The protective oocyte envelope of threespine stickleback fish

Emily Erskine Killingbeck

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2020

Reading Committee:

Willie J. Swanson, Chair

Catherine L. Peichel

Michael J. MacCoss

Program Authorized to Offer Degree:

Genome Sciences

University of Washington

Abstract

The protective oocyte envelope of threespine stickleback fish

Emily Erskine Killingbeck

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

After the end of the last ice age, ancestrally marine threespine stickleback fish (*Gasterosteus aculeatus*) have undergone an adaptive radiation into freshwater environments throughout the Northern Hemisphere, creating an excellent model system for studying molecular adaptation and speciation. Stickleback populations are reproductively isolated to varying degrees, despite the fact that they can be crossed in the lab to produce viable offspring. Ecological and behavioral factors have been suggested to underlie incipient stickleback speciation. However, reproductive proteins represent a previously unexplored driver of speciation. As mediators of gamete recognition during fertilization, reproductive proteins both create and maintain species boundaries. Gamete recognition proteins are also frequently found to be rapidly evolving, and their divergence may culminate in reproductive isolation and ultimately speciation. As an initial investigation into the contribution of reproductive proteins to stickleback reproductive isolation,

we have characterized the egg coat proteome of threespine stickleback eggs. In agreement with other teleosts, we find that stickleback egg coats are comprised of homologs to the zona pellucida (ZP) proteins ZP1 and ZP3. We explore aspects of stickleback ZP protein biology, including glycosylation, disulfide bonding, and sites of synthesis, and find many substantial differences compared to their mammalian homologs. Furthermore, molecular evolutionary analyses indicate that *ZP3*, but not *ZP1*, has experienced positive Darwinian selection across teleost fish. Taken together, these changes to stickleback ZP protein architecture suggest that the egg coats of stickleback fish, and perhaps fish more generally, have evolved to fulfill a more protective functional role than their mammalian counterparts.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

There are many people I am grateful to for helping make this dissertation possible.

The Swanson Lab has been a wonderful place to do research, and I want to thank all members, past and present: Jan Aagaard, Jennifer McCreight, Alberto Rivera, Jolie Carlisle, and Damien Wilburn. Thank you especially to Damien for teaching me (almost) everything I know about protein biochemistry, and for all the helpful conversations and feedback on my research. Thank you to my advisor, Willie Swanson, for being a fantastic and supportive mentor. Thank you for seeing me as a whole person rather than just a graduate student, and for your unwavering positivity amidst difficult times over the years. Thank you for letting me bring threespine stickleback fish research to the lab, a project we first discussed when I was just interviewing with the Genome Sciences graduate program. And thank you for sending me to conferences in places as far flung as Stony Brook, New York; Queensland's Gold Coast, Australia; Portland, Oregon; Austin, Texas; Kyoto, Japan; Yokohama, Japan; and Manchester, England to present my research – incredible opportunities, especially for a young scientist.

Thank you to the Genome Sciences graduate program for giving me an excellent scientific foundation to build on. The collegial nature of this department first stood out to me during graduate school interviews, and I can confirm that it truly is as collaborative and welcoming a place as it seemed after five and a half years here. I particularly want to thank my committee members Katie Peichel, Mike MacCoss, Debbie Nickerson, Christine Queitsch, and Toby Bradshaw for their guidance. I am especially grateful to Katie for her mentorship, and for helping bring stickleback research to our lab – literally, by giving us traps for collecting fish, and figuratively, by sharing her expertise – and for providing invaluable feedback on numerous grants and manuscripts over

the years. Finally, I want to thank Brian Giebel, administrative wizard, for deftly guiding me through the Ph.D. process.

There are many funding sources that contributed to my graduate training. I was supported by a NIH/NHGRI Genome Training Grant HG000035 and a NIH/NICHD Individual Fellowship F31 HD093441, and my research was funded by a UW Royalty Research Fund award A111769 and by NIH/NICHD R01 HD076862. I am also grateful to the numerous travel and registration awards that helped me attend conferences throughout my Ph.D., exposing me to cutting edge research and allowing me to present my work.

The mentorship of other scientists has been indispensable in my scientific training. I want to thank Craig Miller in particular, who took a chance on an undergraduate with no lab experience and introduced me to the threespine stickleback fish model system. Beyond his skills as an advisor, Craig also had the foresight to pair me with Andrew Glazer as my graduate student mentor. Andrew is a brilliant scientist who taught me almost everything I know about being a scientist, and much about scientific programming as well. His suggestion that I take a class in R programming was particularly prescient, and learning the quirks of the language from him while working on my STAT 133 homework – which you could typically find us completing in the wee hours of the morning at People's Café – has been invaluable training for the computational direction that the genomics field in particular (and society in general) is taking. Another wonderful scientific mentor and friend from the Miller Lab, Nick Ellis, stood with me on my wedding day as our Best Man. And finally, as an intern at Sangamo BioSciences (a position that Craig helped me find), I was extremely fortunate to work with Patrick Li. In addition to being a gifted scientist and valued friend, Patrick patiently listened to my debriefings after each graduate school interview weekend, and he knew that I would be attending UW Genome Sciences even before I did.

Thank you to my fellow classmates for undertaking this crazy Ph.D. journey together. Thank you to my friends for your constant love and support, and for letting me talk science while pretending to be interested in what I was saying. I especially want to thank Arian Motamedi, Spencer Hubbard, and Michael Woody for reminding me of the world outside of graduate school, and for being the best friends that anyone could ask for.

I also want to thank my family, who instilled a love of science in me and my identical twin sister Sarah from a very early age. My grandmother, Dorothea Stacker, was a master's student in zoology at the University of Wisconsin, Madison in the 1940s. Her daughter, my mother Ann Killingbeck, became an accomplished veterinarian. Her other two daughters, Debbie Stacker and Jill Martinez, pursued sociology degrees (Debbie) and a degree in computer science (Jill) before there was such a thing, so she majored in math instead. I am extremely grateful for the female role models my sister and I had, such that neither of us ever questioned our ability to pursue a Ph.D. in science. I think it is no coincidence that growing up in this environment led Sarah to a Ph.D. program as well, studying immunology at UC Davis. I have Sarah to thank for my fascination with genetics, too: it has always amazed me that despite our identical DNA, environmental happenstance has created two very distinct individuals. I love you fiercely my twin Sissa.

Mama and Dadda, thank you for exposing your daughters to the wonders of the natural world and for inspiring their curiosity. Thank you for the strength of your love, which emboldened us to take on any and all challenges, including graduate school. Thank you to my Aunt Debba for applying your sociologist's mind toward making the bewildering world a little less painful for your nieces, giving them the tools in their mental toolboxes that they needed to enable their hopes and dreams. And Debba, thank you for holding the family together during the darkest time of my life, when we were losing my beloved father, Greg Killingbeck, to cancer.

Dadda, I miss you every day. But the love you gave me has shaped me as a person and will be with me always, and nothing can take that away.

Thank you to my incredible husband, Jim Schneidereit, for being there for me during that time, and the difficult times that have come since. Thank you for supporting me endlessly, loving me always, and bringing me so much joy. There is no way I could have made it through this dissertation if I did not have you.

You, and our wonderful black cat Poe – the best lap cat any graduate student staying up late nights could ask for. Together, we are a family.

DEDICATION

For my father, Gregory Erskine Killingbeck (1952-2008).

Erskine, party of two.

Chapter 1. EGG COAT PROTEINS ACROSS METAZOAN EVOLUTION

All animal oocytes are surrounded by a glycoproteinaceous egg coat, a specialized extracellular matrix that serves both structural and species-specific roles during fertilization. Egg coat glycoproteins polymerize into the extracellular matrix of the egg coat using a conserved protein-protein interaction module – the zona pellucida (ZP) domain – common to both vertebrates and invertebrates, suggesting that the basic structural features of egg coats have been conserved across hundreds of millions of years of evolution. Egg coat proteins, as with other proteins involved in reproduction, are frequently found to be rapidly evolving. Given that gamete compatibility must be maintained for the fitness of sexually reproducing organisms, this finding is somewhat paradoxical and suggests a role for adaptive diversification in reproductive protein evolution. In this chapter we review the structure and function of metazoan egg coat proteins, with an emphasis on the potential role their evolution has played in the creation and maintenance of species boundaries (1).

1.1 INTRODUCTION

Fertilization, the union of a single sperm and an egg, is essential to metazoan reproduction. The first contact in fertilization is between sperm and the extracellular matrix of the egg coat, a maternally-derived glycoprotein envelope present in all sexually reproducing animals as well as many asexual metazoans (2-4). Egg coats vary in size from a few tens of microns to over 15 centimeters (5), and are called different names in each major vertebrate lineage: the chorion in fish, the vitelline envelope in amphibians, the perivitelline envelope in reptiles and birds, and the

zona pellucida in mammals (4, 6). For simplicity, however, we will collectively refer to these terms as the "egg coat" throughout this chapter.

Despite their varied nomenclature, the overall structure and function of the egg coat is conserved across vertebrates and invertebrates (6, 7). Egg coats mediate fertilization via sperm recognition and binding, establish blocks to polyspermy, and protect the embryo from biotic (e.g. pathogens, predators) and abiotic (e.g. dehydration, UV radiation, salinity, pollutants) threats (4). Egg coats affect embryonic performance by providing a dispersal and attachment medium in aquatic taxa, and in viviparous species they protect the developing embryo until the egg coat hatches and implants in the wall of the uterus (4, 8, 9). In addition to their conserved function, egg coat ultrastructure is also conserved across metazoans, consisting of a fibrous matrix of conserved components with common protein domains (2, 6, 9, 10). Notably, the domain composition of egg coat proteins and the number of genes encoding them is more variable in invertebrates (2, 4) than in vertebrates (2, 9, 11, 12).

While the basic structure of the egg coat has been conserved for more than 600 million years (7, 9, 10, 12), the proteins that make up the egg coat, as with many proteins involved in reproduction, are frequently found to be rapidly evolving (13-17). This rapid evolution of reproductive proteins is somewhat paradoxical: given the fundamentality of fertilization to species propagation, sperm and egg proteins might be expected to be highly conserved to maintain compatibility. However, this rapid evolution suggests a role for positive Darwinian evolution in creating and maintaining species-specificity during sperm-egg interaction (14, 16, 18, 19).

In this chapter we will discuss the composition and evolutionary history of metazoan egg coat proteins, with an emphasis on the role these factors play in the structure, function, and evolution of the metazoan egg coat in fertilization.

1.2 EGG COAT PROTEINS

Animal egg coat proteins share a common polymerization module called the zona pellucida (ZP) domain (2, 11, 20-23). ZP domain-containing proteins (ZP proteins) are found in the egg coats of all vertebrate taxa (2, 24) and some invertebrate species, including gastropod mollusks (9, 13, 25), cephalochordates (26, 27), and urochordates (28, 29).

Phylogenetic analyses of ZP proteins suggest that the last common ancestor of vertebrates had at least one ancestral ZP gene, with all major ZP gene subfamilies emerging before the divergence of fish and amphibians ~360 million years ago (30, 31). ZP glycoprotein subfamilies have a historically complicated nomenclature, but recent consensus defines six subfamilies that evolved through gene duplication and pseudogenization: ZP1, ZP2/ZPA, ZP3/ZPC, ZP4/ZPB, ZPAX, and ZPD (Figure 1) (2-4, 6, 19). ZP4, for instance, is a pseudogene in mouse, ZP1 is a pseudogene in dog, pig, cat, and cow, and ZPD and ZPAX have been pseudogenized or lost in all mammals (4, 6, 18). Such findings suggest that the evolution of ZP genes occurs mainly by gene death, with the accumulation of stop codons and/or insertions/deletions that disrupt reading frame and cause the loss of protein-coding ability (6).

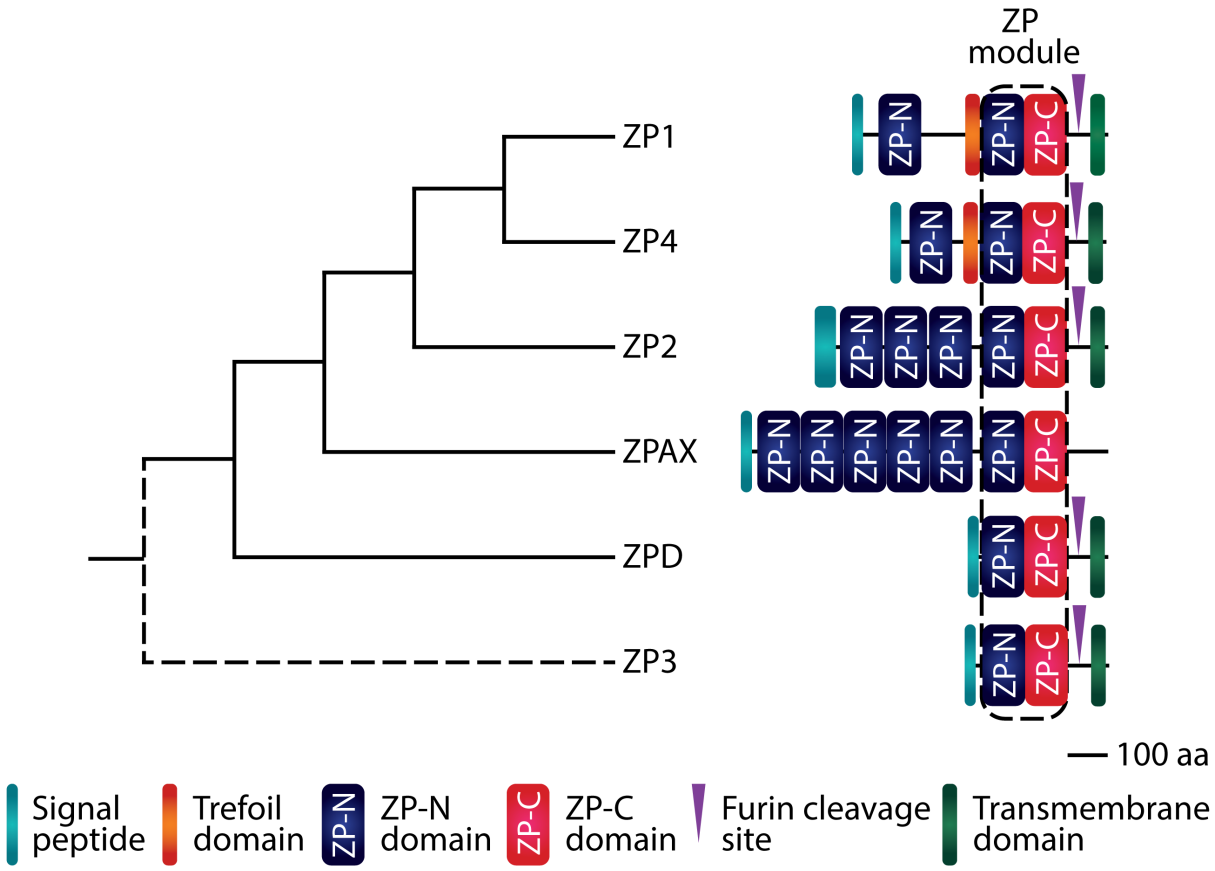

Figure 1. Phylogeny and domain structure of zona pellucida (ZP) glycoproteins. ZP3 is thought to be the ancestral ZP gene, but its position in the tree is unknown (as indicated by the dashed line). Abbreviation: aa, amino acid. Schematics for ZP1, ZP2, ZP3, and ZP4 are based on the human homologs, and ZPD and ZPAX are based on the homologs from *Xenopus tropicalis*. Adapted from (6, 19, 32, 33).

Vertebrate taxa differ in the number and type of ZP proteins incorporated in their egg coat matrix, with anywhere from zero to many copies of each ZP subfamily represented in each lineage (4). ZP3, however, is the only universal ZP gene in vertebrates, suggesting it may be the ancestral gene to all other ZP gene families, in agreement with its more minimal architecture (see Figure 1) (4, 6, 21, 34). ZP3, it has been proposed, duplicated several times hundreds of millions of years ago, giving rise to 3-4 ZP genes in fish (ZP1, ZP3, ZPAX, variants of ZP1 and ZP3), 4-5 ZP genes in amphibians (ZP2-4, ZPD, ZPAX), 6 ZP genes in birds (ZP1-4, ZPD, ZPAX), and 3-4 ZP genes (ZP1-3, sometimes ZP4) in mammals (Figure 2) (34). ZP4 shares a common ancestral gene with ZP1, and is present in rats as well as humans and other primates but is pseudogenized in the mouse genome (3, 6, 9, 34, 35). In summary, ZP1-4 are found in mammals and other vertebrates, ZPD only in amphibians and birds, and ZPAX only in fish, amphibians, and birds (34).

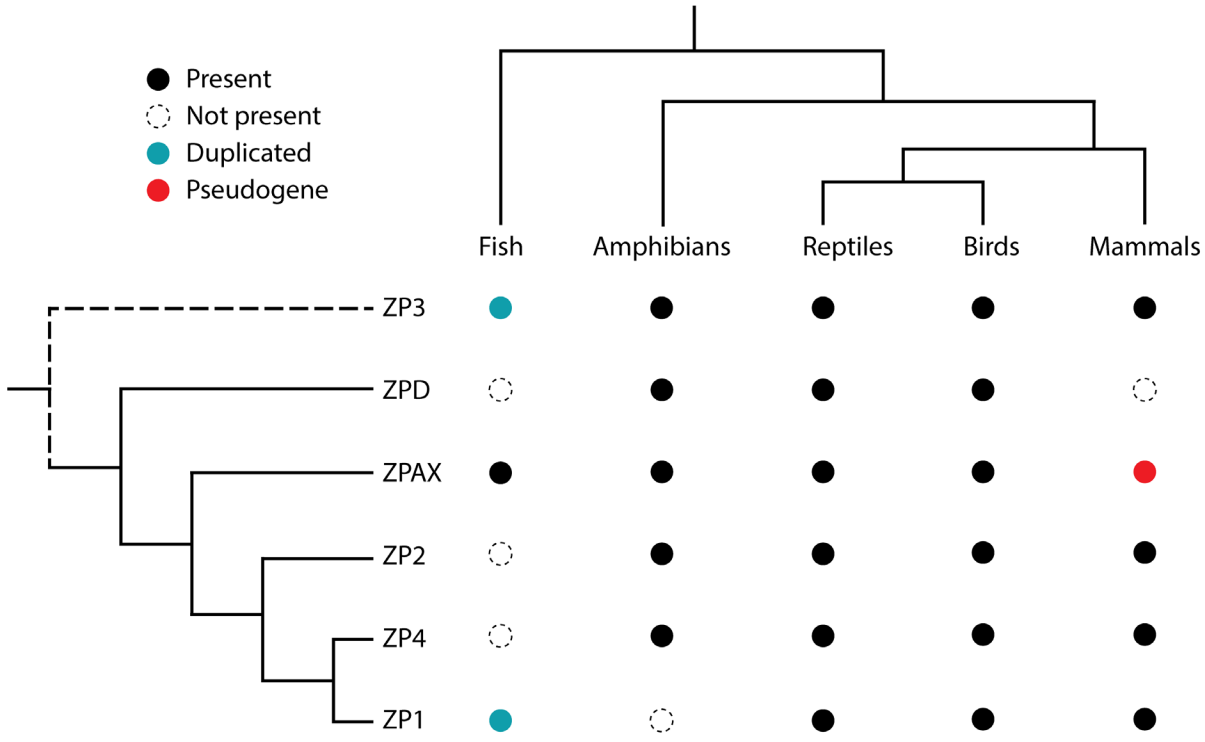

Figure 2. Patterns of ZP gene duplication and loss are highly variable among the major vertebrate lineages; the fish clade comprises teleost fish. Black circle: gene present in genome; dashed circle: gene not present in genome; teal circle: gene duplicated in genome; red circle: gene pseudogenized in genome (6, 18). Adapted from (2, 4, 6, 31).

The ZP domain, despite being named for its abundance in mammalian egg coats, is not found solely in reproductive proteins (11, 12, 21). ZP subfamilies share an ancestral gene with the CUZD1/DMBT1 gene subfamily (CUB and zona pellucida-like domains 1/Deleted in Malignant Brain Tumors 1), which includes proteins that incorporate two domains, the CUB domain and the ZP domain (6). The CUB domain is found almost exclusively in extracellular and plasma membrane-associated proteins, many of which are involved in developmental processes such as

embryogenesis and organogenesis (36, 37). The CUB domain and ZP domain are also present in two families of proteins involved in sperm-egg recognition: the CUB domain in male spermadhesins, and the ZP domain in female ZP proteins (6, 9, 38). CUZD1/DMBT1 proteins are known to be expressed in the female reproductive tract, consistent with a role in fertilization (6).

1.2.1 *ZP gene losses among vertebrates*

Given the diversity of ZP pseudogenes across the vertebrate phylogeny, ZP genes are thought to have evolved mainly through lineage-specific gene losses (6, 18, 25). For instance, the presence of both ZP1 and ZP4 in chicken, rat, chimpanzee, and human implies that the gene duplication that permitted the divergence of the ZP1 and ZP4 occurred early in the vertebrate lineage, before the separation of birds and mammals (~310 Mya), but after the divergence of fish (3, 6). ZP4 is a pseudogene in mouse, indicating that the loss of ZP4 occurred after the divergence of mouse and rat (6). In mammals, only primates and rodents have a ZP1 gene, although ZP1 is present as a pseudogene in cow and dog, suggesting that the death of ZP1 in those species happened after the divergence between primate and rodent groups and other mammals (6). The persistence of both ZP1 and ZP4 across the higher vertebrates suggests that there is a functional importance to these paralogs, as both have been retained (3).

After the divergence of birds and mammals, the ZPAX and ZPD genes seem to have been lost in mammals but not in birds (6). Loss of ZPAX in mammals is predicted to have occurred before the divergence of humans and monkeys, as similar mutations were observed in human and chimpanzee ZPAX pseudogenes (6).

In fish, the phylogeny of ZP genes is less well known due to both genome and gene duplications, particularly of ZP3: for instance, there are four copies of ZP3 in *Oryzias latipes* and three in *Danio rerio* (6, 18, 30, 39).

The persistence of the ZP2 and ZP3 subfamilies across vertebrate lineages suggests that both genes are functionally significant (6). In fact, the egg coats of all mammals contain ZP2 and ZP3 proteins, along with one or both of the ZP1 and ZP4 proteins (6). These findings may imply that sperm-egg interactions in mammals requires the presence of ZP2 and ZP3 as well as one or both of ZP1 and ZP4 (6).

It has been proposed that the pervasive loss of ZP genes in mammals could be due to taxon-specific differences in selective environments (2, 6). ZP genes may be lost in mammals because they no longer play a role in egg coat matrix formation or sperm-egg interactions (4, 6, 18). However, another possibility is that in animals with internal fertilization, embryos no longer develop in external environments and are thus no longer subject to the ecological aspects of natural selection that embryos of animals with external fertilization and/or external development (e.g. fish, amphibians, birds) are subject to (2, 4, 6). This loss of selection on embryonic performance could result in the genes encoding additional structures or functions of egg coats being pseudogenized (2, 4, 6).

1.2.2 *Structure of ZP proteins*

ZP proteins share a common structural organization with four main features: 1) a N-terminal secretory signal peptide (SP) that marks them as secreted proteins, 2) the ZP domain, a conserved sequence of ~260 amino acids including 8 or 10 invariant cysteine residues that adopt two alternative disulfide bond connectivities, 3) a recognition site for members of the proprotein convertase family of proteolytic enzymes called a consensus furin cleavage site (CFCS), and 4) a C-terminal propeptide (CTP) that includes a single-spanning transmembrane (TM) domain (9, 11, 22). These elements play crucial roles in the secretion and assembly of ZP subunits (9).

1.3 THE ZP MODULE

All animal egg coat proteins share a common molecular basis, the ZP domain (2, 4, 11, 34). Egg coat subunits polymerize using this conserved structural motif, suggesting that the basic architecture of animal egg coats has been conserved over hundreds of millions of years of evolution (7, 9, 10, 12).

The ZP domain, the structural element that gives ZP proteins their name, was first identified in ZP2 and ZP3 by pattern-based sequence analysis (34, 40). ZP domains are conserved protein-protein interaction modules comprised of two related immunoglobulin-like domains, ZP-N and ZP-C, that each contain characteristic disulfide bonding patterns (10, 41). ZP-N (~120 amino acids) and ZP-C (~130 amino acids) both have four conserved cysteine residues present as intramolecular disulfide bonds (11, 22, 35, 42).

Biochemical data indicates that only ZP-N is required for protein polymerization (41), and many ZP proteins contain tandem arrays of ZP-N repeats that have evolved independently of each other and from their associated ZP-C motifs (22, 33, 35, 43). While the combined ZP-N/ZP-C pair have classically been referred to as the "ZP domain," ZP-N should be considered a domain of its own independent of ZP-C, and we will use the term "ZP module" as put forth by Bokhove et al. to refer to the combined ZP-N/ZP-C unit (Figure 3; see also Figure 1) (10, 22, 33, 44).

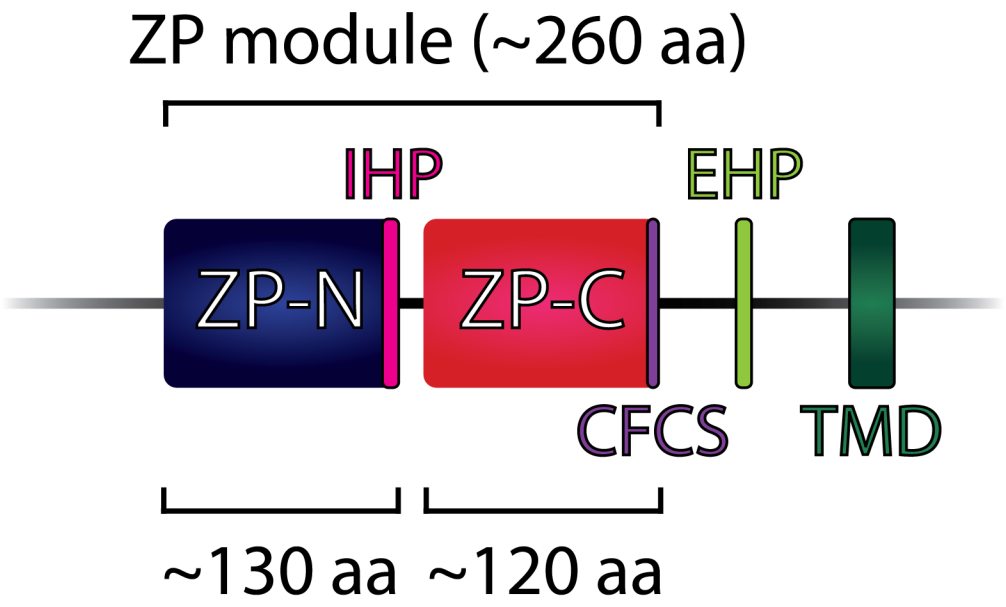

Figure 3. Schematic of the ZP module, which comprises adjacent ZP-N and ZP-C structural domains. The location of other common ZP protein structural features are indicated on the schematic, including the consensus furin cleavage site (CFCS), the transmembrane domain (TMD) if present, the external hydrophobic patch (EHP), and the internal hydrophobic patch (IHP). The EHP and IHP are involved in ZP protein polymerization. Abbreviation: aa, amino acid. Adapted from (11, 20).

The ZP module consists of two β sheets whose strands enclose a hydrophobic core comprising 8 or 10 cysteine residues. In Type I ZP modules the hydrophobic core contains 8 invariant cysteines, a structure homologous to ZP3. In Type II ZP modules, the hydrophobic core contains 10 invariant cysteines and is homologous to ZP1/ZP2/ZP4 (10, 11, 19, 40).

Additional ZP-N domains are present in single or multiple copies at the N-terminus of ZP1, ZP2, ZP4, and ZPAX (see Figure 1) (6, 11, 22, 33, 34, 41, 45, 46). After the crystal structure of mouse ZP3 ZP-N was solved, the sequences of the other ZP proteins N-terminal to the ZP module were threaded through this three-dimensional structure (45). By this analysis, it was determined that the N-terminal regions of ZP1 and ZP4 each contain an additional copy of the ZP-N domain, and the N-terminal region of ZP2 has three additional ZP-N domain repeats in tandem, connected by short linkers (10, 45). ZPAX, too, has several additional ZP-N repeats in its N-terminus – for instance, there are five additional copies in *Xenopus tropicalis* (6, 33).

1.3.1 *Other roles for ZP module-containing proteins*

The presence of the ZP module is not limited to egg coat proteins, and is found in hundreds of extracellular proteins with diverse functions in mammals, amphibians, birds, fish, flies, worms, mollusks, and urochordates (11, 12, 21). Proteins containing ZP modules are structural components of animal tissues, serve as receptors, mechanotransducers, and antimicrobials, and are involved in cell signaling, differentiation, and morphogenesis (11, 12, 34, 40, 47-50). ZP module-containing proteins organize and shape highly specialized apical structures in epithelial cells and are involved in the functioning of taste and smell (34). Examples of ZP module-containing proteins include TGF-β receptor III (betaglycan), uromodulin, tectorin-α and -β, endoglin, vomeroglandin, hensin, cuticlins, oikosins, and mesoglein (21, 34).

The presence of ZP modules in hundreds of polymeric extracellular proteins in eukaryotes, from jellyfish to humans, suggests that the structure has been conserved through at least 600 million years of evolution (9, 34). In fact, a *Saccharomyces cerevisiae* mating protein called α-agglutinin/Sag1p Ig III adopts a three-dimensional fold similar to ZP-N, so it is possible that the ZP module has been conserved for closer to 1 billion years of evolution (34, 43).

Mutations in ZP module-containing proteins cause severe human pathologies, including deafness, vascular disease, renal disease, cancer, and potentially infertility (34).

1.3.2 *Not all egg coat proteins are ZP proteins*

While ZP proteins appear to be the core building blocks of egg coats in vertebrates, the genes encoding egg coat proteins in invertebrates are not as conserved across taxa (2, 4). Although the egg coats of some marine invertebrates do contain ZP modules, several different egg coat genes are found in other invertebrates. Examples include EBR1 and rendezvin in sea urchins (2, 17); OBi1 in sea stars (51); chorion genes in *Drosophila* (52, 53), silk moths (53), other lepidopterans (54), and mosquitoes (55); and the Brownie and Citrus genes in the cockroach *Blattella germanica* (56, 57). These genes function as structural components, facilitate sperm-egg interactions, and protect embryos (4).

1.4 SYNTHESIS AND POLYMERIZATION OF ZP PROTEINS

As has been previously noted, the basic molecular structure of the egg coat has been conserved through hundreds of millions of years of evolution. For instance, recombinant mouse egg coat subunits can incorporate into the egg coats of *Xenopus* oocytes due to their common polymerization domain, the ZP module (9, 58). How do ZP proteins polymerize to form the extracellular matrix of the egg coat?

After cleavage of the signal peptide, ZP protein precursors are transported through the endoplasmic reticulum (ER) and the Golgi, remaining bound to the membrane of these organelles by their transmembrane (TM) domain (9). In the ER/Golgi the ZP proteins form disulfide bonds and are modified with *N*- and *O*-linked oligosaccharides (9). The membrane-anchored proteins are then packaged into large vesicles (~2 μm in diameter), which fuse with the plasma membrane of

the oocyte (9). After membrane fusion, or potentially prior to fusion within the *trans*-Golgi, ZP precursors are cleaved at their CFCS (9). This C-terminal processing is dependent on the TM domain, and releases mature ZP proteins into the perivitelline space where they incorporate into the innermost layer of the growing egg coat via their ZP module (9).

ZP modules have been shown to interact with each other directly in the polymerization of ZP proteins (46). The protofilaments formed by ZP proteins are organized in a right-handed double helix with frequent branching, creating a reticular network (2). ZP proteins can interact heterospecifically, permitting a diverse assembly of proteins within the reticular network of protofilaments (2). For instance, both urinary and cochlear ZP proteins can incorporate into the mouse egg coat if the whole ZP module and adjacent C-terminus are intact (46). The auto-aggregation and polymerization of ZP proteins is advantageous to extracellular matrix formation, as additional motifs associated with the ZP module can be incorporated without interfering with matrix assembly (2).

Other structural features of ZP proteins appear to relate to the specific functions of each subunit: cross-links between ZP filaments are thought to be established by the N-terminal region of ZP1, and regions N- and C-terminal to the ZP module of ZP2 and ZP3, respectively, are thought to be involved in sperm-egg interaction (9, 10).

ZP proteins vary across species in their sites of synthesis. In mammals and amphibians, ZP proteins are synthesized solely in the ovary by oocytes and/or follicle cells, whereas in fish and birds ZP proteins are synthesized in the ovary and/or liver in response to estrogen and transported via the bloodstream to the ovary, where they self-assemble around eggs (11, 21, 34, 39). Consequently, TM domains are not present in fish ZP proteins synthesized by the liver (2, 59, 60). The timing of ZP gene expression in teleosts likely coincides with vitellogenesis, such that soluble

ZP proteins lacking a TM domain can be transported to the ovarian follicles along with vitellogenic (egg yolk) proteins, limiting ZP protein precipitation in circulation and ensuring the movement of proteins essential to oogenesis (61-64).

In terms of the sites of ZP protein synthesis, it has been proposed that simpler egg coats may contain only oocyte-derived proteins, whereas more elaborate egg coat matrices may require additional contributions from somatic tissue (2). More complex egg coats are often associated with mechanically protective roles, such as resistance to environmental hazards like osmotic shock and desiccation (2). In support of this hypothesis, in animals whose egg coats are very robust, such as fish and birds, ZP proteins are often synthesized in the liver (2, 60, 65-67).

Proteins that assemble in the extracellular space must have mechanisms to avoid premature association within the cell as they are synthesized (9). ZP protein precursors are stabilized in a soluble, nonpolymerization competent conformation by two short, conserved motifs: an external hydrophobic patch (EHP) in the C-terminus between the CFCS and the TM domain, if present, and an internal hydrophobic patch (IHP) between the ZP-N and ZP-C of the ZP module (see Figure 3) (9, 23). Cleavage of ZP precursors at the CFCS, an event required for secretion of mammalian ZP proteins and incorporation of both fish and mammalian subunits into the inner layer of the growing egg coat, dissociates mature polypeptides from the EHP and activates them for polymerization (23, 59, 68, 69). Thus, despite fish and mammalian egg coat proteins differing in their sites of synthesis and C-terminal architecture, they share a common assembly mechanism (9). Because this mechanism relies on highly conserved elements such as the presence of the EHP/IHP and cleavage at the CFCS, it is likely common to all ZP proteins (9, 11, 23).

Notably, in experiments where ZP proteins are truncated just upstream of the TM domain, proteins lacking a TM domain are secreted normally but are not cleaved at the CFCS or

incorporated into the egg coat (23, 46). This suggests that TM domains in ZP proteins are not involved in specific interactions, but help with proper localization and/or topological orientation of nascent proteins for proteolytic processing and assembly (12). Whereas mammalian ZP precursors lacking a TM domain are not assembled into the egg coat, ZP precursors from fish or birds that lack a TM domain endogenously undergo cleavage at the CFCS and assemble into the egg coat normally upon reaching the ovary (59, 70).

Ultrastructural analyses of egg coats suggests that they consist of filaments of similar dimensions across organisms, but how egg coat subunits are organized into these polymers is less clear (9, 11, 71, 72). Biochemical, electron microscopy, and gene knockout studies are most consistent with a model in which the filaments are a linear repetition of ZP2/ZP3 heterodimers, with the interface between ZP2 and ZP3 running perpendicular to the axis of the filaments (73-75). Filament formation is therefore dependent on the interaction between Type I (ZP3) and Type II (ZP1/ZP2/ZP4) ZP modules (9).

ZP1 is thought to be responsible for cross-linking these filaments into a three-dimensional matrix (9, 76, 77). In the mammalian egg coat, ZP1 is expressed at much lower levels than the other subunits, so ZP1 (and ZP4, if present) would be incorporated only rarely in place of ZP2 (9). By contrast, in fish, the presence of two or more ZP1 homologs (at least one of which is highly expressed (78)), as well as the lack of a ZP2 homolog, predict an egg coat with a much higher number of cross-links (9). This is in keeping with the significant resistance to mechanical and chemical stress that fish egg coats display, even prior to hardening (9). Thus the composition of egg coats in terms of the number of ZP1 homologs they contain may suggest physical properties, by giving an estimate of the number of cross-links (9, 11). Notably, the formation of additional

intra- and intermolecular disulfide bonds has been implicated in the hardening of the mammalian egg coat (9, 79).

1.5 EGG COAT STRUCTURE

1.5.1 *Mammals*

Mouse egg coats are the best understood of all vertebrates, and the most well-studied of mammals specifically (4, 9). Mouse egg coats are comprised of homologs of ZP1 (~200 kDa), ZP2 (~120 kDa), and ZP3 (~83 kDa) (6, 12, 20). These proteins assemble into ~2-3 μm long filaments, with a structural repeat of ~150 Å, and are cross-linked into a highly porous, 6.5 μm thick elastic network (9). The egg coats of other mammals, including humans, are thought to have a similar structure with the inclusion of an additional ZP1-like subunit, ZP4 (6, 10).

By electron microscopy, it was suggested that mouse ZP2 and ZP3 assemble into micron-long polymers which are cross-linked into a three-dimensional matrix by disulfide-bonded ZP1 homodimers (10, 74, 77). This model is confirmed by the phenotypes of mice lacking the genes for the individual ZP subunits (10). Homozygous ZP1 knockout mice produce an egg coat, but it is loose and insufficiently cross-linked and mutant females are less fertile than wild-type, suggesting that ZP1 interconnects ZP fibrils and that a structurally intact egg coat is integral to fertilization (19, 76, 80). Homozygous knockout female mice lacking either ZP2 or ZP3 fail to construct an egg coat and are infertile, indicating that ZP2 and ZP3 depend on each other for incorporation into the egg coat (34, 80-83). Mutant female mice with a single ZP3 allele assemble an egg coat and reproduce normally, but their egg coat is less than half the thickness of wild-type (34, 80, 84).

1.5.2 *Birds*

The egg coats of birds consist of two layers separated by a thin continuous membrane, with the inner or perivitelline layer representing a 1-3.5 μm thick network of fibers analogous to the mammalian egg coat as it mediates the species-specific binding of sperm (9). After fertilization, oocytes acquire the continuous membrane (0.1-0.5 μm thick) and the outer layer (3-8 μm thick), which are thought to be involved in blocks to polyspermy (9, 85).

As with the mammalian egg coat, the avian perivitelline layer contains several glycoproteins: homologs to ZP1, ZP2, ZP3, ZP4, and ZPD have been found in quail, and genes for ZP1, ZP2, ZP3, ZP4, ZPD, and ZPAX are present in the chicken genome (6, 9, 18, 30, 85, 86).

Avian ZP3 is ~32-42 kDa, and avian ZP1 exists as both a ~97 kDa monomer and a homodimer held together by intermolecular disulfide bonds (9). ZP1 is secreted by the liver in response to estrogens, and is characterized by a proline/glutamine-rich repeat region N-terminal to the trefoil domain and a short CTP lacking a TM domain, as would be predicted by its liver synthesis (9, 65, 87). Chicken ZPD (~42 kDa) loosely associates with the perivitelline layer. Together with dimeric ZP1, ZPD has been implicated in sperm activation (9, 66, 85). ZP3 is thought to be responsible for sperm binding in the perivitelline layer of both chicken and quail (85). Upon binding of sperm, avian ZP1 is degraded and hole-like structures appear in the perivitelline layer (9, 88).

1.5.3 *Amphibians*

Amphibians include anurans (frogs), which reproduce by external fertilization in water, and urodeles (salamanders), which reproduce by internal fertilization in the female cloaca (9, 89). The ~1 μm thick egg coat of *Xenopus laevis* is the best studied of the anurans, and consists of five

major ZP glycoproteins that are synthesized by the oocyte: ZP2 (gp69/64), ZP3 (gp43/41), ZP4 (gp37), ZPAX (gp120/112), and ZPD (gp80) (6, 9, 90-96).

Xenopus ZPAX contains a ~600 amino acid N-terminal region and a short CTP lacking a predicted TM domain, and is related to ZP2 with its Type II ZP module and absence of a trefoil domain (9). In agreement with homology to ZP2, *Xenopus tropicalis* ZPAX has five additional ZP-N repeats in its N-terminus (33).

Xenopus ZPD has a simple architecture, consisting of a SP, a ZP module, a CFCS, and a TM domain, and appears to represent a subfamily of its own: although its ZP module sequence is most similar to a Type II ZP module, it lacks two of the 10 conserved cysteines (9, 93).

Fertilization in urodels has not been as well studied, with the exception of the newt *Cynops pyrrhogaster*, whose eggs can undergo polyspermy since no fertilization envelope forms after fertilization (89). A transcriptome assembly from ovary, testis, and oviduct found homologs to all six ZP subfamilies – ZP1, ZP2, ZP3, ZP4, ZPAX, and ZPD – expressed in the ovaries of *C. pyrrhogaster* (97). Notably, the authors found six distinct paralogs of ZP3 (97).

1.5.4 *Teleost fish*

Teleost fish are highly diverse, constituting almost half of the total number of vertebrates (9). This diversity is reflected in the architecture of teleost egg coats, which vary in thickness, structure, and number of layers both between and within species (9). A single layer of follicle cells surrounds oocytes as they grow (9). In response to external signals, follicle cells produce 17β-estradiol to induce synthesis of both egg yolk (vitellogenin) and ZP protein precursors (9). In most teleosts, soluble ZP protein precursors are secreted by hepatocytes and travel in the blood to the oocyte for incorporation into the egg coat (98, 99). ZP precursors are deposited in the perivitelline space, at the base of long microvilli that stretch from the plasma membrane of the oocyte to the

follicle cells (9). Egg coat assembly proceeds as the oocyte accumulates egg yolk, resulting in a radially striated structure of helicoidal glycoprotein bundles separated by extended microvilli (9).

The egg coats of rainbow trout (*Oncorhynchus mykiss*) have a thin outer layer and a ~50 μm thick inner layer of three major subunits: VEα (~58 kDa), VEβ (~52 kDa), and VEγ (~47 kDa) (9, 12, 100). Like avian ZP1, these subunits have a N-terminal proline/glutamine-rich repeat region and a short CTP lacking a predicted TM domain (9, 78). VEα and VEβ are very similar in sequence and contain a trefoil domain immediately before a Type II ZP module, suggesting homology to mammalian ZP1/ZP2/ZP4, whereas VEγ contains a sperm combining site-like sequence C-terminal to its Type I ZP module, suggesting homology to ZP3 (9, 12, 78, 101). Rainbow trout egg coats consist of VEα/γ and VEβ/γ heterodimers (12).

The egg coats of other teleost fish have similar compositions to rainbow trout, although in species such as carp, goldfish, and zebrafish, ZP genes are synthesized in the ovaries and thus contain a predicted TM domain in their CTP (12, 102, 103). In medaka, ZP subunits are synthesized by both the liver and oocytes (104). Additionally, ZP3 genes are often duplicated within teleosts: for instance, there are four ZP3 genes in medaka and three in zebrafish (6).

No ZP2 orthologs have been identified in teleosts, despite the classification of some fish egg coat genes as ZP2 (105). These genes are in fact ZP1 homologs, as evidenced by the presence of a trefoil domain in these proteins (see Figure 1) (105, 106). In agreement with the absence of the ZP2 subfamily in teleosts, there are no N-terminal ZP-N repeat regions in teleost ZP proteins, except for some homologs of ZPAX (90).

Teleost sperm lack an acrosome, a secretory vesicle that facilitates sperm-egg contact and egg coat dissolution. Instead, sperm reach the egg plasma membrane through the micropyle, a funnel-shaped, narrow channel through the egg coat (2, 9, 12, 107, 108). The micropyle attracts

sperm by chemotaxis, and its precise diameter prevents polyspermy by permitting one sperm to pass at a time (5, 109, 110). In contrast to mammals, birds, amphibians, mollusks, echinoderms, and urochordates, teleost sperm do not need to dissolve the egg coat to reach the egg and fuse with its plasma membrane (5, 111). Therefore, it has been argued that fish egg coat proteins play a purely structural role in fertilization (9, 12, 111).

The structure of the micropyle has convergently evolved in at least two animal orders with different modes of reproduction, the dipterans and the teleosts (2). In fish the micropyle is formed after the deposition of the egg coat, by the retraction of the cytoplasmic process of a specialized follicle cell called the micropylar cell that extends to the oocyte surface (5, 111). In zebrafish it has been shown that mutants in the gene bucky ball have an excessive number of follicle cells that develop as micropylar cells, leading to multiple functional micropyles and polyspermic fertilization (112). Micropyle architecture differs across teleost species, from simple channels traversing the egg coat to more elaborate structures with outer sperm catchment areas that funnel sperm into the micropyle (109, 111, 113-115).

After fertilization, teleost egg coats undergo a cortical reaction and secrete a transglutaminase, which hardens the egg coat by introducing cross-links between egg coat subunits, likely via their proline/glutamine-rich repeat regions (9, 99). The hardened egg coat has a different morphology, and protects the developing embryo against environmental hazards and pathogens. Notably, mammalian ZP proteins lack an N-terminal proline/glutamine-rich repeat region, and covalent linkages between ZP subunits have not been detected in mammalian eggs or embryos (12, 75).

1.5.5 *Mollusks*

Marine invertebrates are some of the first model systems in the study of fertilization, facilitated by their numerous, accessible gametes as an externally fertilizing group (14). Much of the work has centered on the gastropod abalone (genus *Haliotis*). The abalone egg coat is known to mediate species-specificity in gamete interactions, and triggers the sperm acrosome reaction (9, 116). This exocytic event releases lysin, a dimeric 16 kDa protein that binds to the vitelline envelope receptor for lysin (VERL), a highly repetitive ~1 MDa egg coat glycoprotein, ~90% of which is 22 repeats of ~153 residues homologous to the vertebrate ZP-N polymerization domain (35, 117, 118). These ZP-N repeats are thought to generate an interlocking network of hydrogen bonds, stabilizing the egg coat supramolecular structure (35). The VERL protein consists of a SP, the array of 22 tandem ZP-N repeats, a Type II ZP module, a CFCS, and a TM domain (9, 118).

Lysin dissolves the abalone egg coat through nonenzymatic means, likely competing for hydrogen bonds between VERL repeats and splaying the fibers of the egg coat, allowing sperm to pass (35, 117). Lysin dissolves conspecific egg coats faster than heterospecific egg coats, suggesting species-specificity in its activity (35, 117). In keeping with this pattern of species-specificity, the sequence of lysin homologs from different species of abalone are extremely divergent as a result of strong adaptive evolution (9, 119). VERL ZP-N repeats 1 and 2 are also rapidly evolving, whereas repeats 3 to 22 evolve neutrally and are homogenized (>98% identical within a given species) by unequal crossing over and concerted evolution (9, 35, 120-122). Lysin and VERL experience correlated rates of evolution, and display intergenic linkage disequilibrium despite not being physically linked (123). This coevolution between a male protein and a female receptor suggests that identification of coevolving proteins in the two sexes could potentially find other interacting fertilization proteins – for instance, mammalian ZP interactors (123, 124).

There are other ZP module-containing proteins in the abalone egg coat in addition to VERL; in fact, at least 30 additional ZP proteins were identified in abalone egg coats by expressed sequence tag sequencing and shotgun proteomics (25). These additional ZP proteins included a paralog of VERL called VEZP14, which also contains a putative lysin-binding motif and is rapidly evolving (25).

The gastropod mollusk *Tegula*, a genus of free-spawning marine snails, also possesses an ortholog to VERL that binds *Tegula* lysin (125). Whereas abalone VERL has 22 tandem ZP-N repeats, *Tegula* VERL contains only one (125). Of note, however, is the presence of a homogenized array of three to four residues, mainly serine-proline-threonine (SPT) or serine-proline-threonine-threonine (SPTT), that is repeated 70 times between *Tegula* VERL's ZP-N repeat and its ZP module (125).

In the basal mollusk chiton, follicle cells surround a gelatinous protective layer around the egg coat called the jelly hull (2, 21, 126). These follicle cells shrink on contact with sea water, leaving behind pores in the hull that are almost continuous with pores in the egg coat (2, 126-128). These tunnels focus sperm and facilitate access to the egg plasma membrane, somewhat akin to micropyles (2, 129).

In the bivalve *Unio*, a freshwater mussel, the egg coat is attached to the egg plasma membrane solely at the vegetal pole (130). As the only site of sperm binding and fusion, this region is the functional equivalent of a micropyle (2). The attachment point is a crater on the egg coat comprised exclusively of the sperm-receptive gp273 glycoprotein, marking the only fusogenic region of the egg; the remainder of the egg coat consists of the structural glycoprotein gp180 (131).

1.5.6 *Sea urchin*

In sea urchin, the egg coat is estimated to contain 25 major glycoproteins (132-134). Two of these have been shown to play a structural role: p160 and rendezvin^{VL} (2). p160 is a 160 kDa protein composed of five CUB domains and a putative transmembrane domain, and is predicted to be an integral membrane protein with protein-protein interaction motifs facing the extracellular matrix of the egg (135). p160 is found at the tips of microvilli, suggesting that it links the egg plasma membrane to the egg coat until fertilization, at which point it is cleaved to separate the two and establish one block to polyspermy (135).

Rendezvin^{VL} is a splice variant of the oocyte-specific rendezvin gene that is trafficked to the vitelline layer, where it reunites at fertilization with another rendezvin splice variant that associates with cortical granules: specialized organelles that are exocytosed into the perivitelline space between the egg coat and the egg that participate in the slow block to polyspermy (19, 136). The two splice variants create the fertilization envelope – modifications to the egg coat that form the block to polyspermy in sea urchins – likely via heterologous interactions between their CUB domains (136).

Analogous to lysin and VERL in the abalone and *Tegula* systems, sea urchins encode a pair of interacting fertilization proteins called bindin and EBR1 (35). Bindin is the sole protein in sea urchin sperm acrosomes, and mediates the species-specific binding of sperm to the egg coat via its receptor, EBR1 (137, 138). EBR1 has a similar architecture to VERL in that it contains ~19 tandem egg bindin receptor (EBR) repeats consisting of alternating CUB and thrombospondin type 1 (TSP-1) domains in *Strongylocentrotus franciscanus*, with the last 10 EBR repeats being highly conserved and species-specific (138). *Strongylocentrotus purpuratus* EBR1 has a slightly modified architecture of ~8 EBR repeats that share 88% identity with the *S. franciscanus* EBR

repeat, but an entirely differently species-specific domain consisting of 11 hyalin-like (HYR) repeats (138). These species-specific domains in *S. purpuratus* and *S. franciscanus* EBR1 are known to function in species-specific sperm adhesion (138).

1.5.7 *Insects*

The egg coats of flies (dipterans) are composed of two layers: an inner vitelline layer and an outer chorion layer (139-142). The chorion is synthesized by the surrounding follicle cells, and protects the egg from desiccation, mechanical stress, and pathogens after it is laid (2, 53).

Dipteran eggs are fertilized internally as the egg travels down the oviduct (2). Sperm are released in a controlled manner from the spermatheca, so the presence of a micropyle in dipterans, as previously noted, is somewhat unexpected (139-141, 143). It is likely that the importance of the egg coat in preventing desiccation takes precedence, making the micropyle less of a structure whose primary role is to block polyspermy and more one that supports gamete interactions while allowing gas exchange during embryogenesis (2, 144).

The *Drosophila* chorion is composed of ~20 structural proteins synthesized by ovarian follicle cells, and contains six major chorion proteins (cp): cp15, cp16, cp18, cp19, cp36, and cp38 (53, 145). cp36 and cp38 are expressed early in chorion formation (145), and cp15, cp16, cp18, and cp19 are expressed late (146-148). These chorion proteins are stabilized in the final egg coat structure by peroxidase-mediated tyrosine cross-links (149).

Several additional genes in *Drosophila* are required for the assembly of the chorion, including fs(2)QJ42 and defective chorion 1 (dec-1) (142). The protein products of these genes are found throughout the egg coat rather than localized at the micropyle, suggesting that they play a role in structural integrity rather than sperm-egg interactions; this hypothesis is supported by the loss-of-function phenotypes associated with these loci (2, 142). Females homozygous for

fs(2)QJ42 produce chorions with altered permeability properties, and females who fail to synthesize dec-1 display morphological abnormalities in the chorion layers (142).

In silk moths, such as the cultivated *Bombyx mori* and the wild oak silk moth *Antheraea polyphemus*, the chorion contains more than 100 distinct components by two-dimensional gel electrophoresis (150). Silk moths possess an additional egg coat structure called the lamellar chorion, which in *Bombyx* contains unusual cysteine-rich proteins termed high-cysteine (Hc) proteins whose presence is unique to *Bombyx* (151, 152). These Hc proteins give the egg coat enhanced hardness and reduced permeability, likely as an adaptation to prolonged winter diapause (152). Cysteine-rich proteins have also been found in the chorions of the mosquitoes *Aedes aegypti* and *Anopheles gambiae* (55, 153).

In the cockroach *Blattella germanica*, two highly abundant chorion genes, Brownie and Citrus, have been recently described (56, 57). Brownie is expressed in follicle cells, forming a structure called the sponge-like body in a cavity left behind by those cells in late choriogenesis, when Brownie expression is at its highest (56). The sponge-like body is a complex structure that combines the micropyle and the aeropyle, a feature of insect eggs that functions in gas exchange (56).

Citrus is involved in chorion formation, as females treated with RNAi to Citrus laid fragile eggs showing discontinuous deposition of chorion proteins, resulting in increased permeability (57). Citrus has a signal peptide and is rich in glycine, tyrosine, and proline, as with other insect chorion proteins where these abundances serve structural roles (such as tyrosine cross-linking) (57). The protein contains 33 repeats of the same motif: 30-40 residues in length, rich in glutamic acid at the N-terminus and glycine, tyrosine, and proline repetitions at the C-terminus (57). The first four N-terminal repeats are the least conserved, and database searches found no homologous

proteins or motifs (57). Both Brownie and Citrus show an absence of cysteines and a high concentration of tyrosines, suggesting they could be cross-linked as has been described in the chorions of dipterans (57, 149).

1.5.8 *Cephalochordates and urochordates*

Cephalochordates, urochordates, and vertebrates represent the three extant groups of chordate animals (154). Ascidians (sea squirts) are mostly hermaphroditic urochordates that produce self-sterile gametes due to the presence of a self-incompatibility system (9, 155). The ascidian egg coat plays a role in species-specific binding of sperm and egg, participates in the slow block to polyspermy, and prevents self-fertilization in self-sterile species (9, 156).

Ascidian egg coats contain multiple homologs to ZP proteins, most of which possess epidermal growth factor (EGF)-like repeats (29). For instance, the major component of the egg coat in *Halocynthia roretzi* is HrVC120 (precursor to HrVC70), which consists of a SP, 13 EGF-like repeats, a Type II-like ZP module, a CFCS, and TM domain (9, 28, 157, 158). HrVC70, the product of proteolytic cleavage of HrVC120 within the last EGF repeat, contains 12 EGF repeats and shows sperm receptor activity (155, 158). Because of its sperm binding capabilities and polymorphic nature, HrVC70 has been proposed as a potential allorecognition molecule mediating self-sterility (155). In *Halocynthia aurantium*, a homolog to HrVC120 called HaVC130 shows high similarity to HrVC120 but has 14 EGF domain repeats, suggesting that the number of EGF repeats could play a role in species-specificity of sperm-egg interactions (157).

Ciona intestinalis, another urochordate, has four homologs of ZP proteins called Vc16, Vc20, Vc182, and Vc569 (158). The ZP proteins consist of a signal peptide, a ZP module, and a TM domain; both Vc182 and Vc16 have a CFCS, and all four have varying numbers of EGF repeats: six for Vc569, two for Vc20, one for Vc16, and none for Vc182 (158). All four ZP proteins

are expressed in the developing oocyte (158). A more recent proteomic characterization of *C. intestinalis* egg coats found an additional seven ZP proteins, bringing the total to 11; all seven contain single or multiple EGF-like domains (29).

ZP proteins are present in cephalochordate egg coats as well: five were identified in *Branchiostoma belcheri* by mass spectrometry, termed BbZP1-5 (26). Each BbZP has a C-terminal ZP module, and the majority contain a low-density lipoprotein receptor domain and a von Willebrand factor type A (vWFA) domain (26). However, none have the EGF-like domain frequently observed in the ZP proteins of urochordates (26). Only BbZP1 has a TM domain, and BbZP1, 3, and 4 have CFCSs (26). The five BbZPs are synthesized primarily by the developing oocytes (26).

Cephalochordate ZP proteins are evolutionary homologs of the ZP1, ZP2, and ZPAX subfamilies, whereas urochordate ZP proteins appear to be more closely related to ZP3 (26). Regardless, analyses of cephalochordate and urochordate ZP homologs suggest that vertebrate ZP proteins have an invertebrate origin, or at least arose at the base of chordate evolution (26).

1.6 ZP PROTEINS IN FERTILIZATION

In species-specific mating, gamete recognition ensures a single sperm fertilizes the egg while preventing polyspermy that can lead to embryo death. The first contact in gamete recognition is mediated by the extracellular matrix (ECM) of the egg coat surrounding ovulated eggs (159, 160). As with the ECM in other tissues, the egg coat ECM is a critical intermediary in cell-cell communication (2). Initial contact with the egg coat triggers a cascade of reactions in sperm, increasing metabolism and motility and initiating the acrosome reaction, releasing the contents of the sperm secretory vesicle into the local environment by exocytosis (161-164). In most animals, these events signify the beginning of successful gamete recognition (2).

The molecules that elicit sperm activation vary significantly among animals, often requiring overlapping receptor-ligand interactions that enhance species-specificity (2). Sperm activation results in a shift to chemotactic motility toward the egg coat and initiates the acrosome reaction, exposing additional sperm-egg binding partners in the form of membrane-associated proteins in the luminal face of the acrosome and soluble proteins released into the local environment (2, 161). Some of these soluble proteins aid sperm progression through the egg coat ECM (2). The acrosome reaction thus signifies sperm activation and the successful achievement of initial gamete contact, except in teleost fish whose sperm lack an acrosome (2, 111).

It is likely that both ZP2 and ZP3 play important roles in sperm binding. Early experiments in the 1970s demonstrated that solubilized ZP glycoproteins from unfertilized hamster and mouse egg coats could inhibit binding of hamster sperm to ovulated eggs *in vitro*, suggesting the presence of receptors in the solubilized ZPs that could bind sperm and prevent their interaction with ovulated eggs (34, 165). In mice, it was found that solubilized ZP3 inhibited sperm binding to ovulated mouse eggs *in vitro*, whereas solubilized ZP1 and ZP2 did not, suggesting that in mice the egg coat receptor for sperm is ZP3 (166). ZP3 is able to trigger the acrosome reaction, and acrosome-reacted sperm bind to ZP2, making ZP3 the primary receptor for sperm and ZP2 the secondary receptor (34, 167-171).

Over time, it was shown that *O*-glycans on serine-332 and -334 in the C-terminus of ZP3, located in a so-called sperm combining site, are essential for gamete recognition, and that their post-fertilization cleavage could account for another observation in mice: the inability of mouse sperm to bind to solubilized ZPs from two-cell embryos (159, 172, 173). However, elimination of glycosyltransferases required for the candidate *O*-glycans did not affect fertility, and mass spectrometry on native mouse egg coats found no evidence of glycosylation at the two serine

residues (159, 174, 175). Furthermore, mutating serine-332 and -334 to preclude modification with *O*-glycans did not alter sperm recognition or fertility, even if endogenous ZP3 was removed, suggesting that these sites cannot be intrinsically involved in sperm binding (176, 177).

Such observations suggest that ZP proteins other than ZP3 may be involved in sperm-egg recognition. Mice lacking ZP1 are fertile, despite reduced fecundity, ruling out an essential role for ZP1 in gamete interactions (76). ZP4 is a pseudogene in mice, and human sperm do not bind to transgenic mouse eggs expressing recombinant human ZP4 (6, 178). This leaves ZP2 as a candidate.

Although human sperm can bind to the egg coats of *Homo sapiens*, *Gorilla gorilla*, and *Hylobates lar* (gibbon), they do not bind to the egg coats of baboon and rhesus macaque or to other mammals, including mice (179, 180). Notably, however, replacing mouse ZP2 with human ZP2 is sufficient to permit human sperm to bind to the recombinant mouse egg coat (159). In this gain of function assay, the site of gamete recognition was localized to a domain of ~115 amino acids in the N-terminus of ZP2 (159).

Following fertilization, the proteolytic cleavage of ZP2 into two fragments of approximately 23 and 90 kDa, which remain disulfide bonded, prevents sperm from binding to two-cell mouse embryos (181). The cleavage of ZP2 is catalyzed by ovastacin, an oocyte-specific member of the astacin family of metalloendoproteases that is released into the perivitelline space upon fusion of the egg cortical granules with the plasma membrane (166, 182-184). Mutating the cleavage site of ZP2 or eliminating the gene encoding ovastacin leaves ZP2 intact following fertilization, allowing sperm to bind to the egg coat despite cortical granule exocytosis (177, 184).

Taken together, these observations suggest a molecular basis of gamete recognition in which sperm bind to an N-terminal domain of ZP2, followed by egg coat penetration and gamete

fusion (185). After fertilization, ovastacin is released from egg cortical granules, cleaving extracellular ZP2 and eliminating the sperm-binding domain, accounting for the inability of sperm to bind to two-cell embryos (177, 184).

While ZP2 has been historically considered a secondary egg coat receptor that binds acrosome-reacted sperm, there is precedence for ZP2 acting in primary gamete recognition in *Xenopus laevis*, where its ZP2 homolog gp69/64 inhibits sperm binding to eggs *in vitro* (186). Following fertilization, *X. laevis* ZP2 is cleaved by a zinc metalloprotease, suggesting that post-fertilization proteolysis of ZP2 could apply more generally across vertebrates as a block to polyspermy (187).

1.6.1 *ZP-N repeats in sperm binding*

In addition to their C-terminal ZP module, some ZP proteins have N-terminal extensions that are thought to be involved in species-specific gamete recognition (33). These extensions are made up of single or multiple copies of the ZP-N domain, which is related to the N-terminal region of the ZP module. The presence of these ZP-N repeats in ZP1, ZP2, ZP4, and ZPAX suggests that ZP proteins may have evolved around a common ZP-N architecture (7, 33).

With the exception of some homologs of the ZPAX subfamily, ZP-N repeats are not found in the egg coat proteins of fish, whose ZP proteins play a structural role like their counterparts in other metazoans but may not be involved in sperm binding due to the presence of the micropyle (10, 12). Further, although the ZP module is conserved in hundreds of extracellular proteins unrelated to fertilization, none of these have been shown to possess ZP-N repeats (11). These considerations suggest that ZP-N repeats could have functional roles specific to fertilization, in agreement with what is known about the ZP-N architecture and sperm binding properties of ZP2 (10, 33, 159, 171, 177, 188).

Notably, the mollusk and ascidian egg coat proteins VERL and HrVC70 both contain C-terminal ZP modules preceded by repeats that have been implicated in sperm binding (117, 118, 155, 157). While the 12 repeats of HrVC70 are epidermal growth factor (EGF)-like domains, the 22 tandem repeats of abalone VERL are known to adopt a ZP-N fold (35, 43, 155).

Furthermore, a ZP-N domain in the N-terminal region of VEZP14, another abalone egg coat protein thought to be involved in sperm binding, facilitated the discovery of a ZP-N-like fold in the Ig III domain of the yeast protein alpha-agglutinin/Sag1p by fold recognition analysis (25, 43). This Sag1p Ig III ZP-N is directly involved in haploid yeast cell interactions during mating, a process that mirrors sperm-egg interactions in higher eukaryotes (10, 189). Despite a separation of almost 1 billion years of evolution, gamete recognition in metazoans and mating in yeast appear to share common structural features, potentially mediated by the ZP-N domain (10, 43).

1.7 Evolution of egg coat proteins

The rapid evolution of reproductive proteins is a recurring observation among natural populations of animals (14, 35). Because reproductive proteins regulate an essential process – fertilization – that fundamentally impacts fitness, the rapid divergence characteristic of reproductive proteins is striking, and suggests that they may evolve under adaptive evolution (190).

The first evidence for rapid evolution of reproductive proteins came from free-spawning marine invertebrates such as abalone and sea urchins (137, 191). Whereas the evolution of gamete interactions in internally fertilizing species can be confounded by physiological and behavioral aspects of mating that may also be under selection, in taxa that release millions of gametes into the external environment for fertilization, gamete interactions can be readily observed and the targets of selection are clear (14, 17, 190, 192). External fertilization is also thought to be the ancestral

mating strategy, providing insight into how gametes generally and gametic compatibility in particular have evolved (192, 193).

Since their inception in marine invertebrates, studies on reproductive protein evolution have accumulated in diverse taxa. Evolutionary sequence analyses from insects to vertebrates have found evidence for the rapid, adaptive evolution of reproductive proteins with functionally diverse roles (14). This rapid evolution of reproductive proteins can contribute to reproductive isolation between diverging taxa, and the evolution of reproductive isolation is integral to the process of speciation (14, 18, 124, 194). For instance, rapid evolution may drive changes to amino acids important in sperm-egg interaction, creating or reinforcing reproductive barriers (18).

Rapid evolution can be tested for by comparing the frequency of nucleotide substitutions at codons within genes, usually between species (35). In the absence of selection, most nucleotide (and thus amino acid) substitutions will occur at a rate that reflects the basal mutation rate. The frequency of synonymous substitutions (d_S) provides an estimate of this rate, and under neutral evolution nonsynonymous substitutions (d_N) should occur at a similar frequency, leading to a d_N/d_S ratio of approximately 1. However, since most nonsynonymous substitutions are likely to alter protein structure and negatively impact function, nonsynonymous substitutions should occur less frequently and d_N/d_S will be less than 1. Sites where nonsynonymous substitutions are disfavored are considered under negative or purifying selection (195-198). Alternatively, in situations where rapid change may be adaptive, nonsynonymous substitutions can occur more frequently than the mutation rate and d_N/d_S will be greater than 1, reflecting positive Darwinian selection (19, 35, 196).

Numerous evolutionary forces have been implicated in the rapid evolution of reproductive proteins, including pathogen resistance, sperm competition, cryptic female choice, sexual conflict, reinforcement, and avoidance of heterospecific fertilization (14, 190, 192, 199, 200). These

mechanisms reflect the potential for selection to act on reproductive proteins at various stages throughout fertilization (14).

To accomplish fertilization, sperm and egg must encounter foreign molecules in the form of the gametes of the opposite sex (14). Gametes are additionally exposed to novel environments in the process of fertilization, whether as sperm travelling through the female reproductive tract or egg and sperm released into the surrounding environment in externally fertilizing species. These exposures render gametes vulnerable to microbial pathogens, making pathogen resistance an exogenous force that can drive the divergence of reproductive proteins (190). For instance, microbial attack may impose a constant selective pressure for diversification of gamete surface proteins, necessitating continual adaptation of reproductive proteins to this changing recognition surface (190, 201).

Interactions between gametes represents another source of selection. In species where females mate with multiple males, females may exhibit cryptic female choice in choosing between sperm of different males (14). In males, competition among sperm can lead to strong selection to rapidly penetrate and fertilize the egg (14). Females, however, favor a lower fertilization rate because the larger energy investment of their gametes means polyspermy is more detrimental to female fitness. Consequently, female gamete proteins will evolve to lower the fertilization rate, and sperm proteins will evolve to increase it (190). Such opposing optimal fertilization rates can lead to sexual conflict at the gamete level, maintaining polymorphism and leading to a coevolutionary arms race between sperm and egg proteins (14, 190, 202-204).

Heterospecific interactions are also subject to selection. For instance, reinforcement is a selective force driven by the negative consequences of hybridization: if gametes from different species meet and the resulting hybrids have reduced fitness, changes to protein interactions that

reduce heterospecific fertilization will be favored, leading to the evolution of reproductive barriers (14, 192). Species that overlap geographically may show only partial gametic incompatibility, so reinforcement is predicted to be strongest in sympatric populations, as opposed to allopatric populations where no heterospecific sperm is present (192, 205, 206).

These and other selective forces acting on reproductive proteins can be challenging to separate, and may occur simultaneously in a single species (14, 194).

1.7.1 *Reproductive proteins as species barriers*

Sperm-egg interaction is a species-specific event mediated by the recognition and binding of complementary molecules on the surface of the egg coat and on the sperm plasma membrane (160, 163, 207). The importance of the egg coat in maintaining species boundaries is evidenced by the fact that its removal permits heterospecific sperm binding to the egg plasma membrane; for instance, guinea pig, mouse, rat, human, rabbit, goat, dolphin, cattle, horse, and pig sperm can all successfully penetrate hamster oocytes lacking an egg coat (207, 208).

As the first step in gamete recognition, mutations affecting the interface between sperm and the egg coat can create barriers to fertilization that may ultimately lead to speciation – particularly because sperm and egg coat proteins from ascidians to mollusks to humans are known to diverge rapidly as a result of adaptive evolution (9, 18, 30, 121, 157, 207, 209-212). Experimental studies have demonstrated that adaptive evolution acting on functional domains within reproductive proteins is sufficient to cause reproductive isolation among closely related species, with implications for the mechanisms of speciation (14, 116, 190, 194, 213).

In support of this hypothesis, variation in gamete recognition proteins within a species influences reproductive compatibility; protein divergence predicts the likelihood of hybrid fertilization better than neutral genetic markers, and polymorphism in sperm-egg recognition

proteins determines reproductive success among individuals within a population (137, 192, 203, 214-216). Gametic compatibility, therefore, influences the effectiveness of reproductive isolation across species boundaries (192, 203, 207, 217).

Furthermore, a comparatively small number of nonsynonymous mutations can alter reproductive protein interactions, with as few as 10 amino acid changes in sea urchins leading to incompatibility (137). Nonsynonymous substitutions occur frequently at sperm-egg binding sites in many species, indicating these substitutions may influence gamete interactions (116, 125, 192, 218-220).

The species-specific reproduction of externally fertilizing marine invertebrates led to the hypothesis that divergence in gamete recognition proteins could lead to speciation (17, 194). In sea urchins, sympatric species with overlapping habitats show the highest rates of sperm bindin evolution and the highest sequence polymorphism, with strong blocks to hybrid fertilization caused by the failure of heterospecific sperm bindin to bind to egg coats (17, 160, 205, 221-223). Notably, in the three sea urchin genera that contain allopatric species, bindin has been found to be evolving slowly (17). These observations provide support for the hypothesis of reinforcement, where the rapid evolution of bindin in sympatric species may be a mechanism to prevent hybridization and reinforce species boundaries (17, 35, 205, 223). Conversely, in allopatric populations with no overlapping habitats, interspecific hybrids cannot occur, eliminating selective pressure to reinforce differences in gamete recognition proteins and thus slowing the evolutionary rate of bindin (17).

Gamete recognition shows species-specificity in abalone as well (14, 116). In *in vitro* egg coat dissolution experiments, significantly more lysin is required to dissolve egg coats in heterospecific pairings than in conspecific pairings, indicating that the interaction of lysin and VERL is the species-specific step in abalone fertilization (14, 117). Swapping lysin amino acid

sequences between two species of abalone to create chimeric proteins determined that the N- and C-termini of lysin – in particular the hypervariable N-terminus, which is always species-unique – mediated species-specific recognition of the abalone egg coat (116). Notably, a recent NMR solution structure of lysin found that its N- and C-termini are close together in physical space, forming a nexus that also contains most of lysin's positively selected residues (224). These regions are thus good candidates for determining which parts of lysin structurally interact with VERL to maintain species boundaries (14, 35).

1.8 Conclusions

The egg coats surrounding all metazoan eggs are closely related and share common structural features, such as the ZP module. Furthermore, the genes encoding egg coat proteins likely share a common ancestral gene, potentially ZP3 (4, 6, 21, 34). Despite the deep homology of the egg's extracellular barrier, however, a frequent observation among egg coat proteins is their rapid evolution.

In humans, approximately 10% of *in vitro* fertilization (IVF) attempts fail, with no known cause (225). Such cases of unexplained infertility may be due to incompatibility between sperm-egg recognition proteins, driven by the rapid evolution of reproductive loci (190). Correlating this sequence divergence with protein function may help to elucidate the molecular basis of reproductive incompatibilities, with implications for infertility, reproductive isolation, and speciation (14, 35, 190).

Anomalies in the structure and thickness of the egg coat are known to affect reproductive fitness, with IVF success being one particularly well-studied aspect; notably, a high degree of sequence variation has been observed in all four human ZP glycoproteins (ZP1-4) in infertile women presenting to IVF clinics (226-230).

On the other hand, ZP glycoproteins are also potential targets for contraception (231). Recently, it was shown that the mouse sperm binding region of ZP2, when conjugated to agarose beads inserted into the uterus of fertile mice, provided a long-acting method of contraception by decoying and binding sperm, preventing their access to the egg (232). This same construct, when paired with the human ZP2 sperm binding region, could select for sperm that are better able to bind and penetrate the egg coat, suggesting the suitability of this method for assisted reproductive technologies as well (232).

Continued research into the structure, function, and evolution of metazoan egg coat proteins has great potential to contribute to both human health and the field of evolutionary biology.

Chapter 2. PROTEOMICS SUPPORT THE THREESPINE STICKLEBACK EGG COAT AS A PROTECTIVE OOCYTE ENVELOPE

2.1 INTRODUCTION

Animal oocytes are surrounded by a specialized glycoprotein extracellular matrix termed the “egg coat” (1, 2, 4, 233). The egg coat is an interface between the egg and its environment, protecting the oocyte from physical, chemical, and biological hazards (2, 4, 115, 234). It is also an interface between gametes during fertilization, playing roles in attracting and activating sperm, mediating sperm recognition and binding, and blocking the detrimental fitness costs of polyspermy (2, 4, 234). The egg coat goes by different names in the major vertebrate lineages, including the zona pellucida (ZP) in mammals, the vitelline envelope in non-mammals, and the chorion in fish (4, 6). Despite historically complicated nomenclature, egg coats are generally comprised of a common set of glycoproteins characterized by the zona pellucida (ZP) module (40). The ZP module is an ~260 residue polymerization element consisting of a N-terminal ZP-N domain and a C-terminal ZP-C domain that both adopt immunoglobulin (Ig)-like folds (10, 40, 235). Beyond the core ZP module, many ZP proteins have more elaborate structures, including trefoil domains, transmembrane domains, consensus furin protease cleavage sites (CFCS), and tandem arrays of ZP-N repeats that have evolved independently of one another and their associated ZP-C (22, 23, 33, 41, 46, 236). Since ZP-N and ZP-C are independent structural domains, we will use the term “ZP module” to refer to the combined ZP-N and ZP-C domains rather than the more generic “ZP domain” (22, 44).

Vertebrate ZP proteins arose from a common ancestral gene through multiple duplication events hundreds of millions of years ago, giving rise to five gene families: *ZP1*/*ZP4*, *ZP2*, *ZP3*, *ZPAX*, and *ZPD* (see Figure 1; (30, 31)). ZP3 proteins, which are typically the smallest ZP protein, contain only the ZP module; this minimal architecture as well as molecular phylogenetics suggest that ZP3 may be most similar to the ancestral ZP protein (4, 6, 21, 34). ZP3 proteins can also have repetitive P/Q residues in relatively short stretches (237). With the exception of ZPD, all other ZP protein families (ZP1/4, ZP2, and ZPAX) contain additional ZP-N domain repeats N-terminal to their ZP module (see Figure 1; (6, 33)). These N-terminal ZP-N domains tend to be less conserved among orthologous proteins of different species (33). ZP1-like proteins typically possess a N-terminal ZP-N domain repeat followed by a P/Q-rich region, a trefoil domain, and a ZP module (33, 237). ZP2 proteins are characterized by multiple N-terminal ZP-N domain repeats prior to their ZP module, and the ZP2 homolog ZPAX has an analogous N-terminal ZP-N domain repeat architecture (33).

ZP proteins are synthesized as precursor polypeptides with a signal sequence (SS) at the N-terminus and a C-terminal propeptide containing a transmembrane domain (TMD) (9, 11, 23). In some fish, however, the TMD is absent (9, 237). The ZP module itself consists of 8, 10, or 12 disulfide-bonded cysteine residues, followed by a CFCS and, if present, a TMD or hydrophobic sequence (9, 11, 237). The dimerization of ZP-N domains between ZP modules facilitates the assembly of the filamentous egg coat ultrastructure (7, 45, 236, 238, 239).

In mammalian egg coats, ZP proteins serve as both structural and sperm-binding proteins (10, 159, 171, 177, 185, 188, 232). In fish, however, the role of ZP proteins in the egg coat is less well characterized and may be purely structural (9, 12, 237). Teleost fish sperm lack an acrosome, a secretory vesicle involved in sperm-egg binding, and teleost fish eggs contain an additional

structure called the micropyle, a funnel-shaped, narrow channel through the egg coat that permits sperm to reach the plasma membrane of the egg (2, 106, 108, 109, 111). The micropyle attracts sperm by chemotaxis, and its precise diameter restricts polyspermy by allowing passage of only one sperm at a time (5, 109-111). Whereas sperm in other animals bind to and dissolve the egg coat at the point of contact, in teleost fish the micropyle is solely responsible for sperm entry through the egg coat (111).

In mammals, ZP proteins are synthesized in the ovary by oocytes and/or their surrounding follicle cells (34). In fish, however, ZP proteins can be expressed in the liver as well as the ovary in response to estrogen, and subsequently transported through the bloodstream to the ovary to assemble around eggs (39, 70, 98, 115). This additional site of ZP synthesis may reflect the comparatively large size of fish egg clutches, necessitating the synthesis of large amounts of protein in a relatively short time (115, 237, 240, 241). ZP1 and ZP3, the most common ZP proteins in fish egg coats, both have paralogous classes of genes with hepatic and ovarian expression (31, 237, 240). Species-specific gene amplifications and losses have resulted in some teleost fish, such as zebrafish, retaining only ovarian expression; others retain both ovary and liver expression, and others solely liver (39, 98). One of the two expression sites typically becomes dominant, with liver synthesis of ZP proteins most common across teleosts (39, 98). Vitellogenin, an egg yolk precursor protein, shows similar hepatic expression and migration to the ovary in the bloodstream of fish, amphibians, and birds (9, 70). In fish, ZP synthesis and vitellogenesis occur simultaneously in response to 17β-estradiol production by follicle cells (9, 12, 98, 115).

Reproductive proteins that mediate gamete recognition during fertilization show species-specificity in both their structure and binding affinities (207, 209). Despite their central role in fertilization, however, reproductive proteins are frequently among the most rapidly evolving genes

in any taxa (13-18, 209). This juxtaposition of rapid evolution and functional constraint suggests a role for positive Darwinian selection in the maintenance of sperm-egg interactions. Furthermore, the molecular evolutionary history of a protein can identify sites under adaptive evolution that may be functionally important (35, 117, 224, 242-245). Signatures of rapid, adaptive evolution characteristic of reproductive proteins suggest that sequence diversification can be beneficial to genes involved in reproduction (209). More formally, when nonsynonymous (d_N) substitutions outweigh synonymous (d_S) substitutions, d_N/d_S (also denoted ω) is greater than one and suggests there was positive selection for changes in amino acid sequence (35, 196, 209). Positive selection on gamete recognition proteins can contribute to reproductive isolation between diverging taxa, with variation between diverging populations creating species barriers that may ultimately lead to speciation (2, 4, 14, 16, 194).

Threespine stickleback fish have been called "Darwin's fishes" in light of their remarkable adaptive radiation throughout the Northern Hemisphere following glacial retreat at the end of the last ice age (~12,000 years ago) (246, 247). Ancestrally marine fish have colonized thousands of freshwater lakes and streams, evolving significant diversity in morphology, behavior, physiology, and life history (246-249). These divergent forms come into contact with each other, but are frequently reproductively isolated, making stickleback an ideal model system for speciation research (247, 249). Speciation, in the sense of sympatric populations of stickleback coexisting without interbreeding, is often rapid, and attributed to differences in male morphology and behavior and female preferences for those traits as well as ecological selection against hybrids (247, 249, 250). Despite nearly complete reproductive isolation in the wild, however, virtually any stickleback can be crossed in the lab to produce viable, fertile offspring (246). Whereas the evolution of reproductive isolation in stickleback has been attributed to divergent natural and

sexual selection, the contribution of rapidly evolving reproductive proteins to stickleback speciation has so far not been considered (249). To begin to address this question from the perspective of female reproductive protein evolution, here we combine proteomics and molecular evolutionary analyses to characterize the egg coat proteome of threespine stickleback fish.

2.2 EXPERIMENTAL PROCEDURES

2.2.1 *Animal statement*

Threespine stickleback fish were collected from a single freshwater site in Lake Union, Washington, USA (47°38'55" N, 122°20'47" W) during their annual breeding season in 2015 (Washington Department of Fish and Wildlife permit 15-033 to C. Peichel). Fish were collected with minnow traps, eggs were obtained from gravid females, and they were euthanized shortly after collection by immersion in 0.2% MS-222. All animals were collected under permits obtained from the Washington Department of Fish and Wildlife, and all animal methods were conducted in accordance with the guidelines of the Fred Hutchinson Cancer Research Center Institutional Animal Care and Use Committee (Protocol 1575 to C. Peichel).

2.2.2 *Egg coat isolation*

Eggs were obtained from gravid female stickleback by gentle abdominal squeezing, and lysed by periodic homogenization in 1% Triton X-100 detergent in Hank's solution (138 mM sodium chloride, 5 mM potassium chloride, 0.25 mM sodium phosphate dibasic, 0.4 mM potassium phosphate monobasic, 1.3 mM calcium chloride, 1 mM magnesium sulfate, 4 mM sodium bicarbonate; adapted from (251)). Insoluble egg coats were isolated by centrifugation (2,000 x *g* for 10 minutes), and contaminating egg cytosolic proteins were removed with repeated washes of 1% Triton X-100 in Hank's solution followed by centrifugation. Some

samples were additionally treated with 7 M urea to remove trace contaminants of vitellogenin without affecting major egg coat proteins.

2.2.3 *Analysis of egg coats by SDS-PAGE*

Stickleback egg coats were analyzed under both reducing and non-reducing conditions by SDS-PAGE with 12% acrylamide gels and a tris-tricine buffering system; electrophoresis was performed at 50 V for 15 minutes, followed by 100 V for 90 minutes (252). Samples were prepared by incubation of solid egg coats in a 1% SDS solution, with or without 2-mercaptoethanol, at 95°C, with insoluble material removed by centrifugation. Proteins were stained with either Coomassie Brilliant Blue R-250 (MilliporeSigma, St. Louis, MO, USA) or SYPRO Ruby (Thermo Fisher Scientific, Waltham, MA, USA). Glycosylation was detected by in-gel periodic acid-Schiff staining using the Pro-Q Emerald 488 glycoprotein staining kit (Invitrogen, Carlsbad, CA, USA) and imaged using a Typhoon FLA 9000 laser bed scanner (GE Healthcare Bio-Sciences, Pittsburgh, PA, USA). To determine if the observed glycosylation of stickleback ZP3 was *N*-linked, egg coats were treated with PNGase F (New England BioLabs, Ipswich, MA, USA) prior to electrophoresis following the manufacturer's protocol.

2.2.4 *Mass spectral characterization of egg coats*

Following SDS-PAGE of stickleback egg coats, individual protein bands were excised using a sterile scalpel blade, cut into ~1 mm3 cubes, and placed in a 1.7 ml tube. Remaining Coomassie dye was extracted through multiple rounds of addition of 50 mM ammonium bicarbonate (with 15 minute incubation), addition of acetonitrile (with 15 minute incubation), removal of supernatant, and drying of the gel pieces in a vacuum centrifuge. After the dye was completely removed, disulfide bonds were reduced by incubating the gel pieces in 20 mM DTT

in 100 mM ammonium bicarbonate at 56°C for 45 minutes, followed by alkylation with 55 mM iodoacetamide in 100 mM ammonium bicarbonate in the dark for 30 minutes. The gel pieces were washed twice with 100 mM ammonium bicarbonate, dehydrated with acetonitrile, and incubated with 1 µg trypsin (Promega, Madison, WI, USA) in 50 mM ammonium bicarbonate overnight at 37°C. The supernatant was then collected, the gel pieces washed twice with 50 mM ammonium bicarbonate and acetonitrile, and the washes added to the collected supernatant. The final collected solution was concentrated by evaporative centrifugation and resolubilized in 10 µl 5% acetonitrile / 0.1% formic acid. Three µl of each sample was loaded onto a 30 cm fused silica 75 µm column and 3.5 cm 150 µm fused silica KASIL 1 frit trap (PQ Corporation, Malvern, PA, USA) loaded with 4 µm Jupiter C12 Proteo reverse-phase resin (Phenomenex, Torrance, CA, USA) and analyzed with Thermo Fisher Scientific EASY-nLC. Buffer A was 0.1% formic acid in water and Buffer B was 0.1% formic acid in acetonitrile. The 60-minute LC gradient consisted of 2 to 40% B in 30 minutes, 40 to 60% B in 10 minutes, 60 to 95% B in 5 minutes, followed by a 15 minute wash and a 15 minute column equilibration. Peptides were eluted from the column and electrosprayed into a Velos Pro Linear Ion Trap Mass Spectrometer (Thermo Fisher Scientific). Data was acquired using data-dependent acquisition (DDA) and analyzed using an in-house version of COMET (253, 254) (with a differential modification of 15.994915 Da for methionine and a static modification of 57.021461 Da for cysteine) for database searching against publicly available stickleback ESTs (retrieved from the UCSC Genome Browser) that were assembled with Trinity (255, 256) and six-frame translated. The search database also contained known contaminants such as trypsin and human keratin. Percolator v.2.09 (257) was used to filter the peptide-spectrum matches with a q-value threshold of ≤ 0.01, and peptides were assembled into protein identifications using an in-house implementation of IDPicker (258).

2.2.5 *Sequencing of stickleback ZP cDNA*

Total RNA was isolated from *G. aculeatus* ovary and liver tissue by lysis in guanidinium isothiocyanate and cesium chloride gradient ultracentrifugation (procedure modified from (259)). Briefly, tissues were homogenized in five volumes of 4 M guanidinium isothiocyanate in a Dounce homogenizer, 10% SDS was added to a final concentration of 0.1%, and the mixture centrifuged for 5 minutes at 5,000 x *g* to remove insoluble debris. The supernatant was then layered over 5.7 M cesium chloride, centrifuged at 154,000 x *g* for 23 hours at 20°C, purified RNA was washed three times with 70% ethanol, and resuspended in RNase-free water. *G. aculeatus* ovary and liver cDNA was prepared from total RNA using the SMARTer cDNA synthesis kit (Clontech, Palo Alto, CA, USA).

ZP1 and ZP3 coding sequences were PCR amplified from *G. aculeatus* liver cDNA (primer sequences in Table 1), cloned into the pCR4-TOPO vector (Invitrogen), transformed into NEB 5-alpha chemically competent *Escherichia coli* (New England BioLabs), and submitted for Sanger sequencing (Eurofins Genomics, Louisville, KY, USA). Sequences were analyzed using the Lasergene DNASTAR package (v.11.1.0; Madison, WI, USA).

Table 1. Gene-specific primers used for ZP RT-PCR. Primers were designed to anneal to the 5' and 3' UTR of stickleback ZP1 and ZP3, and included an EcoRI overhang (GAATTC).

Primer	Sequence
ZP1 forward	GAATTCATGGCAAAGCTGGCCACCTCTTCT
ZP1 reverse	GAATTCTTACTGCGCATTCAGCTCCACCGG
ZP3 forward	GAATTCATGGTGATGAAGTGGACTGTGTGC
ZP3 reverse	GAATTCTCAAACAACCATCTTCTCTGCAATGTTGAT

2.2.6 *Disulfide bond characterization*

To investigate the disulfide bonding pattern of stickleback ZP1 and ZP3, egg coat samples were prepared under different reduction and alkylation conditions prior to trypsin proteolysis and mass spectral characterization: (1) no reduction or alkylation (disulfide identification), (2) alkylation without reduction (reduced cysteine identification), (3) alkylation followed by reduction (disulfide identification with potentially better trypsin cleavage site accessibility), and (4) reduction followed by alkylation (traditional peptide fingerprinting). To volumetrically match samples, 100 mM ammonium bicarbonate was substituted in place of reagents as necessary. Briefly, an initial reduction was performed with 100 mM BME in 7 M urea in 100 mM ammonium bicarbonate, and the samples were incubated at 60°C for 45 minutes. Samples were then alkylated with 200 mM iodoacetamide in 7 M urea in 100 mM ammonium bicarbonate and incubated for 45 minutes in the dark. A final reduction was performed with 400 mM BME, and all four samples were diluted 1:4 with ammonium bicarbonate to reduce urea concentration. Trypsin (2 µg; Promega) was added to the samples before incubation at 37°C overnight. The samples were then acidified with 1% TFA, desalted by C18 ZipTip (MilliporeSigma), concentrated by evaporative centrifugation, and resolubilized in 10 µl 5% acetonitrile / 0.1% formic acid. Three µl of each sample was loaded onto a 30 cm fused silica 75 µm column and 3.5 cm 150 µm fused silica KASIL 1 frit trap (PQ Corporation) loaded with 3 µm Reprosil-Pur C18 reverse-phase resin (Dr. Maisch, Ammerbuch, Germany) and analyzed with Thermo Fisher Scientific EASY-nLC. Buffer A was 0.1% formic acid in water and buffer B was 0.1% formic acid in acetonitrile. The 100-minute LC gradient consisted of 0 to 16% B in 15 minutes, 16 to 35% B in 60 minutes, 35 to 75% B in 15 minutes, 75 to 100% B in 5 minutes, followed by a 5 minute wash and a 25 minute column equilibration. Peptides were eluted from the column on a 50°C heated source (CorSolutions,

Ithaca, NY, USA) and electrosprayed into an Orbitrap Fusion Lumos Mass Spectrometer (Thermo Fisher Scientific). Data was acquired using data-dependent acquisition (DDA) with dynamic exclusion turned off. Mass spectral data was analyzed with MassMatrix v.3.0.10.25 to detect disulfide-linked peptides, with ZP1 and ZP3 coding sequences (cloning described above) used as the search database (260-265).

2.2.7 *Molecular evolution of teleost ZP proteins*

To assess ZP gene evolution within teleost fish, 30 species were chosen spanning the teleost phylogeny, with ZP1 and ZP3 open reading frames (ORFs) identified by homology to stickleback ZP1 and ZP3 using TBLASTX (266, 267). For the 31 total species (including *G. aculeatus*), sequences for each gene were aligned separately using Clustal Omega and a concatenated gene tree was constructed using RAxML with the PROTGAMMALG substitution model (268, 269). Rates of molecular evolution were calculated using PAML v.4.8, with site models M8a (nearly neutral) and M8 (positive selection) compared by likelihood ratio test (195, 270). Sites under positive selection were defined as coding positions with a Bayes empirical Bayes (BEB) posterior probability of > 50% under M8 (271).

A homology model of stickleback ZP3 was generated using Rosetta by threading of the stickleback ZP3 sequence to the available chicken ZP3 structure (PDB ID: 3NK4; (235)) (aligned using Clustal Omega), loop modeling using cyclic coordinate descent (CCD) with refinement by kinetic closure (KIC), and full atom minimization using the relax function (269, 272, 273). *N*-glycosylation was modeled using GlycanBuilder (274).

2.3 RESULTS

2.3.1 *Egg coat glycoprotein characterization*

To characterize the proteome of threespine stickleback egg coats, egg coats were isolated and examined by SDS-PAGE (Figure 4). Individual bands were excised and analyzed by liquid chromatography-tandem mass spectrometry (LC-MS/MS), with the two main protein components of stickleback egg coats identified as ZP1 and ZP3 (Table 2). The remaining bands represent carryover of vitellogenin from the egg yolk during egg coat isolation. Treatment with 7 M urea removes the contaminating vitellogenin bands, with no apparent loss in intensity of ZP1 or ZP3 (Figure 5).

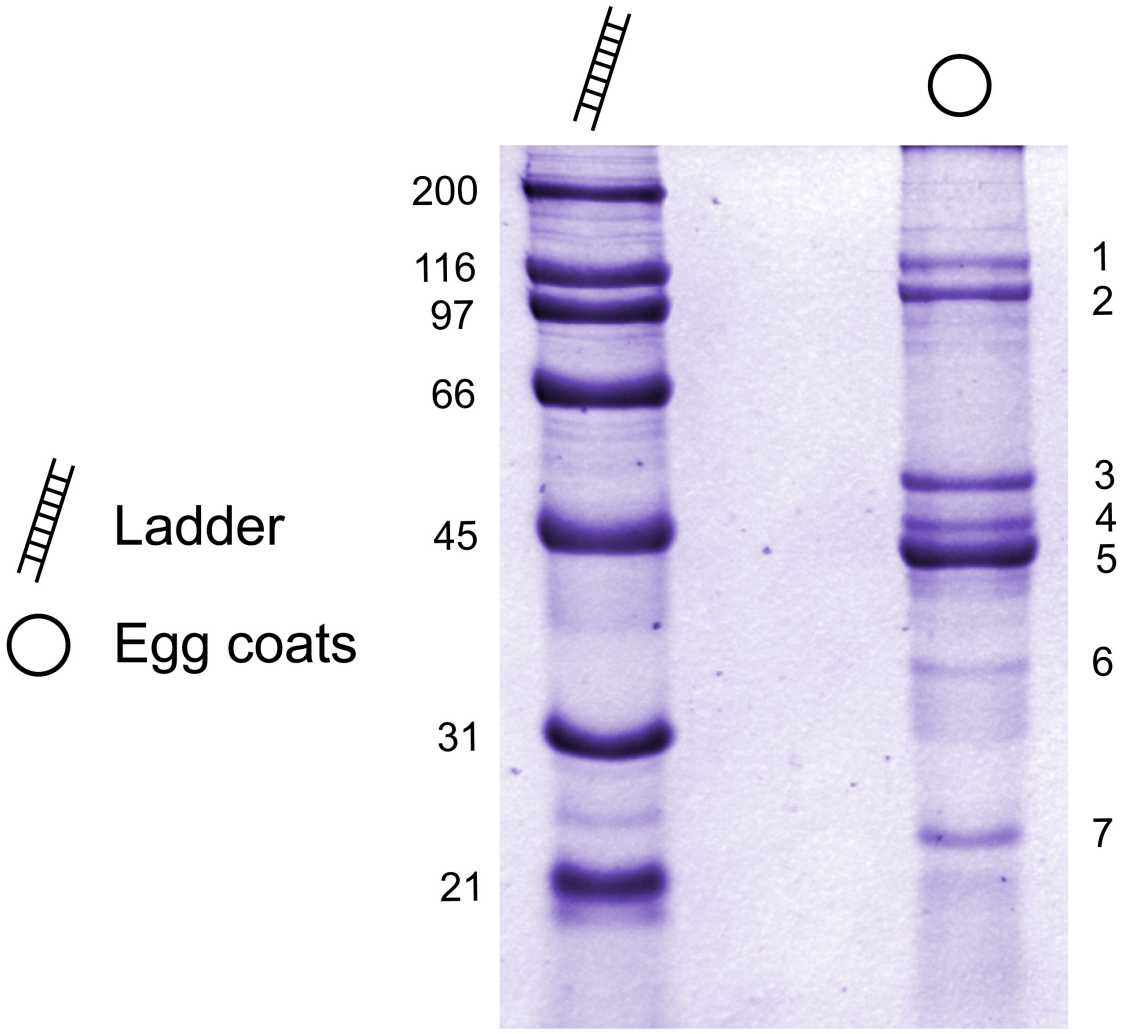

Figure 4. SDS-PAGE of stickleback egg coats for gel extraction-mass spectrometry. Stickleback egg coats were separated with SDS-PAGE, and individual bands (indicated by number, see Table 2) were gel extracted and analyzed by LC-MS/MS. ZP1 and ZP3 were found to be the two major components of stickleback egg coats, with the remaining bands representing vitellogenin contamination likely carried over from egg coat isolation.

Table 2. Summary of mass spectral data for SDS-PAGE-extracted stickleback egg coat proteins. Table of LC-MS/MS identifications for gel extracted stickleback egg coat proteins (see Figure 4). Sequence matches were sorted by normalized spectral abundance factor (NSAF), with the most abundant Trinity contig match reported (275). Note that for Gel Band 3, the two Trinity contig matches have identical protein sequence.

Gel band number	Accession ID(s)	Sequence match	Number of peptides	Number of spectra	Percent total coverage	NSAF
1	TR556|c0_g1_i1_1+	Vitellogenin	62	296	76	0.00385
2	TR556|c0_g1_i1_1+	Vitellogenin	89	726	78	0.08779
3	TR6893|c1_g15_i1_3- TR6893|c1_g17_i1_1-	ZP1	38	329	54	0.01256
4	TR7203|c0_g14_i1_3+	ZP3	39	244	50	0.02208
5	TR7203|c0_g14_i1_3+	ZP3	56	916	54	0.03546
6	–	Unknown	–	–	–	–
7	TR7069|c0_g6_i1_1-	Vitellogenin	51	220	44	0.02319

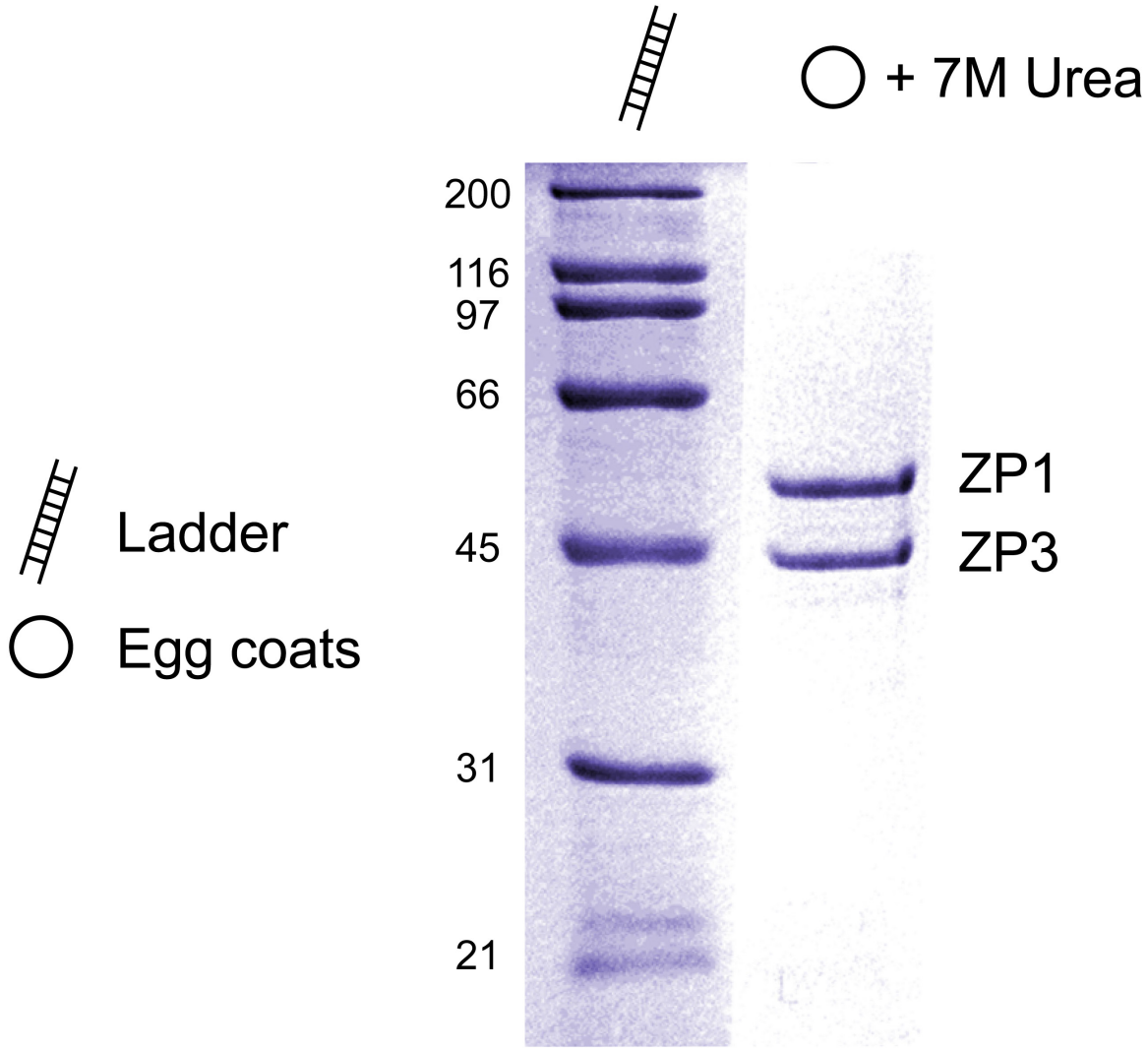

Figure 5. SDS-PAGE of stickleback egg coats treated with 7 M urea. Stickleback egg coats were treated with 7 M urea to remove contaminating vitellogenin, likely carried over from egg coat isolation, and separated by SDS-PAGE. Individual bands were excised from the gel and analyzed by LC-MS/MS, with ZP1 and ZP3 identified as the two major protein components of stickleback egg coats.

Reproductive proteins are frequently glycosylated (276, 277). These post-translational modifications affect protein solubility and stability, and are thought to play a role in gamete recognition (35, 278). Glycosylation analysis of stickleback egg coats indicates that of the two main egg coat proteins, only ZP3 is glycosylated (Figure 6 and Figure 7). Stickleback ZP3 has a single putative *N*-glycosylation motif at N_{160}, and treatment with PNGase F confirmed that the glycan is *N*-linked (Figure 7). Notably, this particular *N*-linked glycosylation site is highly conserved from fish to mammals (101).

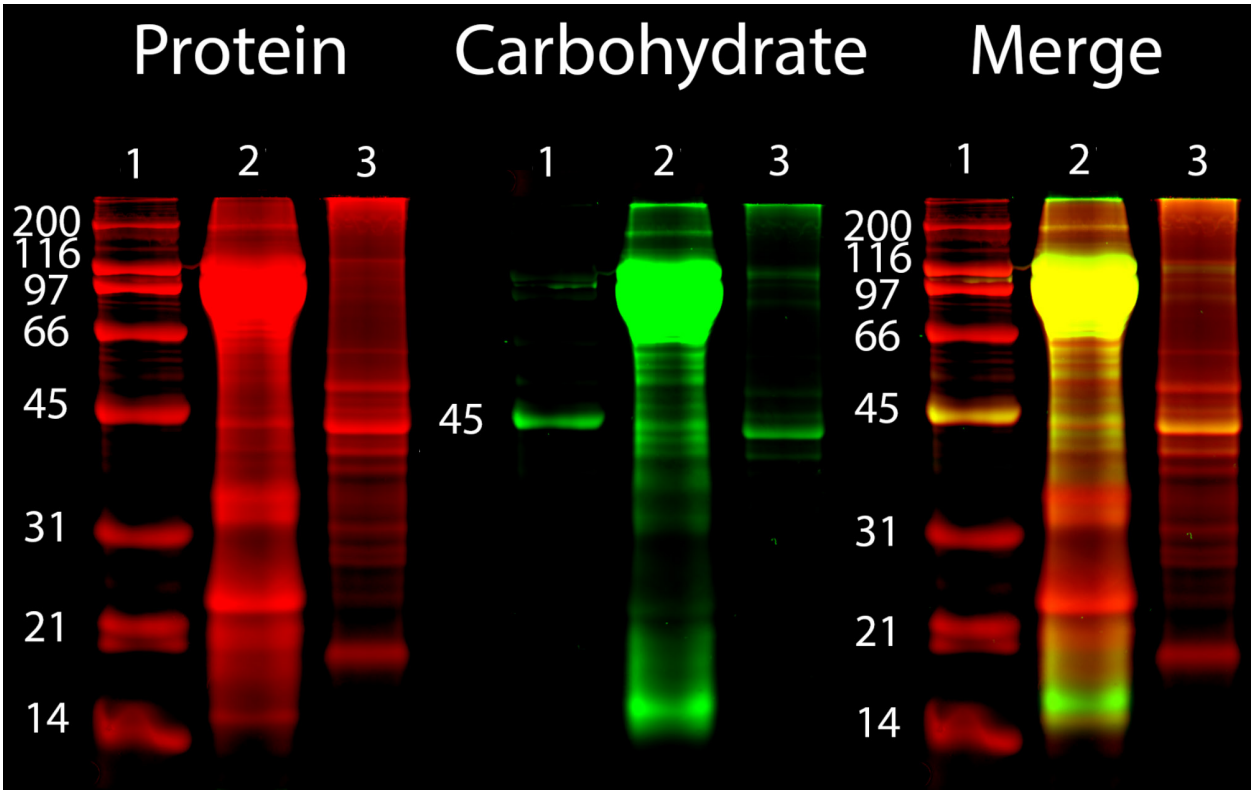

Figure 6. Glycosylation analysis of stickleback egg coats. Stickleback eggs were treated with 1% Triton X-100 detergent to separate the soluble egg lysate (lane 2) from the insoluble egg coats (lane 3). Samples were separated by SDS-PAGE, the gel was stained with fluorescent carbohydrate

and protein dyes, and the images were overlaid to visualize glycoprotein staining. Of the two stickleback egg coat proteins, only ZP3 is glycosylated. The large blob of high molecular weight in the egg lysate lane is vitellogenin, a glycolipoprotein that is a major component of egg yolk. Note that the 45 kDa band in the ladder (lane 1) is ovalbumin, a glycoprotein which serves as a positive control. Lane 1: ladder; lane 2: soluble egg lysate; lane 3: insoluble egg coats.

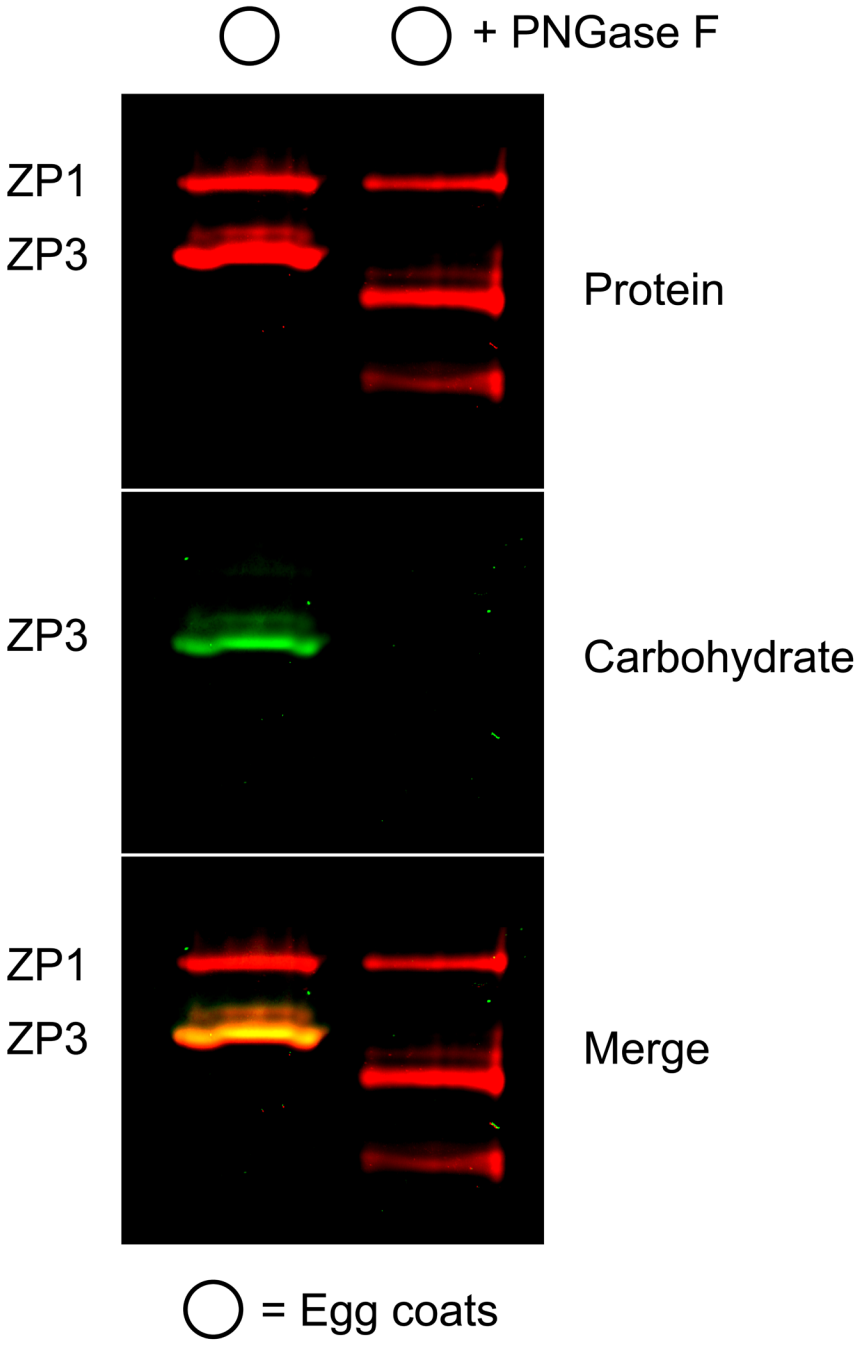

Figure 7. Glycosylation analysis of stickleback egg coats treated with 7 M urea. Stickleback egg coats were treated with 7 M urea, deglycosylated with PNGase F, and separated by SDS-PAGE. The gel was stained with fluorescent carbohydrate and protein dyes, and the images were overlaid

to visualize glycoprotein staining. Of the two stickleback egg coat proteins, only ZP3 is glycosylated, and the glycosylation appears to be *N*-linked as the protein no longer stains for carbohydrate after PNGase F treatment. Based on the mass shift after deglycosylation and the single *N*-linked glycosylation site in ZP3, the glycan appears to be ~4 kDa. Note that the low molecular weight band in the egg coat + PNGase F lane is PNGase F.

2.3.2 *ZP disulfide bond characterization*

The insoluble nature of stickleback egg coats in the absence of reducing conditions (even in 7 M urea) suggests that intermolecular disulfide bonds may stabilize the egg coat structure. To determine the disulfide bonding patterns of stickleback ZP1 and ZP3, egg coats were subjected to differing reduction and alkylation conditions prior to performing LC-MS/MS, with dynamic exclusion turned off to permit more quantitative peptide spectral counting. Reverse Transcriptase PCR (RT-PCR) of ZP1 and ZP3 from stickleback ovary and liver RNA showed strong amplification from liver RNA only, suggesting that in stickleback ZP genes are transcribed in the liver (Figure 8). ZP sequences obtained from RT-PCR were used as the database for LC-MS/MS searches, resulting in 66% sequence coverage from 25 peptides for ZP1 and 69% sequence coverage from 24 peptides for ZP3.

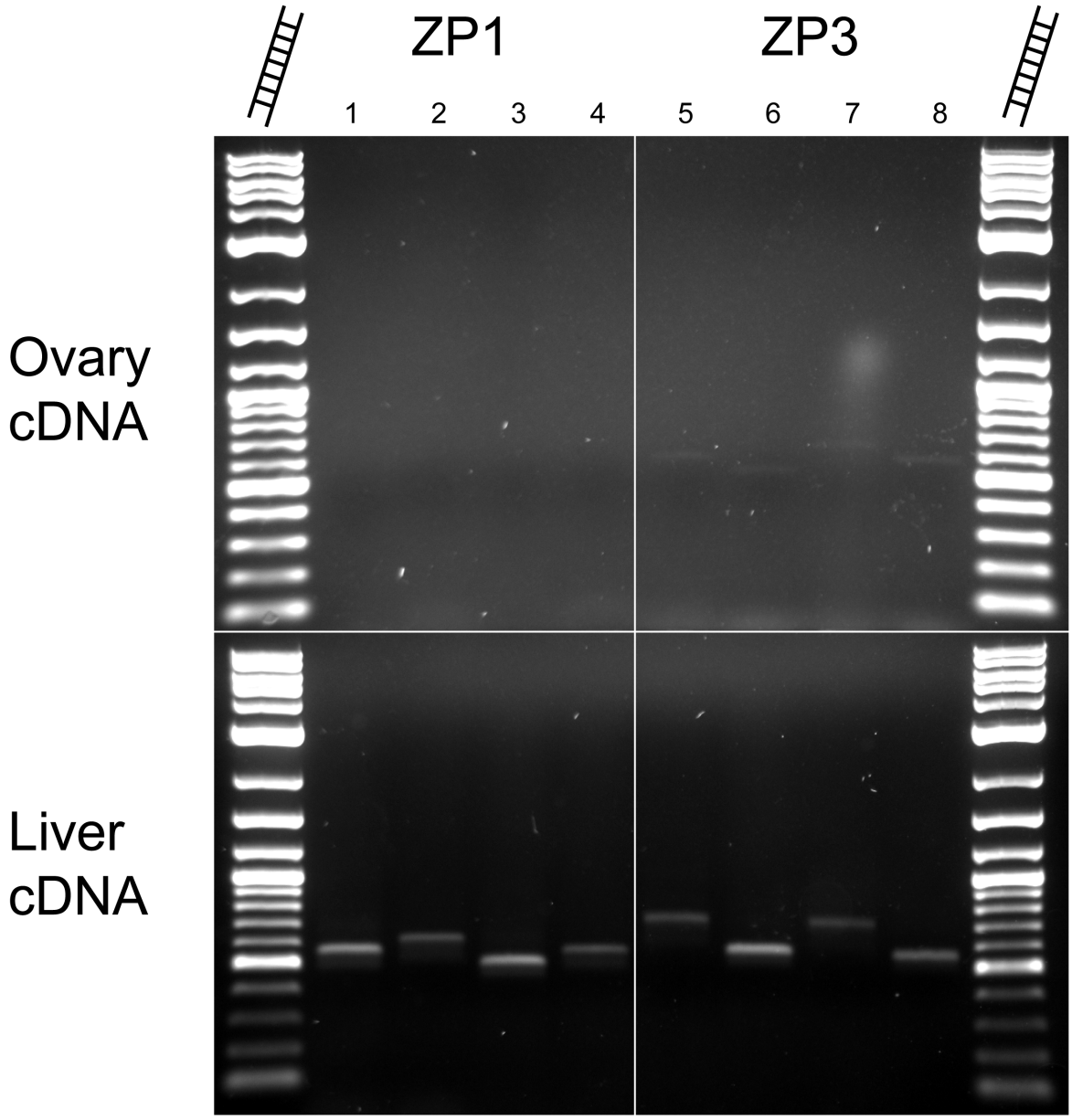

Figure 8. Amplification of ZP1 and ZP3 from stickleback liver cDNA. To determine the site(s) of ZP expression in stickleback, primers were designed against ZP1 and ZP3 and amplified from stickleback liver and ovary cDNA (primer sequences in Table 3, expected product size by lane in Table 4; ladder is TriDye 2-log DNA ladder (NEB)). Strong amplification occurred from liver

cDNA only (although some signal for ZP3 from ovary cDNA is present), suggesting that in stickleback, ZP genes are transcribed in the liver.

Table 3. Gene-specific primers used for ZP ovary and liver PCRs. Primers were designed against the most abundant Trinity contig by LC-MS/MS NSAF.

Primer	Sequence
ZP1_F1	AGCCTGGAAGAGCAATTGAA
ZP1_F2	CCGATGTACCTAGCCTGGAA
ZP1_R1	CCAATCTGGCCTCCTAATCA
ZP1_R2	TCCCAAATTGGTCTCCACAT
ZP3_F1	GCAACTGCAGGATTGTCTCA
ZP3_F2	GAAACTGTGCTGCTGTTGGA
ZP3_R1	TGGCAGGTGATGTAGAGCAG
ZP3_R2	TGGAACTGCAGCTTGTTGTC

Table 4. Expected product sizes for ZP primer pairs (see Figure 8).

Lane	Forward primer	Reverse primer	Expected product size (bp)
1	ZP1_F1	ZP1_R1	579
2	ZP1_F1	ZP1_R2	725
3	ZP1_F2	ZP1_R1	590
4	ZP1_F2	ZP1_R2	716
5	ZP3_F1	ZP3_R1	669
6	ZP3_F1	ZP3_R2	612
7	ZP3_F2	ZP3_R1	724
8	ZP3_F2	ZP3_R2	647

ZP proteins have characteristic disulfide bonding patterns within the ZP-N and ZP-C domains of their ZP modules. The crystal structure of chicken ZP3, for instance, shows a C_1-C_4, C_2-C_3 connectivity for ZP-N and a C_5-C_7, C_6-C_{11}, C_8-C_9, C_{10}-C_{12} connectivity for ZP-C (PDB ID: 3NK4; (235)). Although our analysis generally found evidence of homologous disulfide bonding in stickleback ZP proteins, we also see evidence for shuffled disulfides and new cysteines that could alter the disulfide bonding of stickleback ZP proteins (summarized in Figure 9). For both ZP1 and ZP3, the majority of cysteines within the ZP module were modifiable with iodoacetamide in the absence of reducing agent, suggesting variable and/or transient disulfide bonding. Consistent with disulfide shuffling, both stickleback ZP1 and ZP3 have an odd number of cysteine residues in their ZP module (11 vs. 12 in chicken ZP3). Furthermore, stickleback ZP1 has two additional cysteines (C_4 and C_5) in the linker between its ZP-N and ZP-C domains, and ZP3 has an additional cysteine (C_9) in its ZP-C domain (see Figure 9). Table 5 provides counts for all potential disulfide-linked peptides by LC-MS/MS for ZP1, and Table 6 provides counts for ZP3.

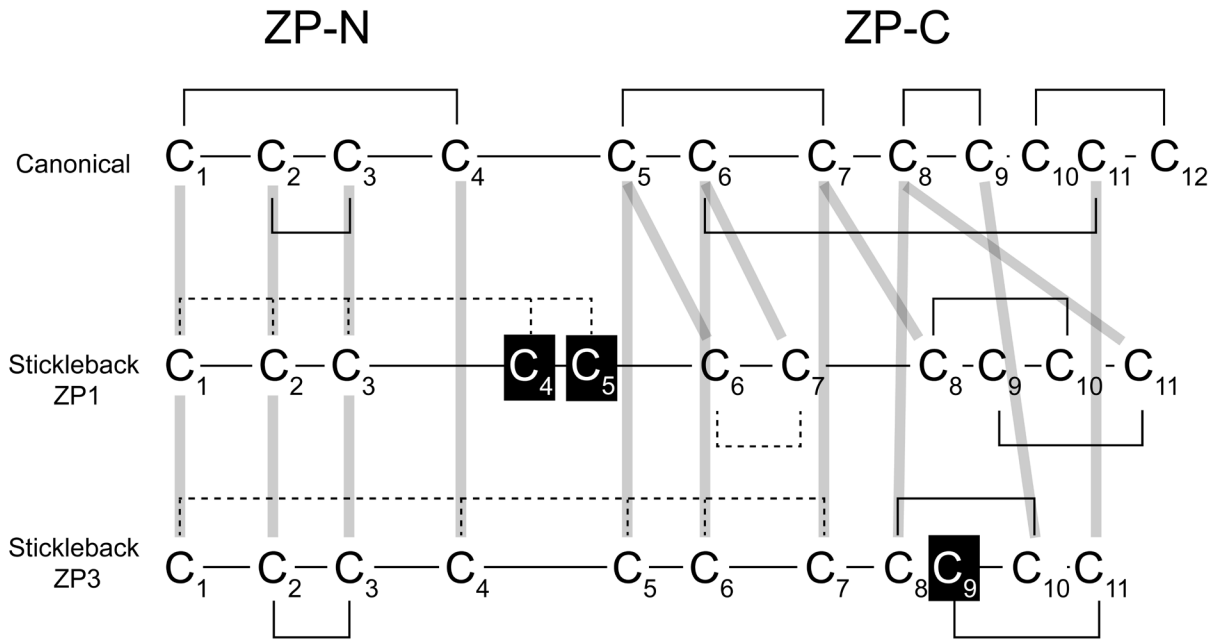

Figure 9. Disulfide bonding pattern of stickleback ZP proteins. Summary of the observed disulfide bonding pattern of stickleback ZP1 and ZP3. The "canonical" disulfide bonding pattern is from the crystal structure of chicken ZP3 (PDB ID: 3NK4; (235)), with cysteines connected by grey lines homologous between stickleback and chicken ZP3. Relative distance between cysteines is indicated by the length of the backbone. Dashed disulfide bonds in stickleback ZP1 and ZP3 denote potential disulfide shuffling, as the indicated cysteines were found to be carbamidomethyl (CAM) modified by mass spectrometry. Cysteines 4 and 5 of stickleback ZP1 (boxed in black) and cysteine 9 of stickleback ZP3 (boxed in black) have no homolog in chicken. Note that cysteines 9 and 10 of stickleback ZP1 are not homologous to chicken ZP3, but do have homologs in chicken ZP1, so are not indicated in black. Stickleback ZP1 is additionally missing its canonical cysteine 4 in its ZP-N domain. For both stickleback ZP1 and ZP3, cysteines 8, 9, 10, and 11 were present on the same peptide so disulfide bonding was inferred by homology to chicken.

Table 5. Stickleback ZP1 disulfide bonding patterns by mass spectrometry. The disulfide bonding pattern of stickleback ZP1 was assessed by detection of cysteine cross-linked peptides with LC-MS/MS; the number of peptides supporting each disulfide bond are indicated in parentheses. Homologous cysteines are from a chicken ZP3 crystal structure (PDB ID: 3NK4; (235)), as no ZP1 structure is currently available; a dash indicates that no homologous cysteine is present in chicken ZP3. Cysteines are considered "IAA reactive" if they were found to be carbamidomethyl (CAM) modified after iodoacetamide (IAA) alkylation. Disulfide bonding patterns were inferred by homology to chicken for cysteines 6 and 7, as no cross-linked peptides were detected, as well as for cysteines 8, 9, 10, and 11, since they were present on the same peptide. Note that stickleback ZP1 contains an odd number of cysteine residues.

Stickleback cysteine number	Homologous cysteine in chicken	IAA reactive	Potential disulfide bonds
1	1	✓	2 (5)
2	2	✓	1 (5), 4 (6)
3	3	✓	4 (4)
4	–	✓	2 (6), 3 (4), 5 (23)
5	–	✓	4 (23)
6	5	✓	7 (none detected)
7	6	✓	6 (none detected)
8	7		10 (3)
9	–		11 (3)
10	–		8 (3)
11	8		9 (3)

Table 6. Stickleback ZP3 disulfide bonding patterns by mass spectrometry. The disulfide bonding pattern of stickleback ZP3 was assessed by detection of cysteine cross-linked peptides with LC-MS/MS; the number of peptides supporting each disulfide bond are indicated in parentheses. Homologous cysteines are from a chicken ZP3 crystal structure (PDB ID: 3NK4; (235)); a dash indicates that no homologous cysteine is present in chicken ZP3. Cysteines are considered "IAA reactive" if they were found to be carbamidomethyl (CAM) modified after iodoacetamide (IAA) alkylation. Cysteines 8, 9, 10, and 11 were present on the same peptide, so bonding patterns of these cysteines were inferred by homology to chicken; additionally, cysteines 8 and 9 are consecutive residues and cannot disulfide bond with each other. Note that stickleback ZP3 contains an odd number of cysteine residues.

Stickleback cysteine number	Homologous cysteine in chicken	IAA reactive	Potential disulfide bonds
1	1	✓	4 (24), 5 (5), 6 (30), 7 (2)
2	2		3 (6)
3	3		2 (6)
4	4	✓	1 (24), 5 (10), 6 (11)
5	5	✓	1 (5), 4 (10), 6 (19)
6	6	✓	1 (30), 4 (11), 5 (19)
7	7	✓	1 (2)
8	8	✓	10 (9)
9	–		11 (9)
10	9		8 (9)
11	11	✓	9 (9)

2.3.3 *ZP molecular evolution*

Molecular evolutionary analyses of stickleback *ZP1* and *ZP3* (concatenated gene tree in Figure 10) suggest that the divergence of *ZP3* across teleosts has been driven by positive Darwinian selection, with 2.3% of sites in *ZP3* under positive selection with $\omega = 1.89$. To test for selection, a model of positive selection (M8) was compared to a model of neutral evolution (M8a) by likelihood ratio test (195, 270). These nested models allow for variation in ω among codons, but the null model M8a restricts ω to 1 while the alternative model M8 allows for adaptive evolution with $\omega > 1$. For *ZP3*, M8 fits the data significantly better than M8a, suggesting that allowing sites with $\omega > 1$ significantly improves the fit of the model to the data ($p = 1.2 \times 10^{-4}$; parameters summarized in Table 7). A similar test for adaptive evolution of *ZP1* across teleosts was not significant, in agreement with previous work where *ZP3* has been found to be under selection in mammals while *ZP1* is not. To our knowledge, this was the first investigation of ZP molecular evolution in fish (14, 279, 280). Residues under positive selection across teleosts are indicated as red spheres in stickleback ZP3 in Figure 11.

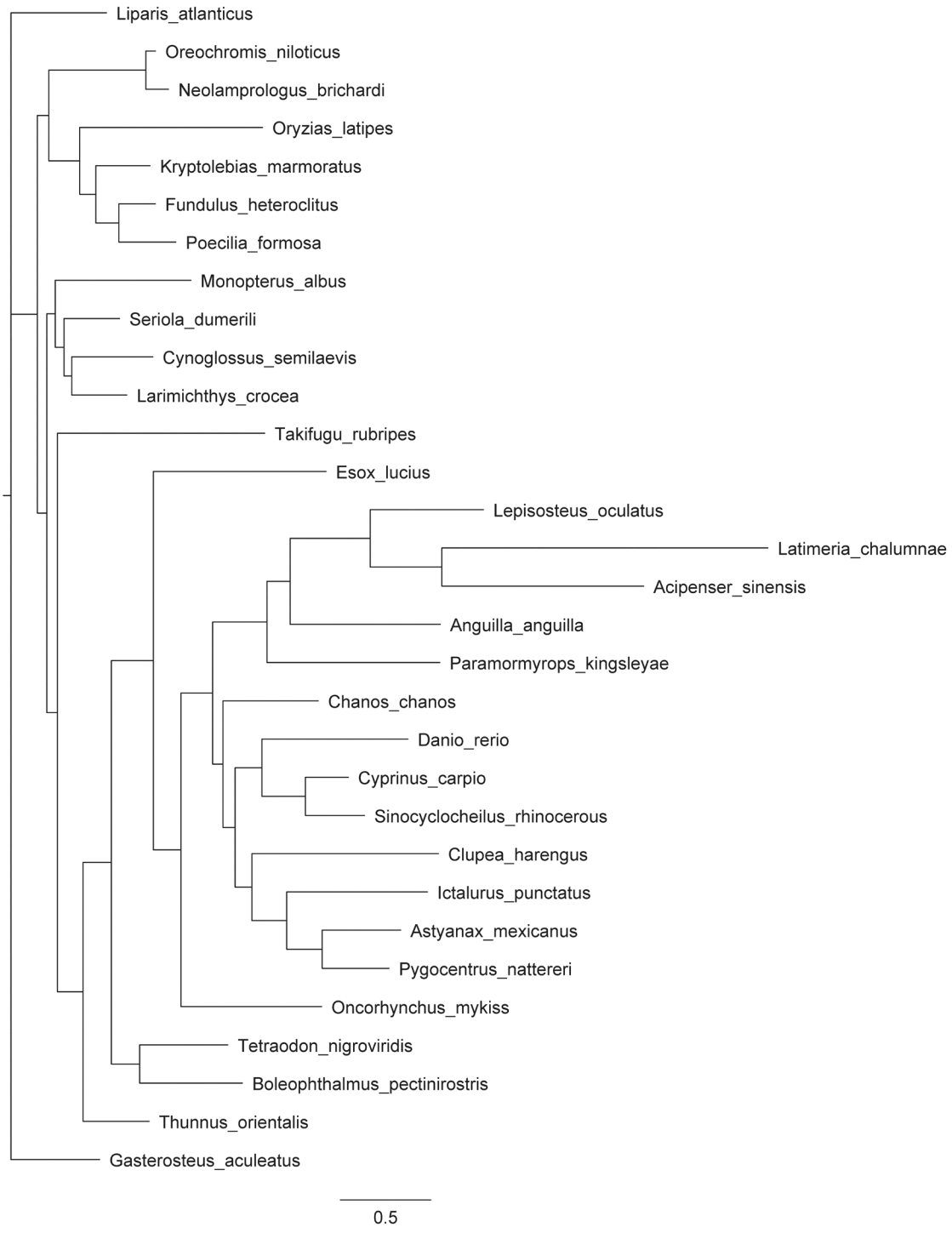

Figure 10. ZP evolution within teleosts. Maximum likelihood evolutionary tree of concatenated *ZP1* and *ZP3* from representative teleost species used for molecular evolutionary analyses.

Table 7. Evolutionary rate analysis of stickleback ZP proteins. The proportion of sites under positive selection (p_1) or under selective constraint (p_0) and the parameters p and q for the beta distribution $B(p, q)$ are given for *ZP1* and *ZP3* across teleosts. P-values for a likelihood ratio test comparing M8 (selection) to M8a (nearly neutral) are shown, with significant results highlighted in bold (195, 270). Sites under selection in *ZP3* are specified with respect to stickleback, with the signal peptide included.

Egg coat protein	M8a (neutral model)	M8 (positive selection)	-2ΔlogL	p-value	Sites under selection
ZP1	p_0 = 0.88429, p = 0.72988, q = 2.60577, p_1 = 0.11571, ω = 1	p_0 = 0.88920, p = 0.72148, q = 2.52174, p_1 = 0.11080, ω = 1.02158	0.071964	0.39	–
ZP3	p_0 = 0.94033, p = 0.65445, q = 1.85936, p_1 = 0.05967, ω = 1	p_0 = 0.97745, p = 0.61635, q = 1.52432, p_1 = 0.02255, **ω = 1.88757**	13.5158	**1.2 x 10^{-4}**	76, 86, 132, 136, 155, 159, 256, 283, 329

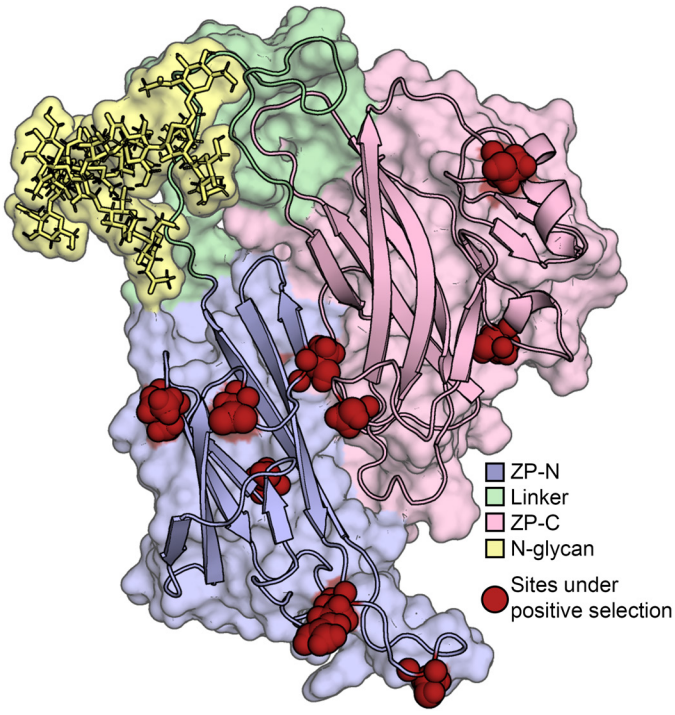

Figure 11. Stickleback ZP3 homology model. ZP-N and ZP-C domains are colored purple and pink, respectively; the linker between the domains is indicated in green; residues under positive selection across teleosts are denoted as red spheres; the single *N*-linked glycosylation site in stickleback ZP3 is shown in yellow. Note that sites under positive selection in *ZP3* tend to cluster, particularly those within the ZP-N domain.

2.4 DISCUSSION

As the first barrier sperm encounter during fertilization, the egg coat is an essential determinant of reproductive isolation in any taxa (16, 207, 281). Egg coat proteins are frequently rapidly evolving, and their divergence contributes to reproductive isolation and suggests a role in speciation (2, 4, 14, 16, 17, 51, 194, 207, 209, 282). With shotgun proteomics using liquid

chromatography-tandem mass spectrometry (LC-MS/MS), we find that the stickleback egg coat is comprised of homologs to the zona pellucida (ZP) glycoproteins ZP1 and ZP3 (Figure 5). Our findings are consistent with egg coat characterization in other fish, where ZP1, ZP3, and occasionally the ZP2 homolog ZPAX are the main structural proteins (101, 237, 240).

Egg coat glycoproteins, as with other reproductive proteins, are frequently glycosylated (276, 277). These carbohydrate modifications are thought to be involved in gamete recognition during fertilization, and to contribute to egg coat solubility (19, 35, 78, 278). Of the two stickleback ZP proteins, we find that only ZP3 is glycosylated. Incubation of stickleback egg coats with a *N*-glycanase resulted in loss of ZP3 carbohydrate staining by SDS-PAGE, and a reduction in its apparent molecular weight of ~4 kDa (Figure 7). This relatively large mass is unusual for an *N*-linked glycan in vertebrates, and may be indicative of a complex, tetra-antennary carbohydrate that could present a recognition surface for sperm, as has been demonstrated in other animals (2, 78, 283-286). There is only one potential *N*-linked glycosylation site in stickleback ZP3, ^{160}NVS^{162} (Figure 11, modeled in yellow), which is the same site found to be *N*-glycosylated in rainbow trout ZP3 (70, 101). In an alignment between rainbow trout and mammalian ZP3 proteins, Darie et al. (101) found that this particular *N*-linked glycosylation site is highly conserved to mammals. While it is interesting to note that ZP1 appears to have lost glycosylation entirely in stickleback, this is consistent with what has been described in other fish (12, 70, 101).

Disulfide bonds play important roles in protein folding and structural stability, particularly for secreted proteins, and are a defining characteristic of ZP module-containing proteins with their 8, 10, or 12 conserved, disulfide bonded cysteines (40, 237, 287). Stickleback egg coats are remarkably insoluble relative to other characterized egg coats, a biochemical feature that seems true of fish egg coats in general (98, 115, 288). For instance, we have found that stickleback egg

coats remain intact in the presence of 7 M urea, but dissolve better with the addition of a reducing agent, suggesting that disulfide bonds could contribute to their significant structural stability. Notably, the ZP module of ZP1-like proteins from fish contains two extra cysteine residues in a linker between the ZP-N and ZP-C domains (see Figure 9; (100, 101)). This interdomain linker has been implicated in homo- and heterodimeric assembly of ZP proteins, and it is possible that these additional cysteines play a role in fish egg coat stability (44, 236). To determine the pattern of disulfide bonding in stickleback ZP proteins, egg coats were treated with differing reduction and alkylation conditions prior to performing LC-MS/MS with dynamic exclusion turned off, to allow more quantitative peptide spectral counting. We found a consistent pattern of alkylatable cysteines present in stickleback egg coats, as detected by carbamidomethyl (CAM) modification of these residues by mass spectrometry (Figure 9, Tables 5 and 6). Typically these cysteines would be expected to participate in disulfide bonds, and should not be modifiable without reduction. Free cysteines suggest the potential for disulfide shuffling throughout the stickleback egg coat – in fact, nearly all cysteines in ZP1 and ZP3 were found to be CAM modified at least some of the time (Figure 9, denoted by dashed disulfide bonds). It is not clear whether these labile disulfide bonds are intra- or intermolecular, but free cysteines imply structural flexibility in the disulfide bonding of stickleback egg coats. Potential disulfide shuffling is especially apparent in the ZP-N domains of ZP1 and ZP3, the region of the ZP module known to be involved in ZP protein polymerization (41, 77). There are an odd number of cysteine residues in both stickleback ZP1 and ZP3, consistent with the proposed prevalence of disulfide shuffling.

In teleost fish, ZP genes are known to exhibit both ovarian and hepatic expression (31, 39, 98, 115, 240). To determine the site(s) of ZP synthesis in stickleback, primers were designed against *ZP1* and *ZP3* and amplified from both ovary and liver cDNA. ZP primers amplified

transcripts from liver cDNA far more robustly than from ovary cDNA, suggesting that in stickleback these genes are transcribed in the liver (Figure 8). This observation is supported by the very small number of ZP transcripts found in a stickleback ovary transcriptome generated with Pacific Biosciences long-read DNA sequencing (see Appendix). ZP protein products secreted from the liver make their way through the bloodstream to the ovary, where they assemble around developing oocytes. Both stickleback ZP1 and ZP3 have lost their canonical transmembrane domain (TMD), in agreement with this altered biosynthesis pattern. Although stickleback ZP proteins lack a TMD, they retain a C-terminal hydrophobic region typical of ZP proteins.

The polymerization of ZP proteins into the higher order structure of the egg coat is best characterized in the mouse, where the egg coat matrix consists of heterodimers of ZP2 and ZP3 that polymerize non-covalently into long fibrils interconnected by cross-links of ZP1 (34, 281, 289). While intramolecular disulfide bonds stabilize the native conformation of secreted ZP proteins, the mouse egg coat matrix also contains intermolecular disulfide bonds in the form of cross-linking ZP1 homodimers (73, 185, 236, 290, 291). Both ZP2 and ZP3 are required for egg coat formation, as ZP2 or ZP3 knockout mice fail to produce egg coats (73, 292). ZP1 knockout mice do form an egg coat, but it is loose and not interconnected and females are less fertile than wild-type (76, 292). It is interesting to note that ZP4 – a ZP1 homolog pseudogenized in mouse – can be substituted in place of ZP2 in transgenic mice so that ZP3/ZP4 heterodimers form the egg coat matrix rather than ZP2/ZP3 (188). This agrees with the observation that the structural function of ZP2 in mammals is performed by ZP1-like subunits in fish, which lack ZP2 (105, 236, 237, 240). It is also consistent with our finding that ZP1 and ZP3 constitute the stickleback egg coat matrix. ZP-N domains within ZP proteins are thought to facilitate egg coat polymerization, with cross-linking between filaments mediated by ZP1 (2, 9, 41).

As alluded to above, there are interesting changes in stickleback ZP protein architecture relative to what is known about other ZP proteins. Classical ZP protein architecture consists of a N-terminal signal sequence (SS) that marks them as secreted proteins; potential sequence upstream of the ZP module containing additional ZP-N domain repeats, or a P/Q rich-region and trefoil domain in ZP1-like proteins; the ZP module, with its paired ZP-N and ZP-C domains; a consensus furin cleavage site (CFCS) that allows cleavage of the C-terminal region; and a hydrophobic region or TMD (9, 11, 237). Changes to stickleback ZP protein architecture are highlighted in approximate order from N- to C-terminus (see Table 8 for summary). First, ZP1 proteins typically have a single N-terminal ZP-N domain repeat upstream of the ZP module, which stickleback ZP1 has lost (33). Stickleback ZP1 has also lost its fourth canonical cysteine in the ZP-N domain of its ZP module (Figure 9). Stickleback ZP1 has two additional cysteine residues, C_4 and C_5, in a linker between the ZP-N and ZP-C domains of its ZP module that are specific to fish (Figure 9, boxed in black; (100, 101, 237)). Stickleback ZP3 also has an additional cysteine residue, C_9, in its ZP-C domain (Figure 9, boxed in black). Both stickleback ZP proteins have lost their TMDs, likely as a consequence of their hepatic expression (12, 237, 293). Stickleback ZP1 appears to have lost all glycosylation, while stickleback ZP3 contains a single *N*-linked glycan in the linker between its ZP-N and ZP-C domains at a site well-conserved from fish to mammals (Figure 11, modeled in yellow; (12, 70, 101)). Finally, disulfide shuffling is prevalent in both stickleback ZP proteins, particularly within the ZP-N domains of their ZP modules, and particularly for ZP1 (see Figure 9). All cysteines in the ZP-N domain of stickleback ZP1 were modifiable with iodoacetamide in the absence of reducing agent, whereas only C_1 and C_4 of ZP3 were – C_2 and C_3 formed a stable disulfide bond.

Table 8. Summary of changes to stickleback ZP protein architecture relative to mammalian ZP proteins.

Egg coat protein	Stickleback architecture
ZP1	– N-terminal ZP-N domain lost – Two extra cysteines (C_4 and C_5) present in fish-specific ZP-N-ZP-C linker – Transmembrane domain lost – Glycosylation lost – Disulfide shuffling prevalent
ZP2	– Not present in fish
ZP3	– Extra cysteine (C_9) in ZP-C – Transmembrane domain lost – *N*-linked glycosylated – Disulfide shuffling prevalent

The role of the ZP-N domain in protein polymerization is not limited to reproductive proteins, and is conserved throughout eukaryotes (41, 43, 44, 46, 236). ZP-N/ZP-N interactions between ZP3 and ZP1/2/4 (depending on which ZP proteins are present) are thought to assemble into the structure of the egg coat, so it is notable that stickleback ZP1 has lost one of its two ZP-N domains with the loss of its canonical N-terminal ZP-N repeat. Similarly, ZP2, with its numerous N-terminal ZP-N repeats, is not found in fish (105, 236, 237, 240). Stickleback egg coats may compensate for the loss of these ZP-N polymerization domains with intermolecular, covalent disulfide cross-links arising from disulfide shuffling, which would be a departure from what has been characterized in other animals. The absence of a TMD in stickleback ZP proteins, and often in fish ZP proteins more generally, suggests that the topology of ZP proteins during egg coat assembly may be different in fish relative to mammals as well (70).

The evolution of stickleback *ZP3* under positive Darwinian selection also has interesting implications for stickleback egg coat architecture. In general, rapid evolution is a hallmark of reproductive proteins (199, 209). Numerous evolutionary forces have been attributed to the rapid evolution of reproductive proteins, including sperm competition, sexual conflict (at the cellular level, cryptic female choice), reinforcement, and pathogen resistance (14, 123, 192, 199, 209). Using a maximum likelihood method to assess ZP protein evolution across teleost fish, we find that *ZP3* has been subjected to positive Darwinian selection along the lineage while *ZP1* has not (Table 7). Rapid evolution in *ZP3* has also been found in mammals (14, 279, 280). ZP3 has nine rapidly evolving residues in teleosts: six that fall within its ZP-N domain, and three that fall within its ZP-C domain (Figure 11, denoted with red spheres on stickleback ZP3; see also Table 7). Notably, as few as ten amino acid changes in a sea urchin reproductive protein can lead to gametic incompatibility (137). Although it is interesting that *ZP1* has not experienced positive selection in teleosts, studies of mammalian *ZP1* similarly find no evidence of positive selection. ZP1 is thought to play a cross-linking role in mammalian egg coats, and stickleback ZP1 may be serving a similar structural function with its parallel evolutionary trajectory. We see many changes in stickleback ZP1 relative to other characterized ZP1 proteins, including the prevalence of disulfide shuffling (even relative to stickleback ZP3), the two extra cysteines that may be involved in homo- or heterodimeric ZP assembly, the loss of its N-terminal ZP-N domain, and the loss of glycosylation. These modifications hint at a conserved structural function, whereas stickleback ZP3 could be playing another role besides contributing to egg coat structure that necessitates evolutionary flexibility. In the mouse, ZP3 has been implicated as a receptor for sperm binding (294). *O*-glycans at S_{332} and S_{334} were identified as sperm ligands, although more recent work has demonstrated that these sites lack glycosylation *in vivo* and are tolerant to mutagenesis without affecting fertility,

calling into question the hypothesis of ZP3 as the primary mouse sperm receptor (73, 172, 174, 185). Regardless, amino acids in and around this "sperm-combining site" have been identified as under positive Darwinian selection in a diverse set of mammals (14, 279, 294). That *ZP3* is maintained under positive selection from stickleback to mammals is intriguing. Although a purely structural role has been suggested for fish ZP proteins given the presence of the micropyle in the egg coat, it is possible that the residues under selection in stickleback ZP3 participate in sperm recognition at the micropyle, particularly given the spatial clustering of the loops containing residues under selection (see Figure 11). These loops of positive selection in ZP3 would therefore be adaptive at the micropyle, but neutral in the remainder of ZP3 molecules forming the rest of the egg coat. The importance of fertilization likely creates a strong selective pressure that could drive rapid evolution, even if this rapid evolution has a functional consequence in only a very small percentage of molecules.

Taken together, our results suggest that the egg coats of stickleback fish are a uniquely protective structure relative to mammalian egg coats. Whereas mammalian sperm secrete acrosomal proteins to bind to the egg coat and create a hole at the point of contact, fish sperm lack an acrosome and enter the egg coat through a specialized channel, the micropyle (2, 12, 35, 111). It is conceivable that the presence of this structure has favored evolutionary events leading to an otherwise impenetrable egg coat: freed from the need to permit sperm access via transient, reversible ZP-N/ZP-N interactions, stickleback egg coats have evolved covalent cross-links arising from disulfide shuffling to stabilize the matrix. Given that fish eggs develop in external environments, such as the bottom of a lake or ocean, subject to high levels of mechanical stress – as well as potential pathogen exposure – a protective structural barrier might be evolutionarily favored (2, 4). Another mechanism for building impenetrable egg coats involves covalent cross-

linking of ZP proteins via the N-terminal P/Q-rich region of ZP1, by the action of a transglutaminase enzyme (70, 101, 115, 237, 288). These heterodimeric cross-links would not be reversed under reducing conditions, however, and so are unlikely to represent a significant contribution to egg coat structural stability the way intermolecular disulfide bonds in stickleback are. Correspondingly, only small amounts of these P/Q cross-linked heterodimers are detected by mass spectrometry in unfertilized rainbow trout eggs (70). On the other hand, these transglutaminase cross-links are likely important after fertilization, where they harden the egg coat to further reinforce the matrix and block polyspermy (2, 9, 12, 98, 115, 237).

In summary, there are unique biochemical attributes of fish ZP proteins that likely create a different set of protein-protein interactions for egg coat assembly and fertilization than has been characterized in other animals. The structure of the micropyle may underlie these changes. In teleost eggs, the inner micropylar opening directly adjoins the egg plasma membrane, creating what may be a specialized site for binding fertilizing sperm (111). The recently described zebrafish egg plasma membrane protein Bouncer – which permits cross-species fertilization between medaka and zebrafish, separated by 200 million years of evolution, expressing the medaka version of Bouncer – represents a possible candidate for sperm recognition at the egg plasma membrane (295). Our findings suggest that ZP3 in the egg coat may also contribute to sperm recognition at the micropyle, given its suggested role as a sperm receptor in mammals and its maintenance under positive Darwinian selection from teleost fish to mammals.

Chapter 3. CONCLUSIONS

The egg coats of metazoans share deep homology over potentially ~1 billion years of evolution as a result of their common structural feature, the ZP module. Despite the remarkable conservation of the egg's extracellular barrier, however, egg coat proteins (and their interacting partners on sperm) are frequently found to be rapidly evolving. Rapid evolution of reproductive proteins is likely to have clinical significance, as in the approximately 10% of *in vitro* fertilization (IVF) attempts that fail, with no known cause (225). It is possible that gametic incompatibilities driven by rapid evolution underlie these failures. Furthermore, identifying sites under positive Darwinian selection can help reveal the molecular basis of reproductive incompatibilities, with implications for infertility, reproductive isolation, and speciation (14, 35, 190). Most notably, adaptive evolution acting on functional domains in gamete recognition proteins can lead to reproductive isolation among closely related species, suggesting a mechanism for speciation to proceed (14, 116, 190, 194, 213).

Threespine stickleback fish have been fantastic model organisms for reproductive isolation and speciation, but so far the emphasis has been on ecological and behavioral explanations for the gradual accumulation of species barriers (249, 250, 296, 297). The fact that we find rapid evolution in a stickleback egg coat protein, ZP3, hints at a possible contribution of reproductive proteins to the process of stickleback speciation as well. Specifically, the presence of the micropyle in the egg coat may facilitate interactions between the rapidly evolving residues on ZP3 and sperm, with the importance of that recognition event driving positive selection across all ZP3 molecules in the egg coat, despite only a small percentage of them being localized to the micropyle and available for sperm binding.

3.1 FUTURE DIRECTIONS

3.1.1 *Reproductive proteins in stickleback species pairs*

Protein-protein interaction incompatibilities during fertilization caused by rapid evolution in sperm-egg recognition proteins can drive the speciation process (209). Stickleback species pairs exist throughout the Northern Hemisphere, as a result of an adaptive radiation of ancestral marine fish into novel freshwater environments that appeared after the end of the last ice age. These replicate potential speciation events show varying degrees of reproductive isolation and times since divergence, creating an ideal system for studying the role of sperm-egg recognition proteins in the speciation process. In our work we examined one stickleback population from Lake Union, Washington, USA, but expanding this research to reproductive proteins in other populations – particularly to species pairs at different stages along the speciation continuum – would give a sense of how general our findings are. Because ZP3 is a highly conserved protein across metazoans, we would expect our observations about ZP3 in Lake Union stickleback to generalize to other stickleback populations – especially because rapid evolution in ZP3 is found across teleosts. However, it would be instructive to investigate whether "speciation genes" exist within stickleback population pairs, as such genes could represent pre-zygotic barriers to hybridization that would reinforce species boundaries (298-303). Such speciation genes would be expected to show different patterns of evolution and introgression as compared to neutrally evolving loci in the genome, suggesting a means for their identification. Furthermore, if reproductive protein incompatibilities are involved in stickleback reproductive isolation, we would expect that population pairs exhibiting stronger reproductive isolation would show correspondingly stronger differentiation at putative speciation genes.

3.1.2 *ZP proteins in publicly available stickleback genomes*

There are a diversity of publicly available stickleback genomes and transcriptomes, reflecting the fact that threespine stickleback fish have become an excellent model system for many biological questions. It would be interesting to use these datasets as an orthogonal analysis to stickleback species pairs for characterizing ZP proteins, as well as other stickleback reproductive proteins. One potential bioinformatic complication is the whole genome duplication that has occurred at the base of the teleost radiation, resulting in many paralogous classes of genes. ZP genes, in particular, are known to be highly duplicated in teleosts as a result of both genome and gene duplications (3, 240). Regardless, assembling a dataset of *ZP3* genes across stickleback populations would allow us to search for rapid evolution in *ZP3* in stickleback specifically, rather than across teleosts as we have done so far.

3.1.3 *ZP3 in stickleback sperm-egg recognition*

Identifying the sperm protein in stickleback that interacts with ZP3 at the egg coat – potentially through affinity purification, or through signatures of molecular coevolution – would lend support to the hypothesis that ZP3 is involved in sperm-egg recognition in stickleback. A method called Evolutionary Rate Covariation (ERC), which uses patterns of coevolution to predict interacting protein partners from correlations between their branch-specific evolutionary rates, could be useful in identifying ZP3's interacting partner on sperm (304). ERC could also be used to find other potential pairs of sperm-egg recognition proteins in stickleback, and patterns of intergenic linkage disequilibrium could verify interacting partners as an independent line of evidence. If ZP3's gamete recognition partner on sperm could be identified, and was also under positive Darwinian selection, it would suggest a role for a coevolutionary arms race between

fertilization proteins in threespine stickleback reproductive isolation and speciation.

3.1.4 *Stickleback ZP3 glycan characterization*

Glycosylation is known to be an important post-translational modification for reproductive proteins, particularly in gamete recognition. Sea urchins, for instance, use oligosaccharides to conspecifically regulate the interaction of gametes (2). In ascidians, incorporation of unusual sugars in the egg coat facilitates the process of allorecognition used to distinguish self from nonself (155). Because stickleback ZP3 contains only one *N*-linked glycosylation site, and the ~4 kDa mass shift after deglycosylation suggests a highly branched and complex sugar, it would be fascinating – and relatively straightforward – to characterize this glycan using mass spectrometry. It would be especially interesting to determine whether the *N*-linked glycan differs between reproductively isolated stickleback populations, as might be expected from the literature, where modification of oligosaccharides (such as by sulfation or methylation) is known to ensure species-specificity in sperm-egg recognition (2).

3.1.5 *Stickleback ovary and testis transcriptomes*

Stickleback ovary and testis transcriptomes from Lake Union, Washington, USA were generated using Pacific Biosciences Iso-Seq technology to capture full-length cDNA transcripts. We had intended to use the ovary transcriptome as a database for mass spectrometry searches, but this was prior to learning that stickleback ZP genes are primarily expressed in the liver. Our long-read transcriptomes from male and female reproductive tissues would be interesting to mine for other reproductive proteins involved in fertilization in stickleback. Secreted proteins containing a N-terminal signal peptide would be of particular interest, as sperm-egg recognition proteins would be expected to be displayed on the surface of the gamete plasma membranes.

3.1.6 *Micropyle characterization*

The micropyle is a fascinating and enigmatic egg coat structure that we would like to better characterize. Micropyles form from a cytoplasmic projection of a specialized cell known as the micropylar cell, which shrinks and retracts after egg coat deposition to leave behind a hollow channel for sperm entry (5, 111). In some fish species, the micropyle is surrounded by a glycoprotein that directs sperm to the micropylar canal in a calcium-dependent manner (110). This substance, known as the micropylar sperm attractant or MISA, can be stained with fluorescein isothiocyanate (FITC) conjugated to wheat germ agglutinin (WGA) (305). We had originally hoped to isolate stickleback micropyles for mass spectrometry characterization using FITC-WGA staining and laser capture microdissection, but stickleback lack a MISA and their micropyles cannot be selectively stained in this manner. Still, it would be possible to perform laser capture microdissection and mass spectrometry characterization of the micropyle in a fish species that does have a MISA, and then to bioinformatically search for homologous proteins in stickleback. With this data, we might be able to determine whether there are any specialized proteins at the micropyle entrance that could facilitate stickleback fertilization and reproductive isolation.

3.1.7 *Bouncer molecular evolution*

Although this work has focused on threespine stickleback egg coat proteins as the first barrier sperm encounter during fertilization, the egg plasma membrane creates another set of protein-protein interactions that are essential to the final step of fertilization, the fusion of the two gametes. The recently described zebrafish egg plasma membrane protein Bouncer, with its role in species-specific fertilization, represents an exciting candidate to follow up on in the evolution of threespine stickleback reproductive isolation (295). Bouncer is a Ly6/uPAR domain-containing

protein, members of which have diverse biological functions including roles in fertilization, envenomation, embryogenesis, morphogenesis, and immunity (306). The recurrence of Ly6/uPAR domains across the genome makes bioinformatic searches for Bouncer homologs in stickleback and other teleosts challenging, particularly since Bouncer is a single exon gene. Synteny could be used to help identify truly homologous genes. Publicly available transcriptomes could also be searched, with the caveat that whole genome duplication at the base of the teleost radiation has created paralogous genes that may complicate analyses – for instance, Bouncer is known to have two paralogs in both medaka and carp (295, 306). If a database of high-confidence Bouncer homologs across teleosts could be generated, it would be fascinating to look for rapid evolution in this essential fertilization protein. It would also be intriguing to investigate Bouncer in stickleback species pairs, as sequence divergence in Bouncer may contribute to threespine stickleback reproductive isolation.

3.1.8 *Sperm competition in stickleback reproductive isolation*

The existence of sperm competition between reproductively isolated stickleback populations would also be interesting to follow up on. Evaluating the extent of fertilization compatibility in light of population divergence, as well as the presence or absence of a conspecific sperm preference, would add to our knowledge of the state of stickleback gametic isolation. Currently, it is known that fertilization in stickleback takes significantly longer than in other teleosts, an observation that has been attributed to sexual conflict (307). Stickleback males build nests during the breeding season and entice females to lay eggs in this nest. After the female has deposited her eggs and left, the courting male enters the nest and fertilizes the eggs. “Sneaker” stickleback males sometimes lie in wait during this courtship, attempting to steal fertilizations by spawning in the nest and introducing sperm competition. From the female reproductive standpoint,

slow fertilization of eggs makes sense to ensure complete fertilization or to encourage multiple paternity. Males, however, benefit from instantaneous fertilization. Given that opportunities for sperm competition are known in stickleback ecology, it would be fascinating to better characterize this phenomenon. Reciprocal crosses between reproductively isolated populations, where an equal amount of sperm from both populations is added to the same clutch of eggs, could be used to evaluate the extent of sperm competition. Determining the final share of paternity between conspecific and heterospecific males would give a sense of the amount of gametic isolation across stickleback population pairs.

BIBLIOGRAPHY

1. Killingbeck EE, Swanson WJ. Egg Coat Proteins Across Metazoan Evolution. Curr Top Dev Biol. 2018;130:443-88.

2. Wong JL, Wessel GM. Defending the zygote: search for the ancestral animal block to polyspermy. Curr Top Dev Biol. 2006;72:1-151.

3. Conner SJ, Lefievre L, Hughes DC, Barratt CL. Cracking the egg: increased complexity in the zona pellucida. Hum Reprod. 2005 May;20(5):1148-52.

4. Shu L, Suter MJ-, Räsänen K. Evolution of egg coats: linking molecular biology and ecology. Mol Ecol. 2015;24(16):4052-73.

5. Lombardi J. Comparative Vertebrate Reproduction. Boston: Kluwer Academic Publishers; 1998.

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

7. Han L, Monne M, Okumura H, Schwend T, Cherry AL, Flot D, et al. Insights into egg coat assembly and egg-sperm interaction from the X-ray structure of full-length ZP3. Cell. 2010 Oct 29;143(3):404-15.

8. Menkhorst E, Selwood L. Vertebrate extracellular preovulatory and postovulatory egg coats. Biol Reprod. 2008 Nov;79(5):790-7.

9. Monne M, Han L, Jovine L. Tracking down the ZP domain: From the mammalian zona pellucida to the molluscan vitelline envelope. Semin Reprod Med. 2006 Sep;24(4):204-16.

10. Monne M, Jovine L. A structural view of egg coat architecture and function in fertilization. Biol Reprod. 2011 Oct;85(4):661-9.

11. Jovine L, Darie CC, Litscher ES, Wassarman PM. Zona pellucida domain proteins. Annu Rev Biochem. 2005;74:83-114.

12. Litscher ES, Wassarman PM. Egg extracellular coat proteins: from fish to mammals. Histol Histopathol. 2007 Mar;22(3):337-47.

13. Aagaard JE, Yi X, MacCoss MJ, Swanson WJ. Rapidly evolving zona pellucida domain proteins are a major component of the vitelline envelope of abalone eggs. Proc Natl Acad Sci U S A. 2006 Nov 14;103(46):17302-7.

14. Turner LM, Hoekstra HE. Causes and consequences of the evolution of reproductive proteins. Int J Dev Biol. 2008;52(5-6):769-80.

15. Findlay GD, Swanson WJ. Proteomics enhances evolutionary and functional analysis of reproductive proteins. Bioessays. 2010 Jan;32(1):26-36.

16. Palumbi SR. Speciation and the evolution of gamete recognition genes: pattern and process. Heredity (Edinb). 2009 Jan;102(1):66-76.

17. Vacquier VD, Swanson WJ. Selection in the rapid evolution of gamete recognition proteins in marine invertebrates. Cold Spring Harb Perspect Biol. 2011 Nov 1;3(11):a002931.

18. Meslin C, Mugnier S, Callebaut I, Laurin M, Pascal G, Poupon A, et al. Evolution of genes involved in gamete interaction: evidence for positive selection, duplications and losses in vertebrates. PLoS One. 2012;7(9):e44548.

19. Claw KG, Swanson WJ. Evolution of the egg: new findings and challenges. Annu Rev Genomics Hum Genet. 2012;13:109-25.

20. Wassarman PM. Zona pellucida glycoproteins. J Biol Chem. 2008 Sep 5;283(36):24285-9.

21. Litscher ES, Wassarman PM. Evolution, structure, and synthesis of vertebrate egg-coat proteins. Trends Dev Biol. 2014;8:65-76.

22. Wilburn DB, Swanson WJ. The "ZP domain" is not one, but likely two independent domains. Mol Reprod Dev. 2017 Apr;84(4):284-5.

23. Jovine L, Qi H, Williams Z, Litscher ES, Wassarman PM. A duplicated motif controls assembly of zona pellucida domain proteins. Proc Natl Acad Sci U S A. 2004 Apr 20;101(16):5922-7.

24. Litscher ES, Williams Z, Wassarman PM. Zona pellucida glycoprotein ZP3 and fertilization in mammals. Mol Reprod Dev. 2009 Oct;76(10):933-41.

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

26. Xu Q, Li G, Cao L, Wang Z, Ye H, Chen X, et al. Proteomic characterization and evolutionary analyses of zona pellucida domain-containing proteins in the egg coat of the cephalochordate, Branchiostoma belcheri. BMC Evol Biol. 2012 Dec 8;12:239,2148-12-239.

27. Putnam NH, Butts T, Ferrier DE, Furlong RF, Hellsten U, Kawashima T, et al. The amphioxus genome and the evolution of the chordate karyotype. Nature. 2008 Jun 19;453(7198):1064-71.

28. Sawada H, Sakai N, Abe Y, Tanaka E, Takahashi Y, Fujino J, et al. Extracellular ubiquitination and proteasome-mediated degradation of the ascidian sperm receptor. Proc Natl Acad Sci U S A. 2002 Feb 5;99(3):1223-8.

29. Yamada L, Saito T, Taniguchi H, Sawada H, Harada Y. Comprehensive egg coat proteome of the ascidian Ciona intestinalis reveals gamete recognition molecules involved in self-sterility. J Biol Chem. 2009 Apr 3;284(14):9402-10.

30. Smith J, Paton IR, Hughes DC, Burt DW. Isolation and mapping the chicken zona pellucida genes: an insight into the evolution of orthologous genes in different species. Mol Reprod Dev. 2005 Feb;70(2):133-45.

31. Spargo SC, Hope RM. Evolution and nomenclature of the zona pellucida gene family. Biol Reprod. 2003 Feb;68(2):358-62.

32. Wilburn DB, Swanson WJ. Gamete Structure: Egg, Comparative Vertebrate. In: Skinner MA, editor. Encyclopedia of Reproduction. 2nd ed. Academic Press; 2018. p. 204-9.

33. Callebaut I, Mornon JP, Monget P. Isolated ZP-N domains constitute the N-terminal extensions of Zona Pellucida proteins. Bioinformatics. 2007 Aug 1;23(15):1871-4.

34. Wassarman PM, Litscher ES. Chapter Thirty-One - A Bespoke Coat for Eggs: Getting Ready for Fertilization. Current Topics in Developmental Biology. 2016 2016;117(Supplement C):539-52.

35. Wilburn DB, Swanson WJ. From molecules to mating: Rapid evolution and biochemical studies of reproductive proteins. J Proteomics. 2016 Mar 1;135:12-25.

36. Bork P. Complement components C1r/C1s, bone morphogenic protein 1 and Xenopus laevis developmentally regulated protein UVS.2 share common repeats. FEBS Lett. 1991 Apr 22;282(1):9-12.

37. Bork P, Beckmann G. The CUB Domain: A Widespread Module in Developmentally Regulated Proteins. Journal of Molecular Biology. 1993 20 May 1993;231(2):539-45.

38. Topfer-Petersen E, Calvete JJ. Sperm-associated protein candidates for primary zona pellucida-binding molecules: structure-function correlations of boar spermadhesins. J Reprod Fertil Suppl. 1996;50:55-61.

39. Sano K, Kawaguchi M, Watanabe S, Nagakura Y, Hiraki T, Yasumasu S. Inferring the Evolution of Teleostean zp Genes Based on Their Sites of Expression. Journal of Experimental Zoology Part B: Molecular and Developmental Evolution. 2013;320(5):332-43.

40. Bork P, Sander C. A large domain common to sperm receptors (Zp2 and Zp3) and TGF-beta type III receptor. FEBS Lett. 1992 Apr 6;300(3):237-40.

41. Jovine L, Janssen WG, Litscher ES, Wassarman PM. The PLAC1-homology region of the ZP domain is sufficient for protein polymerisation. BMC Biochem. 2006 Apr 6;7:11,2091-7-11.

42. Wilburn DB, Swanson WJ. From molecules to mating: Rapid evolution and biochemical studies of reproductive proteins. J Proteomics. 2015 Jun 11.

43. Swanson WJ, Aagaard JE, Vacquier VD, Monne M, Sadat Al Hosseini H, Jovine L. The molecular basis of sex: linking yeast to human. Mol Biol Evol. 2011 Jul;28(7):1963-6.

44. Bokhove M, Nishimura K, Brunati M, Han L, de Sanctis D, Rampoldi L, et al. A structured interdomain linker directs self-polymerization of human uromodulin. Proc Natl Acad Sci U S A. 2016 Feb 9;113(6):1552-7.

45. Monne M, Han L, Schwend T, Burendahl S, Jovine L. Crystal structure of the ZP-N domain of ZP3 reveals the core fold of animal egg coats. Nature. 2008 Dec 4;456(7222):653-7.

46. Jovine L, Qi H, Williams Z, Litscher E, Wassarman PM. The ZP domain is a conserved module for polymerization of extracellular proteins. Nat Cell Biol. 2002 Jun;4(6):457-61.

47. Chung YD, Zhu J, Han Y, Kernan MJ. nompA encodes a PNS-specific, ZP domain protein required to connect mechanosensory dendrites to sensory structures. Neuron. 2001 Feb;29(2):415-28.

48. Fernandes I, Chanut-Delalande H, Ferrer P, Latapie Y, Waltzer L, Affolter M, et al. Zona pellucida domain proteins remodel the apical compartment for localized cell shape changes. Dev Cell. 2010 Jan 19;18(1):64-76.

49. Heiman MG, Shaham S. DEX-1 and DYF-7 establish sensory dendrite length by anchoring dendritic tips during cell migration. Cell. 2009 Apr 17;137(2):344-55.

50. Plaza S, Chanut-Delalande H, Fernandes I, Wassarman PM, Payre F. From A to Z: apical structures and zona pellucida-domain proteins. Trends Cell Biol. 2010 Sep;20(9):524-32.

51. Hart MW, Sunday JM, Popovic I, Learning KJ, Konrad CM. Incipient speciation of sea star populations by adaptive gamete recognition coevolution. Evolution. 2014 May;68(5):1294-305.

52. Jagadeeshan S, Singh RS. Rapid evolution of outer egg membrane proteins in the Drosophila melanogaster subgroup: a case of ecologically driven evolution of female reproductive traits. Mol Biol Evol. 2007 Apr;24(4):929-38.

53. Papantonis A, Swevers L, Iatrou K. Chorion genes: a landscape of their evolution, structure, and regulation. Annu Rev Entomol. 2015 Jan 7;60:177-94.

54. Carter JM, Baker SC, Pink R, Carter DR, Collins A, Tomlin J, et al. Unscrambling butterfly oogenesis. BMC Genomics. 2013 Apr 26;14:283,2164-14-283.

55. Marinotti O, Ngo T, Kojin BB, Chou SP, Nguyen B, Juhn J, et al. Integrated proteomic and transcriptomic analysis of the Aedes aegypti eggshell. BMC Dev Biol. 2014 Apr 5;14:15,213X-14-15.

56. Irles P, Belles X, Piulachs MD. Brownie, a gene involved in building complex respiratory devices in insect eggshells. PLoS One. 2009 Dec 16;4(12):e8353.

57. Irles P, Piulachs MD. Citrus, a key insect eggshell protein. Insect Biochem Mol Biol. 2011 Feb;41(2):101-8.

58. Doren S, Landsberger N, Dwyer N, Gold L, Blanchette-Mackie J, Dean J. Incorporation of mouse zona pellucida proteins into the envelope of Xenopus laevis oocytes. Dev Genes Evol. 1999 Jun;209(6):330-9.

59. Sugiyama H, Murata K, Iuchi I, Nomura K, Yamagami K. Formation of mature egg envelope subunit proteins from their precursors (choriogenins) in the fish, Oryzias latipes: loss of partial C-terminal sequences of the choriogenins. J Biochem. 1999 Mar;125(3):469-75.

60. Hyllner SJ, Westerlund L, Olsson PE, Schopen A. Cloning of rainbow trout egg envelope proteins: members of a unique group of structural proteins. Biol Reprod. 2001 Mar;64(3):805-11.

61. Callard IP, Riley D, Perez L. Vertebrate vitellogenesis: molecular model for multihormonal control of gene regulation. Prog Clin Biol Res. 1990;342:343-8.

62. Callard IP, Riley D, Perez L. Vitellogenesis in reptiles as a model for mammalian sex-differentiated hepatic protein synthesis. J Exp Zool Suppl. 1990;4:106-11.

63. Polzonetti-Magni AM, Mosconi G, Soverchia L, Kikuyama S, Carnevali O. Multihormonal Control of Vitellogenesis in Lower Vertebrates. International Review of Cytology. 2004 2004;239(Supplement C):1-46.

64. Schneider WJ. Vitellogenin Receptors: Oocyte-Specific Members of the Low-Density Lipoprotein Receptor Supergene Family. International Review of Cytology. 1996 1996;166(Supplement C):103-37.

65. Bausek N, Waclawek M, Schneider WJ, Wohlrab F. The major chicken egg envelope protein ZP1 is different from ZPB and is synthesized in the liver. J Biol Chem. 2000 Sep 15;275(37):28866-72.

66. Okumura H, Kohno Y, Iwata Y, Mori H, Aoki N, Sato C, et al. A newly identified zona pellucida glycoprotein, ZPD, and dimeric ZP1 of chicken egg envelope are involved in sperm activation on sperm-egg interaction. Biochem J. 2004 Nov 15;384(Pt 1):191-9.

67. Chang YS, Lu LF, Lai CY, Kou YH, Huang FL. Purification, characterization, and molecular cloning of an outer layer protein of carp fertilization envelope. Mol Reprod Dev. 1999 Oct;54(2):186-93.

68. Qi H, Williams Z, Wassarman PM. Secretion and assembly of zona pellucida glycoproteins by growing mouse oocytes microinjected with epitope-tagged cDNAs for mZP2 and mZP3. Mol Biol Cell. 2002 Feb;13(2):530-41.

69. Litscher ES, Qi H, Wassarman PM. Mouse zona pellucida glycoproteins mZP2 and mZP3 undergo carboxy-terminal proteolytic processing in growing oocytes. Biochemistry. 1999 Sep 21;38(38):12280-7.

70. Darie CC, Biniossek ML, Gawinowicz MA, Milgrom Y, Thumfart JO, Jovine L, et al. Mass spectrometric evidence that proteolytic processing of rainbow trout egg vitelline envelope proteins takes place on the egg. J Biol Chem. 2005 Nov 11;280(45):37585-98.

71. Iconomidou VA, Chryssikos DG, Gionis V, Pavlidis MA, Paipetis A, Hamodrakas SJ. Secondary structure of chorion proteins of the teleostean fish Dentex dentex by ATR FT-IR and FT-Raman spectroscopy. J Struct Biol. 2000 Nov;132(2):112-22.

72. Nara M, Yonezawa N, Shimada T, Takahashi K, Tanokura M, Yumoto F, et al. Fourier transform infrared spectroscopic analysis of the intact zona pellucida of the mammalian egg: changes in the secondary structure of bovine zona pellucida proteins during fertilization. Exp Biol Med (Maywood). 2006 Feb;231(2):166-71.

73. Dean J. Reassessing the molecular biology of sperm-egg recognition with mouse genetics. Bioessays. 2004 Jan;26(1):29-38.

74. Wassarman PM, Mortillo S. Structure of the mouse egg extracellular coat, the zona pellucida. Int Rev Cytol. 1991;130:85-110.

75. Wassarman PM. Zona pellucida glycoproteins. Annu Rev Biochem. 1988;57:415-42.

76. Rankin T, Talbot P, Lee E, Dean J. Abnormal zonae pellucidae in mice lacking ZP1 result in early embryonic loss. Development. 1999 Sep;126(17):3847-55.

77. Greve JM, Wassarman PM. Mouse egg extracellular coat is a matrix of interconnected filaments possessing a structural repeat. J Mol Biol. 1985 Jan 20;181(2):253-64.

78. Brivio MF, Bassi R, Cotelli F. Identification and characterization of the major components of the Oncorhynchus mykiss egg chorion. Mol Reprod Dev. 1991 Jan;28(1):85-93.

79. Iwamoto K, Ikeda K, Yonezawa N, Noguchi S, Kudo K, Hamano S, et al. Disulfide formation in bovine zona pellucida glycoproteins during fertilization: evidence for the involvement of cystine cross-linkages in hardening of the zona pellucida. J Reprod Fertil. 1999 Nov;117(2):395-402.

80. Wassarman PM, Litscher ES. Biogenesis of the mouse egg's extracellular coat, the zona pellucida. Curr Top Dev Biol. 2013;102:243-66.

81. Liu C, Litscher ES, Mortillo S, Sakai Y, Kinloch RA, Stewart CL, et al. Targeted disruption of the mZP3 gene results in production of eggs lacking a zona pellucida and infertility in female mice. Proc Natl Acad Sci U S A. 1996 May 28;93(11):5431-6.

82. Rankin TL, O'Brien M, Lee E, Wigglesworth K, Eppig J, Dean J. Defective zonae pellucidae in Zp2-null mice disrupt folliculogenesis, fertility and development. Development. 2001 Apr;128(7):1119-26.

83. Rankin T, Familari M, Lee E, Ginsberg A, Dwyer N, Blanchette-Mackie J, et al. Mice homozygous for an insertional mutation in the Zp3 gene lack a zona pellucida and are infertile. Development. 1996 Sep;122(9):2903-10.

84. Wassarman PM, Qi H, Litscher ES. Mutant female mice carrying a single mZP3 allele produce eggs with a thin zona pellucida, but reproduce normally. Proc Biol Sci. 1997 Mar 22;264(1380):323-8.

85. Ichikawa Y, Matsuzaki M, Hiyama G, Mizushima S, Sasanami T. Sperm-Egg Interaction during Fertilization in Birds. The Journal of Poultry Science. 2016;advpub.

86. Serizawa M, Kinoshita M, Rodler D, Tsukada A, Ono H, Yoshimura T, et al. Oocytic expression of zona pellucida protein ZP4 in Japanese quail (Coturnix japonica). Anim Sci J. 2011 Apr;82(2):227-35.

87. Sasanami T, Pan J, Mori M. Expression of perivitelline membrane glycoprotein ZP1 in the liver of Japanese quail (Coturnix japonica) after in vivo treatment with diethylstilbestrol. J Steroid Biochem Mol Biol. 2003 Jan;84(1):109-16.

88. Takeuchi Y, Cho R, Iwata Y, Nishimura K, Kato T, Aoki N, et al. Morphological and biochemical changes of isolated chicken egg-envelope during sperm penetration: degradation of the 97-kilodalton glycoprotein is involved in sperm-driven hole formation on the egg-envelope. Biol Reprod. 2001 Mar;64(3):822-30.

89. Watanabe A, Onitake K. The urodele egg-coat as the apparatus adapted for the internal fertilization. Zoolog Sci. 2002 Dec;19(12):1341-7.

90. Hedrick JL. Anuran and pig egg zona pellucida glycoproteins in fertilization and early development. Int J Dev Biol. 2008;52(5-6):683-701.

91. Tian J, Gong H, Lennarz WJ. Xenopus laevis sperm receptor gp69/64 glycoprotein is a homolog of the mammalian sperm receptor ZP2. Proc Natl Acad Sci U S A. 1999 Feb 2;96(3):829-34.

92. Kubo H, Matsushita M, Kotani M, Kawasaki H, Saido TC, Kawashima S, et al. Molecular basis for oviductin-mediated processing from gp43 to gp41, the predominant glycoproteins of Xenopus egg envelopes. Dev Genet. 1999;25(2):123-9.

93. Lindsay LL, Yang JC, Hedrick JL. Identification and characterization of a unique Xenopus laevis egg envelope component, ZPD. Dev Growth Differ. 2002 Jun;44(3):205-12.

94. Kubo H, Kawano T, Tsubuki S, Kotani M, Kawasaki H, Kawashima S. Egg envelope glycoprotein gp37 as a Xenopus homolog of mammalian ZP1, based on cDNA cloning. Dev Growth Differ. 2000 Aug;42(4):419-27.

95. Kubo H, Kawano T, Tsubuki S, Kawashima S, Katagiri C, Suzuki A. A major glycoprotein of Xenopus egg vitelline envelope, gp41, is a frog homolog of mammalian ZP3. Dev Growth Differ. 1997 Aug;39(4):405-17.

96. Lindsay LL, Wallace MA, Hedrick JL. A hatching enzyme substrate in the Xenopus laevis egg envelope is a high molecular weight ZPA homolog. Dev Growth Differ. 2001 Jun;43(3):305-13.

97. Watanabe A, Takayama-Watanabe E. In silico identification of the genes for sperm-egg interaction in the internal fertilization of the newt Cynops pyrrhogaster. Int J Dev Biol. 2014;58(10-12):873-9.

98. Arukwe A, Goksoyr A. Eggshell and egg yolk proteins in fish: hepatic proteins for the next generation: oogenetic, population, and evolutionary implications of endocrine disruption. Comp Hepatol. 2003 Mar 6;2(1):4.

99. Sugiyama H, Iuchi I. Molecular structure and hardening of egg envelope in fish. Recent Research Developments in Comparative Biochemistry & Physiology. 2000;1(1):139-61.

100. Darie CC, Litscher ES, Wassarman PM. Structure, Processing, and Polymerization of Rainbow Trout Egg Vitelline Envelope Proteins. In: Popescu C, Zamfir AD, Dinca N, editors. Applications of Mass Spectrometry in Life Safety. Dordrecht: Springer Netherlands; 2008. p. 23-36.

101. Darie CC, Biniossek ML, Jovine L, Litscher ES, Wassarman PM. Structural characterization of fish egg vitelline envelope proteins by mass spectrometry. Biochemistry. 2004 Jun 15;43(23):7459-78.

102. Mold DE, Kim IF, Tsai CM, Lee D, Chang CY, Huang RC. Cluster of genes encoding the major egg envelope protein of zebrafish. Mol Reprod Dev. 2001 Jan;58(1):4-14.

103. Chang YS, Hsu CC, Wang SC, Tsao CC, Huang FL. Molecular cloning, structural analysis, and expression of carp ZP2 gene. Mol Reprod Dev. 1997 Mar;46(3):258-67.

104. Kanamori A, Naruse K, Mitani H, Shima A, Hori H. Genomic organization of ZP domain containing egg envelope genes in medaka (Oryzias latipes). Gene. 2003 Feb 13;305(1):35-45.

105. Hughes DC. ZP genes in avian species illustrate the dynamic evolution of the vertebrate egg envelope. Cytogenet Genome Res. 2007;117(1-4):86-91.

106. Berois N, Arezo MJ, Papa NG. Gamete interactions in teleost fish: the egg envelope. Basic studies and perspectives as environmental biomonitor. Biol Res. 2011;44(2):119-24.

107. Iwamatsu T, Onitake K, Matsuyama K, Satoh M, Yukawa S. Effect of Micropylar Morphology and Size on Rapid Sperm Entry into the Eggs of the Medaka. Zool Sci. 1997 04/01; 2017/12;14(4):623-8.

108. Hart NH, Donovan M. Fine structure of the chorion and site of sperm entry in the egg of Brachydanio. J Exp Zool. 1983;227(2):277-96.

109. Amanze D, Iyengar A. The micropyle: a sperm guidance system in teleost fertilization. Development. 1990 Jun;109(2):495-500.

110. Yanagimachi R, Cherr G, Matsubara T, Andoh T, Harumi T, Vines C, et al. Sperm attractant in the micropyle region of fish and insect eggs. Biol Reprod. 2013 Feb 28;88(2):47.

111. Hart NH. Fertilization in teleost fishes: mechanisms of sperm-egg interactions. Int Rev Cytol. 1990;121:1-66.

112. Marlow FL, Mullins MC. Bucky ball functions in Balbiani body assembly and animal-vegetal polarity in the oocyte and follicle cell layer in zebrafish. Dev Biol. 2008 Sep 1;321(1):40-50.

113. Hart NH, Pietri R, Donovan M. The structure of the chorion and associated surface filaments in Oryzias--evidence for the presence of extracellular tubules. J Exp Zool. 1984 May;230(2):273-96.

114. Cherr GN, Clark WH,Jr. Induction of the acrosomal reaction in sperm from the white sturgeon, Acipenser transmontanus. Adv Exp Med Biol. 1986;207:235-49.

115. Yamagami K, Hamazaki TS, Yasumasu S, Masuda K, Iuchi I. Molecular and cellular basis of formation, hardening, and breakdown of the egg envelope in fish. Int Rev Cytol. 1992;136:51-92.

116. Lyon JD, Vacquier VD. Interspecies Chimeric Sperm Lysins Identify Regions Mediating Species-Specific Recognition of the Abalone Egg Vitelline Envelope. Developmental Biology. 1999 1 October 1999;214(1):151-9.

117. Swanson WJ, Vacquier VD. The abalone egg vitelline envelope receptor for sperm lysin is a giant multivalent molecule. Proc Natl Acad Sci U S A. 1997 Jun 24;94(13):6724-9.

118. Galindo BE, Moy GW, Swanson WJ, Vacquier VD. Full-length sequence of VERL, the egg vitelline envelope receptor for abalone sperm lysin. Gene. 2002 Apr 17;288(1-2):111-7.

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

120. Swanson WJ, Vacquier VD. Concerted evolution in an egg receptor for a rapidly evolving abalone sperm protein. Science. 1998 Jul 31;281(5377):710-2.

121. Galindo BE, Vacquier VD, Swanson WJ. Positive selection in the egg receptor for abalone sperm lysin. Proc Natl Acad Sci U S A. 2003 Apr 15;100(8):4639-43.

122. Swanson WJ, Aquadro CF, Vacquier VD. 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. 2001 Mar;18(3):376-83.

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

124. Hart MW, Stover DA, Guerra V, Mozaffari SV, Ober C, Mugal CF, et al. Positive selection on human gamete-recognition genes. PeerJ. 2018 jan;6:e4259.

125. Hellberg ME, Dennis AB, Arbour-Reily P, Aagaard JE, Swanson WJ. The Tegula tango: a coevolutionary dance of interacting, positively selected sperm and egg proteins. Evolution. 2012 Jun;66(6):1681-94.

126. Buckland-Nicks J, Koss R, Chia FS. Fertilization in a chiton: acrosome-mediated sperm-egg fusion. Gamete Res. 1988 Nov;21(3):199-212.

127. Alliegro MC, Wright DA. Polyspermy inhibition in the oyster, Crassostrea virginica. J Exp Zool. 1983 Jul;227(1):127-37.

128. Mozingo NM, Vacquier VD, Chandler DE. Structural features of the abalone egg extracellular matrix and its role in gamete interaction during fertilization. Mol Reprod Dev. 1995 Aug;41(4):493-502.

129. Buckland-Nicks J. Hull Cupules of Chiton Eggs: Parachute Structures and Sperm Focusing Devices? Biol Bull. 1993;184(3):269-76.

130. Focarelli R, Renieri T, Rosati F. Polarized site of sperm entrance in the egg of a freshwater bivalve, Unio elongatulus. Developmental Biology. 1988 June 1988;127(2):443-51.

131. Focarelli R, Rosati F. The 220-kDa Vitelline Coat Glycoprotein Mediates Sperm Binding in the Polarized Egg of Unio elongatulus through O-Linked Oligosaccharides. Developmental Biology. 1995 October 1995;171(2):606-14.

132. Gache C, Niman HL, Vacquier VD. Monoclonal antibodies to the sea urchin egg vitelline layer inhibit fertilization by blocking sperm adhesion. Experimental Cell Research. 1983 August 1983;147(1):75-84.

133. Longo FJ. Effects of concanavalin A on the fertilization of sea urchin eggs. Developmental Biology. 1981 February 1981;82(1):197-202.

134. Niman HL, Hough-Evans BR, Vacquier VD, Britten RJ, Lerner RA, Davidson EH. Proteins of the sea urchin egg vitelline layer. Developmental Biology. 1984 April 1984;102(2):390-401.

135. Haley SA, Wessel GM. Proteolytic cleavage of the cell surface protein p160 is required for detachment of the fertilization envelope in the sea urchin. Developmental Biology. 2004 1 August 2004;272(1):191-202.

136. Wong JL, Wessel GM. Rendezvin: An Essential Gene Encoding Independent, Differentially Secreted Egg Proteins That Organize the Fertilization Envelope Proteome after Self-Association. Mol Biol Cell. 2006 09/18;17(12):5241-52.

137. Zigler KS, McCartney MA, Levitan DR, Lessios HA. Sea urchin bindin divergence predicts gamete compatibility. Evolution. 2005 Nov;59(11):2399-404.

138. Kamei N, Glabe CG. The species-specific egg receptor for sea urchin sperm adhesion is EBR1,a novel ADAMTS protein. Genes Dev. 2003 Oct 15;17(20):2502-7.

139. Degrugillier ME, Leopold RA. Ultrastructure of sperm penetration of house fly eggs. Journal of Ultrastructure Research. 1976 September 1976;56(3):312-25.

140. Mouzaki DG, Zarani FE, Margaritis LH. Structure and morphogenesis of the eggshell and micropylar apparatus in the olive fly, Dacus oleae (Diptera: Tephritidae). J Morphol. 1991 Jul;209(1):39-52.

141. Turner FR, Mahowald AP. Scanning electron microscopy of Drosophila embryogenesis: 1. The structure of the egg envelopes and the formation of the cellular blastoderm. Developmental Biology. 1976 May 1976;50(1):95-108.

142. Pascucci T, Perrino J, Mahowald AP, Waring GL. Eggshell Assembly in Drosophila: Processing and Localization of Vitelline Membrane and Chorion Proteins. Developmental Biology. 1996 1 August 1996;177(2):590-8.

143. Neubaum DM, Wolfner MF. Wise, winsome, or weird? Mechanisms of sperm storage in female animals. Curr Top Dev Biol. 1999;41:67-97.

144. Li J, Hodgeman BA, Christensen BM. Involvement of peroxidase in chorion hardening in Aedes aegypti. Insect Biochemistry and Molecular Biology. 1996 March 1996;26(3):309-17.

145. Parks S, Wakimoto B, Spradling A. Replication and expression of an X-linked cluster of Drosophila chorion genes. Dev Biol. 1986 Sep;117(1):294-305.

146. Spradling AC, Mahowald AP. Amplification of genes for chorion proteins during oogenesis in Drosophila melanogaster. Proc Natl Acad Sci U S A. 1980 Feb;77(2):1096-100.

147. Spradling AC. The organization and amplification of two chromosomal domains containing Drosophila chorion genes. Cell. 1981 Nov;27(1 Pt 2):193-201.

148. Griffin-Shea R, Thireos G, Kafatos FC. Organization of a cluster of four chorion genes in Drosophila and its relationship to developmental expression and amplification. Dev Biol. 1982 Jun;91(2):325-36.

149. Konstandi OA, Papassideri IS, Stravopodis DJ, Kenoutis CA, Hasan Z, Katsorchis T, et al. The enzymatic component of Drosophila melanogaster chorion is the Pxd peroxidase. Insect Biochemistry and Molecular Biology. 2005 September 2005;35(9):1043-57.

150. Kafatos FC, Spoerel N, Mitsialis SA, Nguyen HT, Romano C, Lingappa JR, et al. Developmental Control and Evolution in the Chorion Gene Families Of Insects. Advances in Genetics. 1987 1987;24(Supplement C):223-42.

151. Hamodrakas SJ, Kamitsos EI, Papanikolaou A. Laser-Raman spectroscopic studies of the eggshell (chorion) of Bombyx mori. International Journal of Biological Macromolecules. 1984 December 1984;6(6):333-6.

152. Nadel MR, Kafatos FC. Specific protein synthesis in cellular differentiation: IV. The chorion proteins of Bombyx mori and their program of synthesis. Developmental Biology. 1980 1 March 1980;75(1):26-40.

153. Amenya DA, Chou W, Li J, Yan G, Gershon PD, James AA, et al. Proteomics reveals novel components of the Anopheles gambiae eggshell. J Insect Physiol. 2010 Oct;56(10):1414-9.

154. Delsuc F, Brinkmann H, Chourrout D, Philippe H. Tunicates and not cephalochordates are the closest living relatives of vertebrates. Nature. 2006 02/23;439:965.

155. Sawada H, Tanaka E, Ban S, Yamasaki C, Fujino J, Ooura K, et al. Self/nonself recognition in ascidian fertilization: vitelline coat protein HrVC70 is a candidate allorecognition molecule. Proc Natl Acad Sci U S A. 2004 Nov 2;101(44):15615-20.

156. Lambert CC, Goode CA. Glycolipid linkage of a polyspermy blocking glycosidase to the ascidian egg surface. Dev Biol. 1992 Nov;154(1):95-100.

157. Ban S, Harada Y, Yokosawa H, Sawada H. Highly polymorphic vitelline-coat protein HaVC80 from the ascidian, Halocynthia aurantium: structural analysis and involvement in self/nonself recognition during fertilization. Dev Biol. 2005 Oct 15;286(2):440-51.

158. Kurn U, Sommer F, Bosch TC, Khalturin K. In the urochordate Ciona intestinalis zona pellucida domain proteins vary among individuals. Dev Comp Immunol. 2007;31(12):1242-54.

159. Baibakov B, Boggs NA, Yauger B, Baibakov G, Dean J. Human sperm bind to the N-terminal domain of ZP2 in humanized zonae pellucidae in transgenic mice. J Cell Biol. 2012 Jun 25;197(7):897-905.

160. Vacquier VD. Evolution of gamete recognition proteins. Science. 1998 Sep 25;281(5385):1995-8.

161. Neill AT, Vacquier VD. Ligands and receptors mediating signal transduction in sea urchin spermatozoa. Reproduction. 2004 Feb;127(2):141-9.

162. Okamura F, Nishiyama H. The passage of spermatozoa through the vitelline membrane in the domestic fowl, Gallus gallus. Cell Tissue Res. 1978 Apr 28;188(3):497-508.

163. Tulsiani DR, Abou-Haila A, Loeser CR, Pereira BM. The biological and functional significance of the sperm acrosome and acrosomal enzymes in mammalian fertilization. Exp Cell Res. 1998 May 1;240(2):151-64.

164. Wassarman PM. Mammalian fertilization: molecular aspects of gamete adhesion, exocytosis, and fusion. Cell. 1999 Jan 22;96(2):175-83.

165. Gwatkin RB, Williams DT. Receptor activity of the hamster and mouse solubilized zona pellucida before and after the zona reaction. J Reprod Fertil. 1977 Jan;49(1):55-9.

166. Bleil JD, Wassarman PM. Mammalian sperm-egg interaction: identification of a glycoprotein in mouse egg zonae pellucidae possessing receptor activity for sperm. Cell. 1980 Jul;20(3):873-82.

167. Arnoult C, Zeng Y, Florman HM. ZP3-dependent activation of sperm cation channels regulates acrosomal secretion during mammalian fertilization. J Cell Biol. 1996 Aug;134(3):637-45.

168. Bleil JD, Wassarman PM. Sperm-egg interactions in the mouse: sequence of events and induction of the acrosome reaction by a zona pellucida glycoprotein. Dev Biol. 1983 Feb;95(2):317-24.

169. Jungnickel MK, Sutton KA, Wang Y, Florman HM. Phosphoinositide-dependent pathways in mouse sperm are regulated by egg ZP3 and drive the acrosome reaction. Dev Biol. 2007 Apr 1;304(1):116-26.

170. Bleil JD, Wassarman PM. Autoradiographic visualization of the mouse egg's sperm receptor bound to sperm. J Cell Biol. 1986 Apr;102(4):1363-71.

171. Bleil JD, Greve JM, Wassarman PM. Identification of a secondary sperm receptor in the mouse egg zona pellucida: role in maintenance of binding of acrosome-reacted sperm to eggs. Dev Biol. 1988 Aug;128(2):376-85.

172. Florman HM, Wassarman PM. O-linked oligosaccharides of mouse egg ZP3 account for its sperm receptor activity. Cell. 1985 May;41(1):313-24.

173. Chen J, Litscher ES, Wassarman PM. Inactivation of the mouse sperm receptor, mZP3, by site-directed mutagenesis of individual serine residues located at the combining site for sperm. Proc Natl Acad Sci U S A. 1998 May 26;95(11):6193-7.

174. Boja ES, Hoodbhoy T, Fales HM, Dean J. Structural characterization of native mouse zona pellucida proteins using mass spectrometry. J Biol Chem. 2003 Sep 5;278(36):34189-202.

175. Williams SA, Xia L, Cummings RD, McEver RP, Stanley P. Fertilization in mouse does not require terminal galactose or N-acetylglucosamine on the zona pellucida glycans. J Cell Sci. 2007 Apr 15;120(Pt 8):1341-9.

176. Liu C, Litscher ES, Wassarman PM. Transgenic mice with reduced numbers of functional sperm receptors on their eggs reproduce normally. Mol Biol Cell. 1995 05;6(5):577-85.

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

178. Yauger B, Boggs NA, Dean J. Human ZP4 is not sufficient for taxon-specific sperm recognition of the zona pellucida in transgenic mice. Reproduction. 2011 12/20;141(3):313-9.

179. Bedford JM. Sperm/egg interaction: the specificity of human spermatozoa. Anat Rec. 1977 Aug;188(4):477-87.

180. Lanzendorf SE, Holmgren WJ, Johnson DE, Scobey MJ, Jeyendran RS. Hemizona assay for measuring zona binding in the lowland gorilla. Mol Reprod Dev. 1992 Apr;31(4):264-7.

181. Bleil JD, Beall CF, Wassarman PM. Mammalian sperm-egg interaction: fertilization of mouse eggs triggers modification of the major zona pellucida glycoprotein, ZP2. Dev Biol. 1981 Aug;86(1):189-97.

182. Moller CC, Wassarman PM. Characterization of a proteinase that cleaves zona pellucida glycoprotein ZP2 following activation of mouse eggs. Dev Biol. 1989 Mar;132(1):103-12.

183. Quesada V, Sanchez LM, Alvarez J, Lopez-Otin C. Identification and characterization of human and mouse ovastacin: a novel metalloproteinase similar to hatching enzymes from arthropods, birds, amphibians, and fish. J Biol Chem. 2004 Jun 18;279(25):26627-34.

184. Burkart AD, Xiong B, Baibakov B, Jimenez-Movilla M, Dean J. Ovastacin, a cortical granule protease, cleaves ZP2 in the zona pellucida to prevent polyspermy. J Cell Biol. 2012 Apr 2;197(1):37-44.

185. Avella MA, Xiong B, Dean J. The molecular basis of gamete recognition in mice and humans. Mol Hum Reprod. 2013 May;19(5):279-89.

186. Tian J, Gong H, Thomsen GH, Lennarz WJ. Gamete Interactions in Xenopus laevis: Identification of Sperm Binding Glycoproteins in the Egg Vitelline Envelope. J Cell Biol. 1997 12/06;136(5):1099-108.

187. Lindsay LL, Hedrick JL. Proteolysis of Xenopus laevis egg envelope ZPA triggers envelope hardening. Biochem Biophys Res Commun. 2004 Nov 12;324(2):648-54.

188. Avella MA, Baibakov B, Dean J. A single domain of the ZP2 zona pellucida protein mediates gamete recognition in mice and humans. J Cell Biol. 2014 Jun 23;205(6):801-9.

189. Dranginis AM, Rauceo JM, Coronado JE, Lipke PN. A biochemical guide to yeast adhesins: glycoproteins for social and antisocial occasions. Microbiol Mol Biol Rev. 2007 Jun;71(2):282-94.

190. Clark NL, Aagaard JE, Swanson WJ. Evolution of reproductive proteins from animals and plants. Reproduction. 2006 Jan;131(1):11-22.

191. Kresge N, Vacquier VD, Stout CD. Abalone lysin: the dissolving and evolving sperm protein. Bioessays. 2001 Jan;23(1):95-103.

192. Kosman ET, Levitan DR. Sperm competition and the evolution of gametic compatibility in externally fertilizing taxa. MHR: Basic science of reproductive medicine. 2014 12/01;20(12):1190-7.

193. Ruppert E, Fox R, Barnes R. Invertebrate Zoology: A Functional Evolutionary Approach. 7th ed. Brooks Cole; 2004.

194. Coyne JA, Orr HA. Speciation. Sunderland, MA: Sinaur Associates, Inc.; 2004.

195. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 2007 Aug;24(8):1586-91.

196. Goldman N, Yang Z. A codon-based model of nucleotide substitution for protein-coding DNA sequences. Mol Biol Evol. 1994 Sep;11(5):725-36.

197. Wong WS, Yang Z, Goldman N, Nielsen R. Accuracy and power of statistical methods for detecting adaptive evolution in protein coding sequences and for identifying positively selected sites. Genetics. 2004 Oct;168(2):1041-51.

198. Yang Z. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. Mol Biol Evol. 1998 May;15(5):568-73.

199. Swanson W, Vacquier V. Reproductive Protein Evolution. ; 2002.

200. Howard DJ. Conspecific Sperm and Pollen Precedence and Speciation. Annu Rev Ecol Syst. 1999 11/01; 2017/12;30(1):109-32.

201. Vacquier VD, Swanson WJ, Lee YH. Positive Darwinian selection on two homologous fertilization proteins: what is the selective pressure driving their divergence? J Mol Evol. 1997;44 Suppl 1:S15-22.

202. Gavrilets S. Rapid evolution of reproductive barriers driven by sexual conflict. Nature. 2000 Feb 24;403(6772):886-9.

203. Palumbi SR. All males are not created equal: Fertility differences depend on gamete recognition polymorphisms in sea urchins. Proceedings of the National Academy of Sciences. 1999 October 26;96(22):12632-7.

204. Rice WR, Holland B. The enemies within: intergenomic conflict, interlocus contest evolution (ICE), and the intraspecific Red Queen. Behav Ecol Sociobiol. 1997 07/01;41(1):1-10.

205. Lessios HA. Reproductive isolation between species of sea urchins. Bull Mar Sci. 2007;81(2):191-208.

206. Dobzhansky T. Speciation as a Stage in Evolutionary Divergence. Am Nat. 1940 07/01; 2017/12;74(753):312-21.

207. Vieira A, Miller DJ. Gamete interaction: is it species-specific? Mol Reprod Dev. 2006 Nov;73(11):1422-9.

208. Yanagimachi R, Phillips DM. The status of acrosomal caps of hamster spermatozoa immediately before fertilization in vivo. Gamete Res. 1984;9(1):1-19.

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

210. Metz EC, Robles-Sikisaka R, Vacquier VD. Nonsynonymous substitution in abalone sperm fertilization genes exceeds substitution in introns and mitochondrial DNA. Proceedings of the National Academy of Sciences. 1998 09/01;95(18):10676-81.

211. Grayson P, Civetta A. Positive Selection and the Evolution of izumo Genes in Mammals. International Journal of Evolutionary Biology. 2012;2012:7.

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

213. Sainudiin R, Wong WSW, Yogeeswaran K, Nasrallah JB, Yang Z, Nielsen R. Detecting Site-Specific Physicochemical Selective Pressures: Applications to the Class I HLA of the Human Major Histocompatibility Complex and the SRK of the Plant Sporophytic Self-Incompatibility System. J Mol Evol. 2005 03/01;60(3):315-26.

214. Levitan DR. Contemporary evolution of sea urchin gamete-recognition proteins: experimental evidence of density-dependent gamete performance predicts shifts in allele frequencies over time. Evolution. 2012 Jun;66(6):1722-36.

215. Levitan DR, Ferrell DL. Selection on gamete recognition proteins depends on sex, density, and genotype frequency. Science. 2006 Apr 14;312(5771):267-9.

216. Levitan DR, Stapper AP. Simultaneous positive and negative frequency-dependent selection on sperm bindin, a gamete recognition protein in the sea urchin Strongylocentrotus purpuratus. Evolution. 2010 Mar 1;64(3):785-97.

217. Evans JP, Marshall DJ. Male-by-female interactions influence fertilization success and mediate the benefits of polyandry in the sea urchin Heliocidaris erythrogramma. Evolution. 2005 Jan;59(1):106-12.

218. Clark NL, Findlay GD, Yi X, MacCoss MJ, Swanson WJ. Duplication and selection on abalone sperm lysin in an allopatric population. Mol Biol Evol. 2007 Sep;24(9):2081-90.

219. Torgerson DG, Kulathinal RJ, Singh RS. Mammalian sperm proteins are rapidly evolving: evidence of positive selection in functionally diverse genes. Mol Biol Evol. 2002 Nov;19(11):1973-80.

220. Springer SA, Crespi BJ. Adaptive gamete-recognition divergence in a hybridizing Mytilus population. Evolution. 2007 Apr;61(4):772-83.

221. Metz EC, Kane RE, Yanagimachi H, Palumbi SR. Fertilization Between Closely Related Sea Urchins Is Blocked by Incompatibilities During Sperm-Egg Attachment and Early Stages of Fusion. Biol Bull. 1994 08/01; 2017/12;187(1):23-34.

222. Palumbi SR, Metz EC. Strong reproductive isolation between closely related tropical sea urchins (genus Echinometra). Mol Biol Evol. 1991 Mar;8(2):227-39.

223. Palumbi SR, Lessios HA. Evolutionary animation: how do molecular phylogenies compare to Mayr's reconstruction of speciation patterns in the sea? Proc Natl Acad Sci U S A. 2005 May 3;102 Suppl 1:6566-72.

224. Wilburn DB, Tuttle LM, Klevit RE, Swanson WJ. Solution structure of sperm lysin yields novel insights into molecular dynamics of rapid protein evolution. Proceedings of the National Academy of Sciences of the United States of America. 2018.

225. Liu DY, Clarke GN, Martic M, Garrett C, Baker HW. Frequency of disordered zona pellucida (ZP)-induced acrosome reaction in infertile men with normal semen analysis and normal spermatozoa-ZP binding. Hum Reprod. 2001 Jun;16(6):1185-90.

226. Oehninger S, Hinsch E, Pfisterer S, Veeck LL, Kolm P, Schill WB, et al. Use of a specific zona pellucida (ZP) protein 3 antiserum as a clinical marker for human ZP integrity and function. Fertil Steril. 1996 Jan;65(1):139-45.

227. Nayernia K, Adham IM, Shamsadin R, Muller C, Sancken U, Engel W. Proacrosin-deficient mice and zona pellucida modifications in an experimental model of multifactorial infertility. Mol Hum Reprod. 2002 May;8(5):434-40.

228. Sun YP, Xu Y, Cao T, Su YC, Guo YH. Zona pellucida thickness and clinical pregnancy outcome following in vitro fertilization. Int J Gynaecol Obstet. 2005 Jun;89(3):258-62.

229. Mannikko M, Tormala RM, Tuuri T, Haltia A, Martikainen H, Ala-Kokko L, et al. Association between sequence variations in genes encoding human zona pellucida glycoproteins and fertilization failure in IVF. Hum Reprod. 2005 Jun;20(6):1578-85.

230. Pokkyla RM, Lakkakorpi JT, Nuojua-Huttunen SH, Tapanainen JS. Sequence variations in human ZP genes as potential modifiers of zona pellucida architecture. Fertil Steril. 2011 Jun 30;95(8):2669-72.

231. Paterson M, Jennings ZA, Wilson MR, Aitken RJ. The contraceptive potential of ZP3 and ZP3 peptides in a primate model. J Reprod Immunol. 2002 Jan;53(1-2):99-107.

232. Avella MA, Baibakov BA, Jimenez-Movilla M, Sadusky AB, Dean J. ZP2 peptide beads select human sperm in vitro, decoy mouse sperm in vivo, and provide reversible contraception. Science Translational Medicine. 2016;8(336):336ra60-.

233. Killingbeck EE, Wilburn DB, Merrihew GE, MacCoss MJ, Swanson WJ. Proteomics support the threespine stickleback egg coat as a protective oocyte envelope. bioRxiv. 2020 Cold Spring Harbor Laboratory:2020.03.04.976316.

234. Murata K, Conte FS, McInnis E, Fong TH, Cherr GN. Identification of the origin and localization of chorion (egg envelope) proteins in an ancient fish, the white sturgeon, Acipenser transmontanus. Biol Reprod. 2014 Jun;90(6):132.

235. Han L, Monne M, Okumura H, Schwend T, Cherry AL, Flot D, et al. Insights into egg coat assembly and egg-sperm interaction from the X-ray structure of full-length ZP3. Cell. 2010 Oct 29;143(3):404-15.

236. Bokhove M, Jovine L. Chapter Thirteen - Structure of Zona Pellucida Module Proteins. Current Topics in Developmental Biology. 2018 2018;130:413-42.

237. Litscher ES, Wassarman PM. The Fish Egg's Zona Pellucida. Curr Top Dev Biol. 2018;130:275-305.

238. Sasanami T, Ohtsuki M, Ishiguro T, Matsushima K, Hiyama G, Kansaku N, et al. Zona Pellucida Domain of ZPB1 controls specific binding of ZPB1 and ZPC in Japanese quail (Coturnix japonica). Cells Tissues Organs. 2006;183(1):41-52.

239. Okumura H, Aoki N, Sato C, Nadano D, Matsuda T. Heterocomplex formation and cell-surface accumulation of hen's serum zona pellucida B1 (ZPB1)with ZPC expressed by a mammalian cell line (COS-7): a possible initiating step of egg-envelope matrix construction. Biol Reprod. 2007 Jan;76(1):9-18.

240. Conner SJ, Hughes DC. Analysis of fish ZP1/ZPB homologous genes--evidence for both genome duplication and species-specific amplification models of evolution. Reproduction. 2003 Sep;126(3):347-52.

241. Sano K, Kawaguchi M, Katano K, Tomita K, Inokuchi M, Nagasawa T, et al. Comparison of Egg Envelope Thickness in Teleosts and its Relationship to the Sites of ZP Protein Synthesis. J Exp Zool B Mol Dev Evol. 2017 May;328(3):240-58.

242. Wilburn DB, Bowen KE, Doty KA, Arumugam S, Lane AN, Feldhoff PW, et al. Structural insights into the evolution of a sexy protein: novel topology and restricted backbone flexibility in a hypervariable pheromone from the red-legged salamander, Plethodon shermani. PLoS One. 2014 May 21;9(5):e96975.

243. Palmer MR, McDowall MH, Stewart L, Ouaddi A, MacCoss MJ, Swanson WJ. Mass spectrometry and next-generation sequencing reveal an abundant and rapidly evolving abalone sperm protein. Mol Reprod Dev. 2013 Jun;80(6):460-5.

244. Findlay GD, Yi X, Maccoss MJ, Swanson WJ. Proteomics reveals novel Drosophila seminal fluid proteins transferred at mating. PLoS Biol. 2008 Jul 29;6(7):e178.

245. Nielsen R. Molecular signatures of natural selection. Annu Rev Genet. 2005;39:197-218.

246. Peichel CL. Fishing for the secrets of vertebrate evolution in threespine sticklebacks. Dev Dyn. 2005 Dec;234(4):815-23.

247. Bell MA, Foster SA. The evolutionary biology of the threespine stickleback. Oxford ; New York : Oxford University Press; 1994.

248. Jones FC, Grabherr MG, Chan YF, Russell P, Mauceli E, Johnson J, et al. The genomic basis of adaptive evolution in threespine sticklebacks. Nature. 2012 04/05;484(7392):55-61.

249. McKinnon JS, Rundle HD. Speciation in nature: the threespine stickleback model systems. Trends in Ecology & Evolution. 2002 10/1;17(10):480-8.

250. Hendry AP, Bolnick DI, Berner D, Peichel CL. Along the speciation continuum in sticklebacks. J Fish Biol. 2009 Nov;75(8):2000-36.

251. Hanks JH, Wallace RE. Relation of oxygen and temperature in the preservation of tissues by refrigeration. Proc Soc Exp Biol Med. 1949 Jun;71(2):196-200.

252. Schagger H, von Jagow G. Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. Anal Biochem. 1987 Nov 1;166(2):368-79.

253. Eng JK, Jahan TA, Hoopmann MR. Comet: an open-source MS/MS sequence database search tool. Proteomics. 2013 Jan;13(1):22-4.

254. Eng JK, Hoopmann MR, Jahan TA, Egertson JD, Noble WS, MacCoss MJ. A Deeper Look into Comet—Implementation and Features. J Am Soc Mass Spectrom. 2015 11/01;26(11):1865-74.

255. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. Nat Biotechnol. 2011 05/15;29(7):644-52.

256. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, et al. De novo transcript sequence reconstruction from RNA-Seq: reference generation and analysis with Trinity. Nature protocols. 2013 07/11;8(8):10.1038/nprot.2013.084.

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

258. Zhang B, Chambers MC, Tabb DL. Proteomic parsimony through bipartite graph analysis improves accuracy and transparency. J Proteome Res. 2007 Sep;6(9):3549-57.

259. MacDonald RJ, Swift GH, Przybyla AE, Chirgwin JM. Isolation of RNA using guanidinium salts. Methods Enzymol. 1987;152:219-27.

260. Xu H, Freitas MA. MassMatrix: a database search program for rapid characterization of proteins and peptides from tandem mass spectrometry data. Proteomics. 2009 Mar;9(6):1548-55.

261. Xu H, Freitas MA. A mass accuracy sensitive probability based scoring algorithm for database searching of tandem mass spectrometry data. BMC Bioinformatics. 2007 Apr 20;8:133,2105-8-133.

262. Xu H, Yang L, Freitas MA. A robust linear regression based algorithm for automated evaluation of peptide identifications from shotgun proteomics by use of reversed-phase liquid chromatography retention time. BMC Bioinformatics. 2008 Aug 19;9:347,2105-9-347.

263. Xu H, Freitas MA. Monte carlo simulation-based algorithms for analysis of shotgun proteomic data. J Proteome Res. 2008 Jul;7(7):2605-15.

264. Xu H, Zhang L, Freitas MA. Identification and characterization of disulfide bonds in proteins and peptides from tandem MS data by use of the MassMatrix MS/MS search engine. J Proteome Res. 2008 Jan;7(1):138-44.

265. Xu H, Hsu PH, Zhang L, Tsai MD, Freitas MA. Database search algorithm for identification of intact cross-links in proteins and peptides using tandem mass spectrometry. J Proteome Res. 2010 Jul 2;9(7):3384-93.

266. Betancur-R R, Wiley EO, Arratia G, Acero A, Bailly N, Miya M, et al. Phylogenetic classification of bony fishes. BMC Evolutionary Biology. 2017 07/06;17(1):162.

267. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990 Oct 5;215(3):403-10.

268. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014 05/01;30(9):1312-3.

269. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Molecular systems biology [Internet]. 2011;7:539.

270. Yang Z. PAML: a program package for phylogenetic analysis by maximum likelihood. Comput Appl Biosci. 1997 Oct;13(5):555-6.

271. Yang Z, Wong WS, Nielsen R. Bayes empirical bayes inference of amino acid sites under positive selection. Mol Biol Evol. 2005 Apr;22(4):1107-18.

272. Qian B, Raman S, Das R, Bradley P, McCoy AJ, Read RJ, et al. High-resolution structure prediction and the crystallographic phase problem. Nature. 2007 Nov 8;450(7167):259-64.

273. Mandell DJ, Coutsias EA, Kortemme T. Sub-angstrom accuracy in protein loop reconstruction by robotics-inspired conformational sampling. Nat Methods. 2009 Aug;6(8):551-2.

274. Ceroni A, Dell A, Haslam SM. The GlycanBuilder: a fast, intuitive and flexible software tool for building and displaying glycan structures. Source Code Biol Med. 2007 Aug 7;2:3,0473-2-3.

275. Paoletti AC, Parmely TJ, Tomomori-Sato C, Sato S, Zhu D, Conaway RC, et al. Quantitative proteomic analysis of distinct mammalian Mediator complexes using normalized spectral abundance factors. Proc Natl Acad Sci U S A. 2006 Dec 12;103(50):18928-33.

276. Gagneux P, Varki A. Evolutionary considerations in relating oligosaccharide diversity to biological function. glycob. 1999;9(8):747-55.

277. Varki A. Nothing in Glycobiology Makes Sense, except in the Light of Evolution. Cell. 2006 8 September 2006;126(5):841-5.

278. Bausek N, Ruckenbauer HH, Pfeifer S, Schneider WJ, Wohlrab F. Interaction of sperm with purified native chicken ZP1 and ZPC proteins. Biol Reprod. 2004 Aug;71(2):684-90.

279. Swanson WJ, Yang Z, Wolfner MF, Aquadro CF. Positive Darwinian selection drives the evolution of several female reproductive proteins in mammals. Proc Natl Acad Sci U S A. 2001 Feb 27;98(5):2509-14.

280. Jansa SA, Lundrigan BL, Tucker PK. Tests for positive selection on immune and reproductive genes in closely related species of the murine genus mus. J Mol Evol. 2003 Mar;56(3):294-307.

281. Nixon B, Aitken RJ, McLaughlin EA. New insights into the molecular mechanisms of sperm-egg interaction. Cell Mol Life Sci. 2007 Jul;64(14):1805-23.

282. Nosil P. Ecological speciation. Oxford: Oxford : Oxford University Press; 2012.

283. North SJ, Hitchen PG, Haslam SM, Dell A. Mass spectrometry in the analysis of N-linked and O-linked glycans. Current Opinion in Structural Biology. 2009 October 2009;19(5):498-506.

284. Wilburn DB, Bowen KE, Feldhoff PW, Feldhoff RC. Proteomic analyses of courtship pheromones in the redback salamander, Plethodon cinereus. J Chem Ecol. 2014 Aug;40(8):928-39.

285. Katsumata T, Noguchi S, Yonezawa N, Tanokura M, Nakano M. Structural characterization of the N-linked carbohydrate chains of the zona pellucida glycoproteins from bovine ovarian and fertilized eggs. Eur J Biochem. 1996 Sep 1;240(2):448-53.

286. Vo LH, Yen T, Macher BA, Hedrick JL. Identification of the ZPC Oligosaccharide Ligand Involved in Sperm Binding and the Glycan Structures of Xenopus laevis Vitelline Envelope Glycoproteins1. Biol Reprod. 2003 12/01; 3/19;69(6):1822-30.

287. Sevier CS, Kaiser CA. Formation and transfer of disulphide bonds in living cells. Nat Rev Mol Cell Biol. 2002 Nov;3(11):836-47.

288. Oppen-Berntsen DO, Helvik JV, Walther BT. The major structural proteins of cod (Gadus morhua) eggshells and protein crosslinking during teleost egg hardening. Dev Biol. 1990 Feb;137(2):258-65.

289. Wassarman PM, Litscher ES. Chapter Ten - The Mouse Egg's Zona Pellucida. Current Topics in Developmental Biology. 2018 2018;130:331-56.

290. Epifano O, Liang L, Dean J. Mouse Zp1 Encodes a Zona Pellucida Protein Homologous to Egg Envelope Proteins in Mammals and Fish. Journal of Biological Chemistry. 1995 November 10;270(45):27254-8.

291. Bleil JD, Wassarman PM. Structure and function of the zona pellucida: Identification and characterization of the proteins of the mouse oocyte's zona pellucida. Developmental Biology. 1980 April 1980;76(1):185-202.

292. Wassarman PM, Litscher ES. Mammalian fertilization: the egg's multifunctional zona pellucida. Int J Dev Biol. 2008;52(5-6):665-76.

293. Wang H, Gong Z. Characterization of two zebrafish cDNA clones encoding egg envelope proteins ZP2 and ZP3. Biochim Biophys Acta. 1999 Jul 7;1446(1-2):156-60.

294. Kinloch RA, Sakai Y, Wassarman PM. Mapping the mouse ZP3 combining site for sperm by exon swapping and site-directed mutagenesis. Proc Natl Acad Sci U S A. 1995 Jan 3;92(1):263-7.

295. Herberg S, Gert KR, Schleiffer A, Pauli A. The Ly6/uPAR protein Bouncer is necessary and sufficient for species-specific fertilization. Science. 2018 Sep 7;361(6406):1029-33.

296. Hendry AP, Peichel C, Matthews B, Boughman JW, Nosil P. Stickleback research: the now and the next. Evol Ecol Res. 2013 Feb 2013;15(2):111-41.

297. Rundle HD, Nagel L, Wenrick Boughman J, Schluter D. Natural selection and parallel speciation in sympatric sticklebacks. Science. 2000 Jan 14;287(5451):306-8.

298. Harrison RG, Larson EL. Hybridization, introgression, and the nature of species boundaries. J Hered. 2014;105 Suppl 1:795-809.

299. Marques DA, Lucek K, Meier JI, Mwaiko S, Wagner CE, Excoffier L, et al. Genomics of Rapid Incipient Speciation in Sympatric Threespine Stickleback. PLoS Genet. 2016 Feb 29;12(2):e1005887.

300. Rieseberg LH, Whitton J, Gardner K. Hybrid zones and the genetic architecture of a barrier to gene flow between two sunflower species. Genetics. 1999 06;152(2):713-27.

301. Beaumont MA, Balding DJ. Identifying adaptive genetic divergence among populations from genome scans. Mol Ecol. 2004 Apr;13(4):969-80.

302. Nosil P, Feder JL. Widespread yet heterogeneous genomic divergence. Mol Ecol. 2012 Jun;21(12):2829-32.

303. Maheshwari S, Barbash DA. The genetics of hybrid incompatibilities. Annu Rev Genet. 2011;45:331-55.

304. Clark NL, Alani E, Aquadro CF. Evolutionary rate covariation reveals shared functionality and coexpression of genes. Genome Res. 2012 Apr;22(4):714-20.

305. Yanagimachi R, Harumi T, Matsubara H, Yan W, Yuan S, Hirohashi N, et al. Chemical and physical guidance of fish spermatozoa into the egg through the micropyledagger,double dagger. Biol Reprod. 2017 Apr 1;96(4):780-99.

306. Leth JM, Leth-Espensen KZ, Kristensen KK, Kumari A, Lund Winther AM, Young SG, et al. Evolution and Medical Significance of LU Domain-Containing Proteins. Int J Mol Sci. 2019 Jun 5;20(11):10.3390/ijms20112760.

307. Bakker TC, Zbinden M, Frommen JG, Weiss A, Largiader CR. Slow fertilization of stickleback eggs: the result of sexual conflict? BMC Ecol. 2006 May 19;6:7.

APPENDIX: DATA RESOURCES GENERATED

Threespine stickleback fish ovary and testis transcriptomes were pooled and sequenced with Pacific Biosciences Single Molecule, Real-Time (SMRT) sequencing, with Iso-Seq analysis used to generate full-length cDNA sequences.

Table 9. Summary of PacBio Iso-Seq data.

Number of consensus reads	Number of poly-A reads	Number of non-full-length reads	Number of full-length reads	Number of full-length non-chimeric reads	Mean full-length non-chimeric read length
77,249	34,191	47,697	29,524	29,271	868

Table 10. Summary of PacBio Iso-Seq transcripts. Note that high-quality isoforms are defined as those with accuracy ≥ 99%.

Number of unpolished consensus isoforms	Number of polished low-quality isoforms	Number of polished high-quality isoforms	Mean unpolished consensus isoforms read length
17,932	15,021	2,849	887

VITA

Emily Erskine Killingbeck was born in La Jolla, California and grew up in the coastal town of Solana Beach, where she and her twin sister Sarah spent many magical days exploring the San Elijo Lagoon, Torrey Pines State Reserve, and the local beaches with their parents. When she and her sister were nine years old the family moved north to the Central Coast of California, replacing beach days with winding drives up Highway 1 and weekend visits to property the family owned in the Santa Lucia mountains, called Madrona after the red madrone tress that grow throughout the property. Their parents (both field biology majors at UC Santa Barbara) instilled in them a love of California and its native flora and fauna, complete with proper taxonomic nomenclature, and both girls grew up loving science. Emily met her future husband, Jim Schneidereit, in the Templeton school district, and all three graduated from Templeton High School in 2008. Emily moved north to UC Berkeley for college, graduating with a Bachelor of Arts in Molecular and Cell Biology with an emphasis in Genetics, Genomics, and Development in 2012. While at UC Berkeley, she spent two years in the lab of Craig T. Miller studying quantitative trait loci underlying threespine stickleback fish skeletal phenotypes. Emily also spent a year as an intern at Sangamo BioSciences in Richmond, California before moving further north up the West Coast to Seattle in 2014, pursuing graduate research at the University of Washington. She received her Doctorate of Philosophy from the Department of Genome Sciences in 2020, still working with threespine stickleback fish but now taking an evolutionary approach to the study of their reproductive proteins in the lab of Willie J. Swanson. In her free time, Emily enjoys reading, hiking, skiing, and spending time with her husband and their adorable black cat, Poe.