from the duplication of the gene coding for the alpha-chain of C4b-binding protein. Genomics 17:185-193.

Sato K. 1979. Polyspermy-preventing mechanisms in mouse eggs fertilized in vitro. J Exp Zool. 210:353-359.

Sievers F, Higgins DG. 2014. Clustal Omega, accurate alignment of very large numbers of sequences. Methods Mol Biol 1079:105-116.

Sinding C, Kemper E, Spornraft-Ragaller P, Hummel T. 2013. Decreased perception of bourgeonal may be linked to male idiopathic infertility. Chem Senses 38:439-445.

Sirot LK, Findlay GD, Sitnik JL, Frasheri D, Avila FW, Wolfner MF. 2014. Molecular characterization and evolution of a gene family encoding both female- and male-specific reproductive proteins in Drosophila. Mol Biol Evol 31:1554-1567.

Slater GS, Birney E. 2005. Automated generation of heuristics for biological sequence comparison. BMC Bioinformatics 6:31.

Spehr M, Gisselmann G, Poplawski A, Riffell JA, Wetzel CH, Zimmer RK, Hatt H. 2003. Identification of a testicular odorant receptor mediating human sperm chemotaxis. Science 299:2054-2058

Stapper AP, Beerli P, Levitan DR. 2015. Assortative mating drives linkage disequilibrium

between sperm and egg recognition protein loci in the sea urchin Strongylocentrotus purpuratus. Mol Biol Evol 32:859-870.

Stears RL, Lennarz WJ. 1997. Mapping sperm binding domains on the sea urchin egg receptor for sperm. Dev Biol 187:200-208.

Stewart-Savage J, Bavister BD. 1988. A cell surface block to polyspermy occurs in golden hamster eggs. Dev Biol 128:150-157.

Strunker T, Goodwin N, Brenker C, Kashikar ND, Weyand I, Seifert R, Kaupp UB. 2011. The CatSper channel mediates progesterone-induced Ca2+ influx in human sperm. Nature 471:382-386.

Styan CA, Butler AJ. 2000. Fitting fertilization kinetics models for free-spawning marine invertebrates. Marine Bio 137:943-951.

Sutton KA, Jungnickel MK, Florman HM. 2008. A polycystin-1 controls postcopulatory reproductive selection in mice. Proc Natl Acad Sci U S A 105:8661-8666.

Sutton KA. Jungnickel MK, Ward CJ, Harris PC, Florman HM. 2006. Functional characterization of PKDREJ, a male germ cell-restricted polycystin. J Cell Physiol 209:493-500.

Sutton KA, Jungnickel MK, Ward CJ, Harris PC, Florman HM. 2006. Functional characterization of PKDREJ, a male germ cell-restricted polycystin. J Cell Physiol 209:493-500.

Suyama M, Torrents D, Bork P. 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res 34:W609-612.

Suzuki N. 1995. Structure, function and biosynthesis of sperm-activating peptides and fucose sulfate glycoconjugate in the extracellular coat of sea urchin eggs. Zoolog Sci 12:13-27.

Swann CA, Cooper SJ, Breed WG. 2007. Molecular evolution of the carboxy terminal region of the zona pellucida 3 glycoprotein in murine rodents. Reproduction 133:697-708.

Swann CA, Cooper SJB, Breed WG. 2017. The egg coat zona pellucida 3 glycoprotein - evolution of its putative sperm-binding region in Old World murine rodents (Rodentia: Muridae). Reprod Fertil Dev 29:2376-2386.

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

Swanson WJ, Aquadro CF, Vacquier VD. 2001. Polymorphism in abalone fertilization proteins is

consistent with the neutral evolution of the egg's receptor for lysin (VERL) and positive Darwinian selection of sperm lysin. Mol Biol Evol 18:376-383.

Swanson WJ, Clark AG, Waldrip-Dail HM, Wolfner MF, Aquadro CF. 2001. Evolutionary EST analysis identifies rapidly evolving male reproductive proteins in Drosophila. Proc Natl Acad Sci 98: 7375-7379.

Swanson WJ, Nielsen R, Yang Q. 2003. Pervasive adaptive evolution in mammalian fertilization proteins. Mol Biol Evol 20:18-20.

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

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

Swanson WJ, Vacquier VD. 1995. Extraordinary divergence and positive Darwinian selection in a fusagenic protein coating the acrosomal process of abalone spermatozoa. Proc Natl Acad Sci U S A 92:4957-4961.

Swanson WJ, Vacquier VD. 2002. The rapid evolution of reproductive proteins. Nat Rev Genet 3:137-144.

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

Teves ME, Guidobaldi HA, Unates DR, Sanchez R, Miska W, Publicover SJ, Morales Garcia AA, Giojalas LC. 2009. Molecular mechanism for human sperm chemotaxis mediated by progesterone. PLoS One 4:e8211.

Tokuhiro K, Dean J. 2018. Glycan-Independent Gamete Recognition Triggers Egg Zinc Sparks and ZP2 Cleavage to Prevent Polyspermy. Dev Cell, 46:627-640.

Trimmer JS, Vacquier VD. 1986. Activation of sea urchin gametes. Annu Rev Cell Biol 2:1-26.

Turner LM, Hoekstra HE. 2006. Adaptive evolution of fertilization proteins within a genus: variation in ZP2 and ZP3 in deer mice (Peromyscus). Mol Biol Evol 23:1656-1669.

Turner LM, Hoekstra. 2008. Causes and consequences of the evolution of reproductive proteins. Int J Dev Biol 52:769-80.

Ulrich AS, Otter M, Glabe CG, Hoekstra D. 1998. Membrane fusion is induced by a distinct peptide sequence of the sea urchin fertilization protein bindin. J Biol Chem 273:16748-16755.

Vacquier VD, Moy GW. 1977. Isolation of bindin: the protein responsible for adhesion of sperm to sea urchin eggs. Proc Natl Acad Sci U S A 74:2456-2460.

Vacquier VD, Moy GW. 1997. The fucose sulfate polymer of egg jelly binds to sperm REJ and is the inducer of the sea urchin sperm acrosome reaction. Dev Biol 192:125-135.

Vacquier VD, Swanson WJ. 2011. Selection in the rapid evolution of gamete recognition proteins in marine invertebrates. Cold Spring Harb Perspect Biol 3:a002931.

Velova H, Gutowska-Ding MW, Burt DW, Vinkler M. 2018. Toll-Like Receptor Evolution in Birds: Gene Duplication, Pseudogenization, and Diversifying Selection. Mol Biol Evol 35:2170-2184.

Villanueva-Diaz C, Arias-Martinez J, Bermejo-Martinez L, Vadillo-Ortega F. 1995. Progesterone induces human sperm chemotaxis. Fertil Steril 64:1183-1188.

Wagstaff BJ, Begun DJ. 2005. Comparative genomics of accessory gland protein genes in Drosophila melanogaster and D. pseudoobscura. Mol Biol Evol 22:818-832.

Wang M, Lv Z, Shi J, Hu Y, Xu C. 2009. Immunocontraceptive potential of the Ig-like domain of

Izumo. Mol Reprod Dev 76:794-801.

Wassarman PM. 2009. Mammalian fertilization: the strange case of sperm protein 56. Bioessays 31:153-158.

Wilburn DB, Arnold, J. A., Houck, L. D., Feldhoff, P. W., Feldhoff, R. C. 2017. Gene Duplication, Co-option, Structural Evolution, and Phenotypic Tango in the Courtship Pheromones of Plethodontid Salamanders. Herpetologica 73:206-219.

Wilburn DB, Swanson WJ. 2016. From molecules to mating: Rapid evolution and biochemical studies of reproductive proteins. J Proteomics 135:12-25.

Wilburn DB, Swanson WJ. 2017. The “ZP domain” is not one, but likely two independent domains. Mol Reprod Dev 84:284-285.

Wilburn DB, Tuttle LM, Klevit RE, Swanson WJ. 2018. Solution structure of sperm lysin yields novel insights into molecular dynamics of rapid protein evolution. Proc Natl Acad Sci U S A 115:1310-1315.

Vacquier VD. 2012. The quest for the sea urchin egg receptor for sperm. Biochem Biophys Res Commun 425:583-587.

Vacquier VD. 2019. My research career on (mainly) sea urchins. Methods Cell Biol 1515:21-26.

Vacquier VD, Lee YH. 1993. Abalone sperm lysin: unusual mode of evolution of a gamete recognition protein. Zygote 1:181-196.

Vicens A, Gomez Montoto L, Couso-Ferroer F, Sutton KA, Roldan ER. 2015. Sexual selction and the adaptive evolution of PKDREJ protein in primates and rodents. Mol Hum Reprod 21:146-156.

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

Ward GE, Brokaw CJ, Garbers DL, Vacquier VD. 1985. Chemotaxis of Arbacia punctilata spermatozoa to resact, a peptide from the egg jelly layer. J Cell Biol 101:2323-2329.

Welborn J, Manahan D. 1995. Taurine metabolism in larvae of marine invertebrate molluscs (Bivalvia Gastropoda). J Exp Biol 198:1791-1799.

Wolf J, Ziyyat A. 2020. Sperm SPACA6 protein is required for mammalian Sperm-Egg Adhesion/Fusion. Sci Rep 10:5335.

Wu J, Li H, Zhang S. 2012. Regulator of complement activation (RCA) group 2 gene cluster in zebrafish: identification, expression, and evolution. Funct Integr Genomics 12:367-377.

Xue F, Wang L, Liu Y, Tang H, Xu W, Xu C. 2016. Vaccination with an epitope peptide of IZUMO1 to induce contraception in female mice. Am J Reprod Immunol 75:474-485.

Yang Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. Comput Appl Biosci 13:555-556.

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol 24:1586-1591.

Yang Z, Nielsen R, Goldman N, Pedersen AM. 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics 155:431-449.

Yoshida M, Murata M, Inaba K, Morisawa M. 2002. A chemoattractant for ascidian spermatozoa is a sulfated steroid. Proc Natl Acad Sci U S A 99:14832-14836.

Zigler KS, Lessios Ha. 2003. Evolution of bindin in the pantropical sea urchin Tripneustes: comparisons to bindin of other genera. Mol Biol Evol. 20:220-231.

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

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

VITA

Jolie Carlisle grew up in California and earned a B. S. in Microbiology from the University of California, Berkeley in 2015. As an undergraduate she worked in the Evolutionary Genetics Laboratory with Dr. Carla Cicero and in Dr. Michael Shapira's lab. She started her PhD work in Genome Sciences at the University of Washington in 2015. Her graduate work with Dr. Willie Swanson focused on the evolution and function of fertilization genes mediating sperm-egg coat interactions. She plans to continue to study the evolution and function of reproductive proteins as a post-doctoral scholar at Cornell in the labs of Dr. Mariana Wolfner and Dr. Andrew Clark.