RESEARCH ARTICLE

Molecular Reproduction & Development

Proteomics support the threespine stickleback egg coat as a protective oocyte envelope

Emily E. Killingbeck | Damien B. Wilburn | Gennifer E. Merrihew | Michael J. MacCoss | Willie J. Swanson

Department of Genome Sciences, University of Washington, Seattle, Washington, USA

Correspondence
Emily E. Killingbeck, Department of Genome Sciences, University of WashingtonFoege Bldg S-250, Box 355065, 3720 15th Ave NE, Seattle, WA 98195, USA.
Email: emilyek@u.washington.edu

Abstract

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. Ecological and behavioral factors have been suggested to underlie stickleback reproductive isolation and incipient speciation, but reproductive proteins mediating gamete recognition during fertilization have so far remained unexplored. To begin to investigate the contribution of reproductive proteins to stickleback reproductive isolation, we have characterized the stickleback egg coat proteome. We find that stickleback egg coats are comprised of homologs to the zona pellucida (ZP) proteins ZP1 and ZP3, as in other teleost fish. Our molecular evolutionary analyses indicate that across teleosts, *ZP3* but not *ZP1* has experienced positive Darwinian selection. Mammalian *ZP3* is also rapidly evolving, and surprisingly some residues under selection in stickleback and mammalian ZP3 directly align. Despite broad homology, however, we find differences between mammalian and stickleback ZP proteins with respect to glycosylation, disulfide bonding, and sites of synthesis. Taken together, the changes we observe in 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.

KEYWORDS
egg coat, fertilization, reproduction, stickleback, zona pellucida

1 | INTRODUCTION

Animal oocytes are surrounded by a specialized glycoprotein extracellular matrix termed the "egg coat" (Killingbeck & Swanson, 2018; Shu et al., 2015; Wong & Wessel, 2006). The egg coat is an interface between the egg and its environment, protecting the oocyte from physical, chemical, and biological hazards (Murata et al., 2014; Shu et al., 2015; Wong & Wessel, 2006; Yamagami et al., 1992). 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 (Murata et al., 2014; Shu et al., 2015; Wong & Wessel, 2006). The egg coat goes by different names in the major vertebrate lineages, including the zona pellucida (ZP) in mammals, the vitelline envelope

in nonmammals, and the chorion in fish (Goudet et al., 2008; Shu et al., 2015). Despite historically complicated nomenclature, egg coats are generally comprised of a common set of glycoproteins characterized by the ZP module (Bork & Sander, 1992). 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 (Bork & Sander, 1992; Han et al., 2010; Monne & Jovine, 2011). Beyond the core ZP module, many ZP proteins have more elaborate structures, including trefoil domains, transmembrane domains (TMDs), 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 (Bokhove & Jovine, 2018; Callebaut et al., 2007; Jovine et al., 2002, 2004, 2006; Wilburn & Swanson, 2017). 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" (Bokhove et al., 2016; Wilburn & Swanson, 2017).

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 (Smith et al., 2005; Spargo & Hope, 2003). ZP3 proteins, which are typically the smallest ZP protein, contain only the ZP module; this minimal architecture as well as molecular phylogenetic observations suggest that ZP3 may be most similar to the ancestral ZP protein (Goudet et al., 2008; Litscher & Wassarman, 2014; Shu et al., 2015; Wassarman & Litscher, 2016). ZP3 proteins can also have repetitive proline/glutamine (P/Q) residues in relatively short stretches (Litscher & Wassarman, 2018). 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 (Callebaut et al., 2007; Goudet et al., 2008). These N-terminal ZP-N domains tend to be less conserved among orthologous proteins of different species (Callebaut et al., 2007). 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 (Callebaut et al., 2007; Litscher & Wassarman, 2018). ZP2 proteins are characterized by multiple N-terminal ZP-N domain repeats before their ZP module, and the ZP2 homolog ZPAX has an analogous N-terminal ZP-N domain repeat architecture (Callebaut et al., 2007).

ZP proteins are synthesized as precursor polypeptides with a signal sequence at the N-terminus and a C-terminal propeptide containing a TMD (Jovine et al., 2004, 2005; Monne et al., 2006). In some fish, however, the TMD is absent (Litscher & Wassarman, 2018; Monne et al., 2006). 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 (Jovine et al., 2005; Litscher & Wassarman, 2018; Monne et al., 2006). The dimerization of ZP-N domains between ZP modules facilitates the assembly of the filamentous egg coat ultrastructure (Bokhove & Jovine, 2018; Han et al., 2010; Monne et al., 2008; Okumura et al., 2007; Sasanami et al., 2006).

In mammalian egg coats, ZP proteins serve as both structural and sperm-binding proteins (Avella et al., 2013, 2014, 2016;

Baibakov et al., 2012; Bleil et al., 1988; Gahlay et al., 2010; Monne & Jovine, 2011). In fish, however, the role of ZP proteins in the egg coat is less well characterized and may be purely structural (Litscher & Wassarman, 2007, 2018; Monne et al., 2006). 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 (Amanze & Iyengar, 1990; Berois et al., 2011; Hart & Donovan, 1983; Hart, 1990; Wong & Wessel, 2006). The micropyle attracts sperm by chemotaxis, and its precise diameter restricts polyspermy by allowing passage of only one sperm at a time (Amanze & Iyengar, 1990; Hart, 1990; Lombardi, 1998; Yanagimachi et al., 2013). 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 (Hart, 1990).

In mammals, ZP proteins are synthesized in the ovary by oocytes and/or their surrounding follicle cells (Wassarman & Litscher, 2016). 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 (Arukwe & Goksoyr, 2003; Darie et al., 2005; Sano et al., 2013; Yamagami et al., 1992). 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 (Conner & Hughes, 2003; Litscher & Wassarman, 2018; Sano et al., 2017; Yamagami et al., 1992). ZP1 and ZP3, the most common ZP proteins in fish egg coats, both have paralogous classes of genes with hepatic and ovarian expression (Conner & Hughes, 2003; Litscher & Wassarman, 2018; Spargo & Hope, 2003). 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 (Arukwe & Goksoyr, 2003; Sano et al., 2013). One of the two expression sites typically becomes dominant, with liver synthesis of ZP proteins most common across teleosts (Arukwe & Goksoyr, 2003; Sano et al., 2013). Vitellogenin, an egg yolk precursor protein, shows similar hepatic expression and migration to the ovary in the bloodstream of fish, amphibians, and birds (Darie et al., 2005; Monne et al., 2006). In fish, ZP synthesis and vitellogenesis occur simultaneously in response to 17β-estradiol production by follicle cells (Arukwe & Goksoyr, 2003; Litscher & Wassarman, 2007; Monne et al., 2006; Yamagami et al., 1992).

Reproductive proteins that mediate gamete recognition during fertilization show species-specificity in both their structure and binding affinities (Swanson & Vacquier, 2002; Vieira & Miller, 2006). Despite their central role in fertilization, however, reproductive proteins are frequently among the most rapidly evolving genes in any taxa (Aagaard et al., 2006; Findlay & Swanson, 2010; Meslin et al., 2012; Palumbi, 2009; Swanson & Vacquier, 2002; Turner & Hoekstra, 2008a; Vacquier & Swanson, 2011). This juxtaposition of rapid evolution and functional constraint suggests a role for positive Darwinian selection in the coevolutionary maintenance of sperm-egg

interactions. Furthermore, the molecular evolutionary history of a protein can identify sites under adaptive evolution that may be functionally important (Findlay et al., 2008; Nielsen, 2005; Palmer et al., 2013; Swanson & Vacquier, 1997; Wilburn & Swanson, 2016; Wilburn et al., 2018; Wilburn, Bowen, Doty, et al., 2014). Signatures of rapid, adaptive evolution characteristic of reproductive proteins suggest that sequence diversification can be beneficial for genes involved in reproduction (Swanson & Vacquier, 2002). More formally, when nonsynonymous (d_N) substitutions outweigh synonymous (d_S) substitutions, d_N/d_S (also denoted ω) is greater than one; this suggests there was positive selection for changes in amino acid sequence (Goldman & Yang, 1994; Swanson & Vacquier, 2002; Wilburn & Swanson, 2016). 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 (Coyne & Orr, 2004; Palumbi, 2009; Shu et al., 2015; Turner & Hoekstra, 2008a; Wong & Wessel, 2006).

Threespine stickleback fish (*Gasterosteus aculeatus*) 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) (Bell & Foster, 1994; Peichel, 2005). Ancestrally marine fish have colonized thousands of freshwater lakes and streams, evolving significant diversity in morphology, behavior, physiology, and life history (Bell & Foster, 1994; Jones et al., 2012; McKinnon & Rundle, 2002; Peichel, 2005). These divergent forms come into contact with each other, but are frequently reproductively isolated, making stickleback an ideal model system for speciation research (Bell & Foster, 1994; McKinnon & Rundle, 2002). 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 (Bell & Foster, 1994; Hendry et al., 2009; McKinnon & Rundle, 2002). Despite nearly complete reproductive isolation in the wild, virtually any stickleback can be crossed in the lab to produce viable, fertile offspring (Peichel, 2005). 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 (McKinnon & Rundle, 2002). To begin to address this question from the perspective of female reproductive protein evolution, we combine proteomics and molecular evolutionary analyses to characterize the egg coat proteome of threespine stickleback fish.

2 | RESULTS

2.1 | Egg coat glycoprotein characterization

To characterize the proteome of threespine stickleback egg coats, egg coats were isolated and examined by sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE; Figure S3). Individual
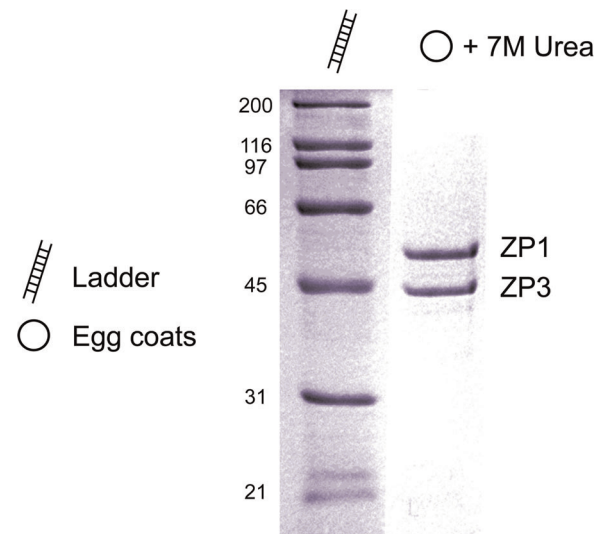

FIGURE 1 SDS-PAGE of stickleback egg coats. Stickleback egg coats were treated with 7 M urea to remove contaminating vitellogenin, likely carried over from egg coat isolation, and separated by SDS-PAGE. Individual bands were excised from the gel and analyzed by LC-MS/MS, with ZP1 and ZP3 identified as the two major protein components of stickleback egg coats. LC-MS/MS, liquid chromatography-tandem mass spectrometry; SDS-PAGE, sodium dodecyl sulphate- polyacrylamide gel electrophoresis

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 S2). 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 1). Similarly, shotgun proteomic analysis of purified egg coats determined that ZP1 and ZP3 comprise ~47% and ~53%, respectively, of the most abundant peptides by normalized spectral abundance factor (data not shown).

Reproductive proteins are frequently glycosylated (Gagneux & Varki, 1999; Varki, 2006). These post-translational modifications affect protein solubility and stability, and are thought to play a role in gamete recognition (Bausek et al., 2004; Wilburn & Swanson, 2016). Glycosylation analysis of stickleback egg coats indicates that of the two main egg coat proteins, only ZP3 is glycosylated (Figure 2). Stickleback ZP3 has a single putative *N*-glycosylation motif at N_{181}, and treatment with PNGase F confirmed that the glycan is *N*-linked. The mass shift in ZP3 after deglycosylation suggests that the glycan is ~4 kDa, representing ~20 sugar moieties (see Figure 2). Notably, this particular *N*-linked glycosylation site is highly conserved from fish to mammals (Darie et al., 2004).

2.2 | 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
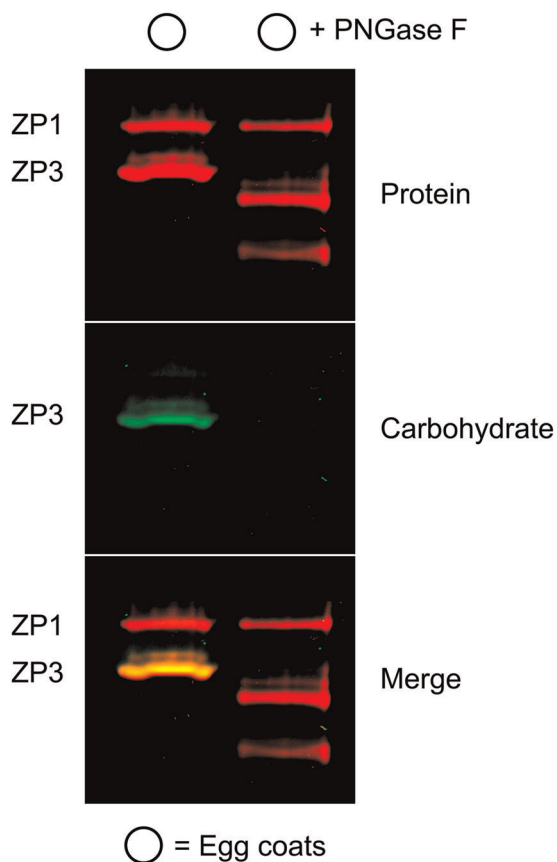
FIGURE 2 Glycosylation analysis of stickleback egg coats. Stickleback egg coats were treated with 7 M urea, deglycosylated with PNGase F, and separated by SDS-PAGE. The gel was stained with fluorescent carbohydrate and protein dyes, and the images were overlaid to visualize glycoprotein staining. Of 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. SDS-PAGE, sodium dodecyl sulphate-polyacrylamide gel electrophoresis

FIGURE 3 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 a crystal structure of chicken ZP3 (PDB ID: 3NK4; Han et al., 2010), 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 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. CAM, carbamidomethyl

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 before performing LC-MS/MS, with dynamic exclusion turned off to permit more quantitative peptide spectral counting. Reverse transcription polymerase chain reaction (RT-PCR) of *ZP1* and *ZP3* from stickleback ovary and liver RNA generated a database of *ZP* sequences for LC-MS/MS searches, resulting in 66% sequence coverage from 25 peptides for ZP1 and 69% sequence coverage from 24 peptides for ZP3. RT-PCR showed strong amplification of *ZP1* and *ZP3* from liver RNA only, suggesting that in stickleback *ZP* genes are transcribed in the liver (Figure S4).

ZP proteins have characteristic disulfide bonding patterns within the ZP-N and ZP-C domains of their ZP modules. A crystal structure
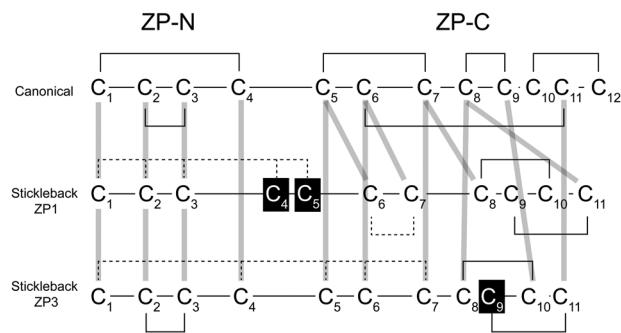
of chicken ZP3, for instance, shows a C_1–C_4, C_2–C_3 connectivity for ZP-N and a C_5–C_7, C_6–C_{11}, C_8–C_9, C_{10}–C_{12} connectivity for ZP-C (PDB ID: 3NK4; Han et al., 2010). Although our analysis generally found evidence of homologous disulfide bonding in stickleback ZP proteins, we also see evidence for shuffled disulfides and new cysteines that could alter the disulfide bonding of stickleback ZP proteins (summarized in Figure 3). For both ZP1 and ZP3, the majority of cysteines within the ZP module were modifiable with the alkylating agent iodoacetamide in the absence of reducing agent, as detected by carbamidomethyl (CAM) modification of these residues by mass spectrometry. Free, alkylatable cysteines not tied up in disulfide bonds suggests the potential for variable and/or transient disulfide bonding in stickleback ZP proteins. Consistent with this proposed 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 Figures 3 and S5). Table 1 provides counts for all potential disulfide-linked peptides by LC-MS/MS for ZP1, and Table 2 provides counts for ZP3.

2.3 | ZP molecular evolution

Molecular evolutionary analyses of stickleback *ZP1* and *ZP3* suggest that the divergence of *ZP3* across teleosts has been driven by positive

TABLE 1 Stickleback ZP1 disulfide bonding patterns by mass spectrometry

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

Note: 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; Han et al., 2010); a dash indicates that no homologous cysteine is present in chicken ZP3. Cysteines are considered "IAA reactive" if they were found to be CAM modified after 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.
Abbreviations: CAM, carbamidomethyl; IAA, iodoacetamide.

TABLE 2 Stickleback ZP3 disulfide bonding patterns by mass spectrometry

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

Note: 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; Han et al., 2010); a dash indicates that no homologous cysteine is present in chicken ZP3. Cysteines are considered "IAA reactive" if they were found to be CAM modified after 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.
Abbreviations: CAM, carbamidomethyl; IAA, iodoacetamide.

Darwinian selection, with 3.4% of sites in *ZP3* under positive selection with $\omega = 1.75$. To test for selection, a model of positive selection (M8) was compared to a model of neutral evolution (M8a) by likelihood ratio test (Swanson et al., 2003; Yang, 1997, 2007). These nested models allow for variation in ω among codons, but the null model M8a has a site class that restricts ω to 1 while the alternative model M8 has a site class with $\omega > 1$ to allow for adaptive evolution. 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 ($p = 2.4 \times 10^{-6}$; parameters summarized in Table 3). A similar test for adaptive evolution in *ZP1* across teleosts was not significant (see Table 3). These results were consistent to tree topology (using the species tree, a single *ZP1* or *ZP3* gene tree, or a concatenated gene tree). To our knowledge, this was the first investigation of *ZP* molecular evolution in fish (Jansa et al., 2003; Swanson et al., 2001; Turner & Hoekstra, 2008a). Residues under positive selection in stickleback ZP3 are modeled on the structure as red spheres in Figure 4.

The fact that we find no evidence of positive selection in teleost *ZP1* agrees with previous work in mammals, where *ZP3* has been found to be under selection while *ZP1* is not (Swanson et al., 2001). To compare sites under selection in *ZP3* across teleosts and mammals, we extended the analysis of (Swanson et al., 2001). In a representative group of eight mammalian species, 6.8% of sites in *ZP3*

were under positive selection with $\omega = 1.68$. 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 ($p = 0.0467$; parameters summarized in Table 4). Our analysis also found overlap in sites under selection between mammalian and teleost *ZP3*— in an alignment between mouse and stickleback ZP3, three rapidly evolving residues were the same ($p = 5.7 \times 10^{-5}$ based on χ^2 test; see Figure S5).

3 | DISCUSSION

The egg coat, as a major barrier encountered by sperm during fertilization, is an essential determinant of reproductive isolation in many taxa (Nixon et al., 2007; Palumbi, 2009; Vieira & Miller, 2006). Egg coat proteins are frequently rapidly evolving, and their divergence contributes to reproductive isolation and suggests a role in speciation (Coyne & Orr, 2004; Hart et al., 2014; Nosil, 2012;

TABLE 3 Evolutionary rate analysis of stickleback ZP proteins

Egg coat protein	M8a (neutral model)	M8 (positive selection)	−2ΔlogL	p value	Sites under selection
ZP1	$p_0 = 0.86205$	$p_0 = 0.87301$	0.728234	0.197	–
	$p = 0.73129$	$p = 0.70923$			
	$q = 2.80750$	$q = 2.57073$			
	$p_1 = 0.13795$	$p_1 = 0.12699$			
	$\omega = 1$	$\omega = 1.05861$			
ZP3	$p_0 = 0.93249$	$p_0 = 0.96646$	20.9473	**2.4 × 10⁻⁶**	71, 76, 86, 132, 136, 155, 159, 256, 259, 283, 329
	$p = 0.67172$	$p = 0.63379$			
	$q = 1.95187$	$q = 1.59312$			
	$p_1 = 0.06751$	$p_1 = 0.03354$			
	$\omega = 1$	**$\omega = 1.75256$**			

Note: 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 (Swanson et al., 2003; Yang, 1997, 2007). Sites under selection in ZP3 are specified with respect to stickleback, with the signal peptide included (see also Figure S5).
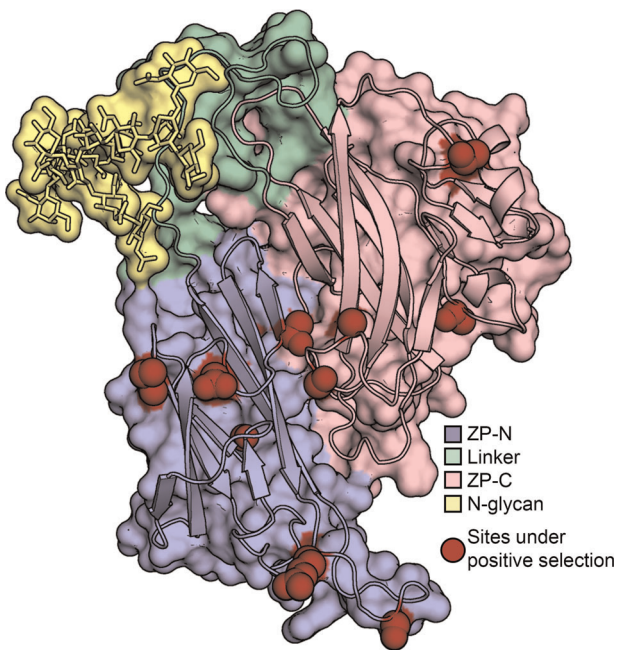

FIGURE 4 Stickleback ZP3 homology model. ZP-N and ZP-C domains are colored purple and pink, respectively; the linker between the domains is indicated in green; residues under positive selection across teleosts are denoted as red spheres; the single *N*-linked glycosylation site in stickleback ZP3 is shown in yellow. Note that sites under positive selection in ZP3 tend to cluster, particularly those within the ZP-N domain

Palumbi, 2009; Shu et al., 2015; Swanson & Vacquier, 2002; Turner & Hoekstra, 2008a; Vacquier & Swanson, 2011; Vieira & Miller, 2006; Wong & Wessel, 2006). With LC-MS/MS, we find that the stickleback egg coat is comprised of homologs to the ZP glycoproteins ZP1 and

TABLE 4 Evolutionary rate analysis of mammalian *ZP3*

M8a (neutral model)	M8 (positive selection)	−2ΔlogL	p value	Sites under selection
$p_0 = 0.80560$	$p_0 = 0.93247$	2.81315	**0.0467**	28, 33, 34, 39, 47, 50, 84, 185, 194, 341, 347, 372, 373
$p = 0.61874$	$p = 0.42651$			
$q = 3.50534$	$q = 1.35756$			
$p_1 = 0.19440$	$p_1 = 0.06753$			
$\omega = 1$	**$\omega = 1.68208$**			

Note: 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 *ZP3* across mammals. *p* values for a likelihood ratio test comparing M8 (selection) to M8a (nearly neutral) are shown, with significant results highlighted in bold (Swanson et al., 2003; Yang, 1997, 2007). Sites under selection in ZP3 are specified with respect to mouse, with the signal peptide included (see also Figure S5).

ZP3 (Figure 1). 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 (Conner & Hughes, 2003; Darie et al., 2004; Litscher & Wassarman, 2018).

Egg coat glycoproteins, as with other reproductive proteins, are frequently glycosylated (Gagneux & Varki, 1999; Varki, 2006). These carbohydrate modifications are thought to be involved in gamete recognition during fertilization, and to contribute to egg coat solubility (Bausek et al., 2004; Brivio et al., 1991; Claw & Swanson, 2012; Wilburn & Swanson, 2016). Of the two stickleback ZP proteins, we find that only ZP3 is glycosylated (Figure 2). 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 (see Figure 2). This relatively large mass is unusual for an *N*-linked glycan in vertebrates (representing

TABLE 5   Summary of changes to stickleback ZP protein architecture

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

Note: Summary of changes to stickleback ZP protein architecture relative to mammalian ZP proteins.

~20 sugar moieties), and may be indicative of a complex branched tetra-antennary carbohydrate that could present a recognition surface for sperm, as has been demonstrated in other animals (Brivio et al., 1991; Katsumata et al., 1996; North et al., 2009; Vo et al., 2003; Wilburn, Bowen, Feldhoff, et al., 2014; Wong & Wessel, 2006). There is only one potential N-linked glycosylation site in stickleback ZP3, [181]NVS[183] (Figure 4, modeled in yellow; see also Figure S5), which correspondingly is the same site found to be N-glycosylated in rainbow trout ZP3 (Darie et al., 2004, 2005). In an alignment between rainbow trout and mammalian ZP3 proteins, Darie et al. (2004) 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 (Darie et al., 2004, 2005; Litscher & Wassarman, 2007).

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 (Bork & Sander, 1992; Litscher & Wassarman, 2018; Sevier & Kaiser, 2002). Stickleback egg coats are remarkably insoluble relative to other characterized egg coats, a biochemical feature that seems true of fish egg coats in general (Arukwe & Goksoyr, 2003; Oppen-Berntsen et al., 1990; Yamagami et al., 1992). 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 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 3; Darie et al., 2004, 2008). 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 (Bokhove & Jovine, 2018; Bokhove et al., 2016). Our mass spectral analyses of ZP1 and ZP3 disulfide bonding revealed that most egg coat cysteines are variably bonded to one another and at least

partially exist in reduced, free sulfhydryl states (Figure 3; Tables 1 and 2). 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 3, 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 (see Figure 3; Greve & Wassarman, 1985; Jovine et al., 2006). There are an odd number of cysteine residues in both stickleback ZP1 and ZP3, suggesting that unpaired cysteines could participate in intra- or intermolecular disulfide shuffling. In particular, intermolecular disulfide bonds between adjacent ZP proteins in the egg coat could be important for stickleback egg coat ultrastructure and structural stability.

In teleost fish, *ZP* genes are known to exhibit both ovarian and hepatic expression (Arukwe & Goksoyr, 2003; Conner & Hughes, 2003; Sano et al., 2013; Spargo & Hope, 2003; Yamagami et al., 1992). To determine the site(s) of ZP synthesis in stickleback, primers were designed against *ZP1* and *ZP3* and amplified from both ovary and liver complementary DNA (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 S4). 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 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 noncovalently into long fibrils interconnected by cross-links of ZP1 (Nixon et al., 2007; Wassarman & Litscher, 2016, 2018). 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 (Avella et al., 2013; Bleil & Wassarman, 1980; Bokhove & Jovine, 2018; Dean, 2004; Epifano et al., 1995). Both ZP2 and ZP3 are required for egg coat formation, as ZP2 or ZP3 knockout mice fail to produce egg coats (Dean, 2004; Wassarman & Litscher, 2008). ZP1 knockout mice do form an egg coat, but it is loose and not interconnected and females are less fertile than wild-type (Rankin et al., 1999; Wassarman & Litscher, 2008). 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 (Avella et al., 2014). This agrees with the observation that the structural function of ZP2 in mammals is performed by ZP1-like subunits in fish, which lack ZP2 (Bokhove & Jovine, 2018; Conner & Hughes, 2003; Hughes, 2007; Litscher & Wassarman, 2018). 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 (Jovine et al., 2006; Monne et al., 2006; Wong & Wessel, 2006).

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 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 CFCS that allows cleavage of the C-terminal region; and a hydrophobic region or TMD (Jovine et al., 2005; Litscher & Wassarman, 2018; Monne et al., 2006). Changes to stickleback ZP protein architecture are highlighted in approximate order from N- to C-terminus (see Table 5 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 (Callebaut et al., 2007). Stickleback ZP1 has also lost its fourth canonical cysteine in the ZP-N domain of its ZP module (see Figure 3). 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 3, boxed in black; Darie et al., 2004, 2008; Litscher & Wassarman, 2018). Stickleback ZP3 also has an additional cysteine residue, C_9, in its ZP-C domain (Figure 3, boxed in black; see also Figure S5). Both stickleback ZP proteins have lost their TMDs, likely as a consequence of their hepatic expression (Litscher & Wassarman, 2007, 2018; Wang & Gong, 1999). 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 4, modeled in yellow; see also Figure S5; Darie et al., 2004, 2005; Litscher & Wassarman, 2007). Loss of glycosylation in ZP1 may be common among teleosts, as this has also been reported in rainbow trout (Darie et al., 2004, 2005). Finally, disulfide shuffling is potentially prevalent in both stickleback ZP proteins, particularly within the ZP-N domains of their ZP modules, and particularly for ZP1 (see Figure 3, dashed lines). 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. It remains to be determined which, if any, of these features are specific to stickleback or are present in all teleosts.

The role of the ZP-N domain in protein polymerization is not limited to reproductive proteins, and is conserved throughout eukaryotes (Bokhove & Jovine, 2018; Bokhove et al., 2016; Jovine et al., 2002, 2006; Swanson et al., 2011). 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 (Bokhove & Jovine, 2018; Conner & Hughes, 2003; Hughes, 2007; Litscher & Wassarman, 2018). 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 (Darie et al., 2005).

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 (Swanson & Vacquier, 2002, 2002). Numerous evolutionary forces have been attributed to the rapid evolution of reproductive proteins, including sperm competition, sexual conflict, cryptic female choice, reinforcement, and pathogen resistance (Clark et al., 2009; Kosman & Levitan, 2014; Swanson & Vacquier, 2002, 2002; Turner & Hoekstra, 2006, 2008a, 2008b). 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 3). Rapid evolution in ZP3 has also been found in mammals, including in our analysis (see Table 4; Jansa et al., 2003; Swanson et al., 2001; Turner & Hoekstra, 2006, 2008a, 2008b). It is notable that the N-terminal portion of ZP3 appears to be under selection in both teleosts and mammals—amazingly, in an alignment between mouse and stickleback ZP3, the location of three positively selected residues in the N-terminus are the same ($p = 5.7 \times 10^{-5}$ based on χ^2 test; see Figure S5). Stickleback ZP3 has eleven rapidly evolving residues, seven that fall within its ZP-N domain and four that fall within its ZP-C domain (Figure 4, denoted with red spheres; see also Table 3 and Figure S5). If these residues are involved in sperm-egg interactions, it is significant to note that as few as 10 amino acid changes in a sea urchin sperm-egg recognition protein bindin were found to lead to gametic incompatibility (Zigler et al., 2005). Although it is interesting that stickleback 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 disulfide shuffling, two extra cysteines that may be involved in homo- or heterodimeric ZP assembly, loss of its N-terminal ZP-N domain, and lack 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 (Kinloch et al., 1995). 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 (Avella et al., 2013; Boja et al., 2003; Dean, 2004; Florman & Wassarman, 1985). Regardless, amino acids in and around this C-terminal "sperm-combining site" have been identified as under positive Darwinian selection in a diverse set

of mammals (Kinloch et al., 1995; Swanson et al., 2001; Turner & Hoekstra, 2006, 2008a). Although the homologous "sperm-combining site" region in teleosts contains no rapidly evolving residues, the three residues maintained under selection from teleosts to mammals are located in the N-terminal portion of ZP3 at the beginning of the ZP-N domain, in a region of the protein that likely forms a structured β strand (see Figure S5). The fact that mammalian and teleost ZP3 both have rapidly evolving residues in this stable region of shared structure could suggest a shared role in sperm interaction. 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 4). 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. Similarly, there may be specific post-translational modifications of the ZP proteins that surround the micropyle, as evidenced by specific lectin staining of the micropyle for some teleost species (Yanagimachi et al., 2013). 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 (Hart, 1990; Litscher & Wassarman, 2007; Wilburn & Swanson, 2016; Wong & Wessel, 2006). Selective pressures specific to the micropyle may underlie the signature of adaptive evolution we find in ZP3, but our current study does not have the resolution to address micropyle-specific differences. It is noteworthy, however, that micropyles from divergent species show differential staining with Coomassie or lectins (Yanagimachi et al., 2013). It is conceivable that the presence of the micropyle 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. Furthermore, the variable disulfide bonding pattern of stickleback ZP proteins likely results in a heterogeneous egg coat surface that would be difficult for a limited set of binding proteins and/or proteases to recognize and dissolve. This structural variation could also lead to variable presentation of rapidly evolving sites, effectively magnifying the effects of positive selection. A similar phenomenon has been observed in the male pheromones of lungless salamanders, where altered disulfide bonding increases conformational sampling of rapidly evolving loops and may provide a selective advantage in recognizing a broader range of female olfactory receptors (Wilburn, Bowen, Doty, et al. (2014). 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 exposure to pathogens that may also evolve rapidly — a protective structural barrier might be evolutionarily favored (Shu et al., 2015; Wong & Wessel, 2006). 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 (Darie et al., 2004, 2005; Litscher & Wassarman, 2018; Oppen-Berntsen et al., 1990; Yamagami et al., 1992). 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 (Darie et al., 2005). 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 (Arukwe & Goksoyr, 2003; Litscher & Wassarman, 2007; Litscher & Wassarman, 2018; Monne et al., 2006; Wong & Wessel, 2006; Yamagami et al., 1992).

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 (Hart, 1990). The recently described zebrafish egg plasma membrane protein Bouncer represents a possible candidate for species-specific sperm recognition at the egg plasma membrane, as zebrafish eggs expressing the medaka version of Bouncer permit fertilization with medaka sperm despite the 200 million years of evolution separating the two species (Herberg et al., 2018). We hypothesize that ZP3 in the egg coat may also contribute to sperm recognition at the micropyle, given its proposed role as a sperm receptor in mammals and its maintenance under positive Darwinian selection, including at the level of shared residues, from teleost fish to mammals.

4 | MATERIALS AND METHODS

4.1 | 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). To minimize the number of breeding females that were trapped and euthanized, egg coat samples from three females were used for the analyses described in this manuscript.

4.2 | 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 (Hanks & Wallace, 1949)). Insoluble egg coats were isolated by centrifugation (2000 x g for 10 min), 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 (Figures 1 and 2).

4.3 | 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 min, followed by 100 V for 90 min (Schagger & von Jagow, 1987). 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) or SYPRO Ruby (Thermo Fisher Scientific). Glycosylation was detected by in-gel periodic acid-Schiff staining using the Pro-Q Emerald 488 glycoprotein staining kit (Invitrogen) and imaged using a Typhoon FLA 9000 laser bed scanner (GE Healthcare Bio-Sciences). To determine if the observed glycosylation of stickleback ZP3 was N-linked, egg coats were treated with PNGase F (New England BioLabs), an enzyme that removes only N-linked glycans, before electrophoresis following the manufacturer's protocol.

4.4 | 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 min incubation), addition of acetonitrile (with 15 min 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 dithiothreitol in 100 mM ammonium

bicarbonate at 56°C for 45 min, followed by alkylation with 55 mM iodoacetamide in 100 mM ammonium bicarbonate in the dark for 30 min. The gel pieces were washed twice with 100 mM ammonium bicarbonate, dehydrated with acetonitrile, and incubated with 1 μg trypsin (Promega) 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) loaded with 4 μm Jupiter C12 Proteo reverse-phase resin (Phenomenex) 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-min LC gradient consisted of 2% to 40% B in 30 min, 40% to 60% B in 10 min, 60% to 95% B in 5 min, followed by a 15 min wash and a 15 min column equilibration. Peptides were eluted from the column and electrosprayed into a Velos Pro Linear Ion Trap Mass Spectrometer (Thermo Fisher Scientific). Data were acquired using data-dependent acquisition (DDA) and analyzed using an in-house version of COMET (Eng et al., 2013, 2015) (with a differential modification of 15.994915 Da for methionine and a static modification of 57.021461 Da for cysteine) for database searching against 277,509 publicly available stickleback ESTs (retrieved from the UCSC Genome Browser) that were assembled with Trinity (Grabherr et al., 2011; Haas et al., 2013) to produce 36,667 putative genes and six-frame translated. The final search database of 220,002 sequences also contained known contaminants such as trypsin and human keratin. Percolator v.2.09 (Kall et al., 2007) was used to filter the peptide-spectrum matches using a decoy database of reversed sequences with a q value threshold of ≤0.01, and peptides were assembled into protein identifications using an in-house implementation of IDPicker with the same q value threshold (Zhang et al., 2007).

4.5 | 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 ultra-centrifugation (procedure modified from MacDonald et al., 1987). 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 min at 5000 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 h 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).

ZP1 and ZP3 coding sequences were PCR amplified from G. aculeatus liver cDNA (primer sequences in Table S1), 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). Sequences were analyzed using the Lasergene DNASTAR package (v.11.1.0).

4.6 | 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 before 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 2-mercaptoethanol in 7 M urea in 100 mM ammonium bicarbonate, and the samples were incubated at 60°C for 45 min. Samples were then alkylated with 200 mM iodoacetamide in 7 M urea in 100 mM ammonium bicarbonate and incubated for 45 min in the dark. A final reduction was performed with 400 mM 2-mercaptoethanol, 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) 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-min LC gradient consisted of 0% to 16% B in 15 min, 16% to 35% B in 60 min, 35% to 75% B in 15 min, 75% to 100% B in 5 min, followed by a 5 min wash and a 25 min column equilibration. Peptides were eluted from the column on a 50°C heated source (CorSolutions) and electrosprayed into an Orbitrap Fusion Lumos Mass Spectrometer (Thermo Fisher Scientific). Data were acquired using DDA with dynamic exclusion turned off. Mass spectral data were 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 (Xu & Freitas, 2007, 2008, 2009; Xu, Yang, et al., 2008; Xu, Zhang, et al., 2008; Xu et al., 2010).

4.7 | Molecular evolution of teleost and mammalian ZP proteins

To assess ZP gene evolution across teleost fish, 30 species were chosen spanning the teleost phylogeny, with ZP1 and ZP3 open reading frames identified by homology to stickleback ZP1 and ZP3 using TBLASTX (Altschul et al., 1990; Betancur-R et al., 2017). Genes identified had E-values on the order of 1e-70, and were at least 10 orders of magnitude better in E-value than the second-best hit. For the 31 total species (including G. aculeatus), sequences for each gene were aligned using FSA, which favors introducing gaps in its

alignments in cases of uncertainty (specifically when the probability that a character is aligned is equal to the probability that it is gapped) (Bradley et al., 2009). A phylogenetic tree was generated from NCBI taxonomy using phlyoT v2 (https://phylot.biobyte.de/; see Figure S1) to represent the currently accepted species phylogeny for the teleost species chosen.

To compare sites under selection between teleost and mammalian ZP3, which has been previously shown to be rapidly evolving, we extended the analysis of (Swanson et al., 2001). Mammalian ZP3 sequences (GenBank accession numbers: M20026 (Mus musculus), Y10823 (Rattus rattus), X56777 (Homo sapiens), S71825 (marmosets), X82639 (Macaca radiata), D45070 (Canis familiaris), D45068 (Felis catus), and D45065 (Sus scrofa)) were aligned using FSA (Bradley et al., 2009) and a phylogenetic tree was generated from NCBI taxonomy using phlyoT v2 (https://phylot.biobyte.de/; see Figure S2).

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 (Swanson et al., 2003; Yang, 1997, 2007). Ambiguous and/or gapped alignment positions were excluded from analysis. Sites under positive selection were defined as coding positions with a Bayes empirical Bayes posterior probability of >50% under M8 (Yang et al., 2005). Results using the accepted species phylogeny were consistent with results using gene trees from each respective locus (ZP1 or ZP3) or a tree from a concatenated ZP1/ZP3 sequence (data not shown).

A homology model of stickleback ZP3 was generated using Rosetta by threading of the stickleback ZP3 sequence to a chicken ZP3 structure (PDB ID: 3NK4; Han et al., 2010) (aligned using Clustal Omega; Sievers et al., 2011), loop modeling using cyclic coordinate descent with refinement by kinetic closure (KIC), and full atom minimization using the relax function (Mandell et al., 2009; Qian et al., 2007; Sievers et al., 2011). N-glycosylation was modeled using GlycanBuilder (Ceroni et al., 2007).

CONFLICT OF INTERESTS

The authors declare that there are no conflict of interests.

DATA AVAILABILITY STATEMENT

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the data set identifiers PXD017488 (disulfide

characterization data) and PXD017489 (egg coat mass spectral characterization data).

REFERENCES

Aagaard, J. E., Yi, X., MacCoss, M. J., & Swanson, W. J. (2006). Rapidly evolving zona pellucida domain proteins are a major component of the vitelline envelope of abalone eggs. *Proceedings of the National Academy of Sciences of the United States of America*, 103(46), 17302–17307. https://doi.org/10.1073/pnas.0603125103

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, 215(3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Amanze, D., & Iyengar, A. (1990). The micropyle: A sperm guidance system in teleost fertilization. *Development*, 109(2), 495–500.

Arukwe, A., & Goksoyr, A. (2003). Eggshell and egg yolk proteins in fish: hepatic proteins for the next generation: oogenetic, population, and evolutionary implications of endocrine disruption. *Comparative Hepatology*, 2(1), 4. https://doi.org/10.1186/1476-5926-2-4

Avella, M. A., Baibakov, B., & Dean, J. (2014). A single domain of the ZP2 zona pellucida protein mediates gamete recognition in mice and humans. *The Journal of Cell Biology*, 205(6), 801–809. https://doi.org/10.1083/jcb.201404025

Avella, M. A., Baibakov, B. A., Jimenez-Movilla, M., Sadusky, A. B., & Dean, J. (2016). ZP2 peptide beads select human sperm in vitro, decoy mouse sperm in vivo, and provide reversible contraception. *Science Translational Medicine*, 8(336), 336ra60. https://doi.org/10.1126/scitranslmed.aad9946

Avella, M. A., Xiong, B., & Dean, J. (2013). The molecular basis of gamete recognition in mice and humans. *Molecular Human Reproduction*, 19(5), 279–289. https://doi.org/10.1093/molehr/gat004

Baibakov, B., Boggs, N. A., Yauger, B., Baibakov, G., & Dean, J. (2012). Human sperm bind to the N-terminal domain of ZP2 in humanized zonae pellucidae in transgenic mice. *The Journal of Cell Biology*, 197(7), 897–905. https://doi.org/10.1083/jcb.201203062

Bausek, N., Ruckenbauer, H. H., Pfeifer, S., Schneider, W. J., & Wohlrab, F. (2004). Interaction of sperm with purified native chicken ZP1 and ZPC proteins. *Biology of Reproduction*, 71(2), 684–690. https://doi.org/10.1095/biolreprod.104.028605

Bell, M. A., & Foster, S. A. (1994). *The evolutionary biology of the threespine stickleback*. Oxford University Press.

Berois, N., Arezo, M. J., & Papa, N. G. (2011). Gamete interactions in teleost fish: The egg envelope. Basic studies and perspectives as environmental biomonitor. *Biological Research*, 44(2), 119–124. https://doi.org/10.4067/S0716-97602011000200002

Betancur-R, R., Wiley, E. O., Arratia, G., Acero, A., Bailly, N., Miya, M., Lecointre, G., & Ortí, G. (2017). Phylogenetic classification of bony fishes. *BMC Evolutionary Biology*, 17(1), 162. https://doi.org/10.1186/s12862-017-0958-3

Bleil, J. D., Greve, J. M., & Wassarman, P. M. (1988). Identification of a secondary sperm receptor in the mouse egg zona pellucida: Role in maintenance of binding of acrosome-reacted sperm to eggs. *Developmental Biology*, 128(2), 376–385. https://doi.org/10.1016/0012-1606(88)90299-0

Bleil, J. D., & Wassarman, P. M. (1980). Structure and function of the zona pellucida: Identification and characterization of the proteins of the mouse oocyte's zona pellucida. *Developmental Biology*, 76(1), 185–202. https://doi.org/10.1016/0012-1606(80)90371-1

Boja, E. S., Hoodbhoy, T., Fales, H. M., & Dean, J. (2003). Structural characterization of native mouse zona pellucida proteins using mass spectrometry. *The Journal of Biological Chemistry*, 278(36), 34189–34202. https://doi.org/10.1074/jbc.M304026200

Bokhove, M., & Jovine, L. (2018). Current topics in developmental biology. In E. S. Litscher, & P. M. Wassarman (Eds.), *Chapter thirteen— Structure of zona pellucida module proteins*. Academic Press. https://doi.org/10.1016/bs.ctdb.2018.02.007

Bokhove, M., Nishimura, K., Brunati, M., Han, L., de Sanctis, D., Rampoldi, L., & Jovine, L. (2016). A structured interdomain linker directs self-polymerization of human uromodulin. *Proceedings of the National Academy of Sciences of the United States of America*, 113(6), 1552–1557. https://doi.org/10.1073/pnas.1519803113

Bork, P., & Sander, C. (1992). A large domain common to sperm receptors (Zp2 and Zp3) and TGF-beta type III receptor. *FEBS Letters*, 300(3), 237–240. https://doi.org/10.1016/0014-5793(92)80853-9

Bradley, R. K., Roberts, A., Smoot, M., Juvekar, S., Do, J., Dewey, C., Holmes, I., & Pachter, L. (2009). Fast statistical alignment. *PLOS Computational Biology*, 5(5), e1000392. https://doi.org/10.1371/journal.pcbi.1000392

Brivio, M. F., Bassi, R., & Cotelli, F. (1991). Identification and characterization of the major components of the Oncorhynchus mykiss egg chorion. *Molecular Reproduction and Development*, 28(1), 85–93. https://doi.org/10.1002/mrd.1080280114

Callebaut, I., Mornon, J. P., & Monget, P. (2007). Isolated ZP-N domains constitute the N-terminal extensions of zona pellucida proteins. *Bioinformatics*, 23(15), 1871–1874. https://doi.org/10.1093/bioinformatics/btm265

Ceroni, A., Dell, A., & Haslam, S. M. (2007). The GlycanBuilder: A fast, intuitive and flexible software tool for building and displaying glycan structures. *Source Code for Biology and Medicine*, 2, 3. https://doi.org/10.1186/1751-0473-2-3

Clark, N. L., Gasper, J., Sekino, M., Springer, S. A., Aquadro, C. F., & Swanson, W. J. (2009). Coevolution of interacting fertilization proteins. *PLOS Genetics*, 5(7), e1000570. https://doi.org/10.1371/journal.pgen.1000570

Claw, K. G., & Swanson, W. J. (2012). Evolution of the egg: New findings and challenges. *Annual Review of Genomics and Human Genetics*, 13, 109–125. https://doi.org/10.1146/annurev-genom-090711-163745

Conner, S. J., & Hughes, D. C. (2003). Analysis of fish ZP1/ZPB homologous genes--evidence for both genome duplication and species-specific amplification models of evolution. *Reproduction*, 126(3), 347–352. https://doi.org/10.1530/rep.0.1260347

Coyne, J. A., & Orr, H. A. (2004). *Speciation*. Sinaur Associates, Inc.

Darie, C. C., Biniossek, M. L., Gawinowicz, M. A., Milgrom, Y., Thumfart, J. O., Jovine, L., Litscher, E. S., & Wassarman, P. M. (2005). Mass spectrometric evidence that proteolytic processing of rainbow trout egg vitelline envelope proteins takes place on the egg. *The Journal of Biological Chemistry*, 280(45), 37585–37598. https://doi.org/10.1074/jbc.M506709200

Darie, C. C., Biniossek, M. L., Jovine, L., Litscher, E. S., & Wassarman, P. M. (2004). Structural characterization of fish egg vitelline envelope proteins by mass spectrometry. *Biochemistry*, 43(23), 7459–7478. https://doi.org/10.1021/bi0495937

Darie, C. C., Litscher, E. S., & Wassarman, P. M. (2008). Structure, processing, and polymerization of rainbow trout egg vitelline envelope proteins. In C. Popescu, A. D. Zamfir, & N. Dinca (Eds.), *Applications of mass spectrometry in life safety* (pp. 23–36). Springer Netherlands.

Dean, J. (2004). Reassessing the molecular biology of sperm-egg recognition with mouse genetics. *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology*, 26(1), 29–38. https://doi.org/10.1002/bies.10412

Eng, J. K., Hoopmann, M. R., Jahan, T. A., Egertson, J. D., Noble, W. S., & MacCoss, M. J. (2015). A Deeper look into comet—Implementation and features. *Journal of the American Society for Mass Spectrometry*, 26(11), 1865–1874. https://doi.org/10.1007/s13361-015-1179-x

Eng, J. K., Jahan, T. A., & Hoopmann, M. R. (2013). Comet: An open-source MS/MS sequence database search tool. *Proteomics*, 13(1), 22–24. https://doi.org/10.1002/pmic.201200439

Epifano, O., Liang, L., & Dean, J. (1995). Mouse Zp1 encodes a zona pellucida protein homologous to egg envelope proteins in mammals and fish. *Journal of Biological Chemistry*, *270*(45), 27254–27258. https://doi.org/10.1074/jbc.270.45.27254

Findlay, G. D., & Swanson, W. J. (2010). Proteomics enhances evolutionary and functional analysis of reproductive proteins. *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology*, *32*(1), 26–36. https://doi.org/10.1002/bies.200900127

Findlay, G. D., Yi, X., Maccoss, M. J., & Swanson, W. J. (2008). Proteomics reveals novel Drosophila seminal fluid proteins transferred at mating. *PLOS Biology*, *6*(7), e178. https://doi.org/10.1371/journal.pbio.0060178

Florman, H. M., & Wassarman, P. M. (1985). O-linked oligosaccharides of mouse egg ZP3 account for its sperm receptor activity. *Cell*, *41*(1), 313–324. https://doi.org/10.1016/0092-8674(85)90084-4

Gagneux, P., & Varki, A. (1999). Evolutionary considerations in relating oligosaccharide diversity to biological function. *Glycobiology*, *9*(8), 747–755. https://doi.org/10.1093/glycob/9.8.747

Gahlay, G., Gauthier, L., Baibakov, B., Epifano, O., & Dean, J. (2010). Gamete recognition in mice depends on the cleavage status of an egg's zona pellucida protein. *Science*, *329*(5988), 216–219. https://doi.org/10.1126/science.1188178

Goldman, N., & Yang, Z. (1994). A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Molecular Biology and Evolution*, *11*(5), 725–736. https://doi.org/10.1093/oxfordjournals.molbev.a040153

Goudet, G., Mugnier, S., Callebaut, I., & Monget, P. (2008). Phylogenetic analysis and identification of pseudogenes reveal a progressive loss of zona pellucida genes during evolution of vertebrates. *Biology of Reproduction*, *78*(5), 796–806. https://doi.org/10.1095/biolreprod.107.064568

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B. W., Nusbaum, C., Lindblad-Toh, K., … Regev, A. (2011). Trinity: Reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nature Biotechnology*, *29*(7), 644–652. https://doi.org/10.1038/nbt.1883

Greve, J. M., & Wassarman, P. M. (1985). Mouse egg extracellular coat is a matrix of interconnected filaments possessing a structural repeat. *Journal of Molecular Biology*, *181*(2), 253–264. https://doi.org/10.1016/0022-2836(85)90089-0

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., Couger, M. B., Eccles, D., Li, B., Lieber, M., MacManes, M. D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C. N., … Regev, A. (2013). De novo transcript sequence reconstruction from RNA-Seq: Reference generation and analysis with Trinity. *Nature Protocols*, *8*(8), 1494–1512. https://doi.org/10.1038/nprot.2013.084

Han, L., Monné, M., Okumura, H., Schwend, T., Cherry, A. L., Flot, D., Matsuda, T., & Jovine, L. (2010). Insights into egg coat assembly and egg-sperm interaction from the X-ray structure of full-length ZP3. *Cell*, *143*(3), 404–415. https://doi.org/10.1016/j.cell.2010.09.041

Hanks, J. H., & Wallace, R. E. (1949). Relation of oxygen and temperature in the preservation of tissues by refrigeration. *Proceedings of the Society for Experimental Biology and Medicine*, *71*(2), 196–200. https://doi.org/10.3181/00379727-71-17131

Hart, M. W., Sunday, J. M., Popovic, I., Learning, K. J., & Konrad, C. M. (2014). Incipient speciation of sea star populations by adaptive gamete recognition coevolution. *Evolution*, *68*(5), 1294–1305. https://doi.org/10.1111/evo.12352

Hart, N. H. (1990). Fertilization in teleost fishes: Mechanisms of sperm-egg interactions. *International Review of Cytology*, *121*, 1–66. https://doi.org/10.1016/s0074-7696(08)60658-0

Hart, N. H., & Donovan, M. (1983). Fine structure of the chorion and site of sperm entry in the egg of Brachydanio. *Journal of Experimental Zoology*, *227*(2), 277–296. https://doi.org/10.1002/jez.1402270212

Hendry, A. P., Bolnick, D. I., Berner, D., & Peichel, C. L. (2009). Along the speciation continuum in sticklebacks. *Journal of Fish Biology*, *75*(8), 2000–2036. https://doi.org/10.1111/j.1095-8649.2009.02419.x

Herberg, S., Gert, K. R., Schleiffer, A., & Pauli, A. (2018). The Ly6/uPAR protein Bouncer is necessary and sufficient for species-specific fertilization. *Science*, *361*(6406), 1029–1033. https://doi.org/10.1126/science.aat7113

Hughes, D. C. (2007). ZP genes in avian species illustrate the dynamic evolution of the vertebrate egg envelope. *Cytogenetic and Genome Research*, *117*(1-4), 86–91. https://doi.org/10.1159/000103168

Jansa, S. A., Lundrigan, B. L., & Tucker, P. K. (2003). Tests for positive selection on immune and reproductive genes in closely related species of the murine genus Mus. *Journal of Molecular Evolution*, *56*(3), 294–307. https://doi.org/10.1007/s00239-002-2401-6

Jones, F. C., Grabherr, M. G., Chan, Y. F., Russell, P., Mauceli, E., Johnson, J., Swofford, R., Pirun, M., Zody, M. C., White, S., Birney, E., Searle, S., Schmutz, J., Grimwood, J., Dickson, M. C., Myers, R. M., Miller, C. T., Summers, B. R., Knecht, A. K., … Kingsley, D. M. (2012). The genomic basis of adaptive evolution in threespine sticklebacks. *Nature*, *484*(7392), 55–61. https://doi.org/10.1038/nature10944

Jovine, L., Darie, C. C., Litscher, E. S., & Wassarman, P. M. (2005). Zona pellucida domain proteins. *Annual Review of Biochemistry*, *74*, 83–114. https://doi.org/10.1146/annurev.biochem.74.082803.133039

Jovine, L., Janssen, W. G., Litscher, E. S., & Wassarman, P. M. (2006). The PLAC1-homology region of the ZP domain is sufficient for protein polymerisation. *BMC Biochemistry*, *7*(11), 11. https://doi.org/10.1186/1471-2091-7-11

Jovine, L., Qi, H., Williams, Z., Litscher, E., & Wassarman, P. M. (2002). The ZP domain is a conserved module for polymerization of extracellular proteins. *Nature Cell Biology*, *4*(6), 457–461. https://doi.org/10.1038/ncb802

Jovine, L., Qi, H., Williams, Z., Litscher, E. S., & Wassarman, P. M. (2004). A duplicated motif controls assembly of zona pellucida domain proteins. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(16), 5922–5927. https://doi.org/10.1073/pnas.0401600101

Kall, L., Canterbury, J. D., Weston, J., Noble, W. S., & MacCoss, M. J. (2007). Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nature Methods*, *4*(11), 923–925. https://doi.org/10.1038/nmeth1113

Katsumata, T., Noguchi, S., Yonezawa, N., Tanokura, M., & Nakano, M. (1996). Structural characterization of the N-linked carbohydrate chains of the zona pellucida glycoproteins from bovine ovarian and fertilized eggs. *European Journal of Biochemistry*, *240*(2), 448–453. https://doi.org/10.1111/j.1432-1033.1996.0448h.x

Killingbeck, E. E., & Swanson, W. J. (2018). Egg coat proteins across metazoan evolution. *Current Topics in Developmental Biology*, *130*, 443–488. https://doi.org/10.1016/bs.ctdb.2018.03.005

Kinloch, R. A., Sakai, Y., & Wassarman, P. M. (1995). Mapping the mouse ZP3 combining site for sperm by exon swapping and site-directed mutagenesis. *Proceedings of the National Academy of Sciences of the United States of America*, *92*(1), 263–267. https://doi.org/10.1073/pnas.92.1.263

Kosman, E. T., & Levitan, D. R. (2014). Sperm competition and the evolution of gametic compatibility in externally fertilizing taxa. *MHR: Basic Science of Reproductive Medicine*, *20*(12), 1190–1197. https://doi.org/10.1093/molehr/gau069

Litscher, E. S., & Wassarman, P. M. (2007). Egg extracellular coat proteins: from fish to mammals. *Histology and Histopathology*, *22*(3), 337–347. https://doi.org/10.14670/HH-22.337

Litscher, E. S., & Wassarman, P. M. (2014). Evolution, structure, and synthesis of vertebrate egg-coat proteins. *Trends in Developmental Biology*, 8, 65–76.

Litscher, E. S., & Wassarman, P. M. (2018). The fish egg's zona pellucida. *Current Topics in Developmental Biology*, 130, 275–305. https://doi.org/10.1016/bs.ctdb.2018.01.002

Lombardi, J. (1998). *Comparative vertebrate reproduction*. Kluwer Academic Publishers.

MacDonald, R. J., Swift, G. H., Przybyla, A. E., & Chirgwin, J. M. (1987). Isolation of RNA using guanidinium salts. *Methods in Enzymology*, 152, 219–227. https://doi.org/10.1016/0076-6879(87)52023-7

Mandell, D. J., Coutsias, E. A., & Kortemme, T. (2009). Sub-angstrom accuracy in protein loop reconstruction by robotics-inspired conformational sampling. *Nature Methods*, 6(8), 551–552. https://doi.org/10.1038/nmeth0809-551

McKinnon, J. S., & Rundle, H. D. (2002). Speciation in nature: the threespine stickleback model systems. *Trends in Ecology & Evolution*, 17(10), 480–488. https://doi.org/10.1016/S0169-5347(02)02579-X

Meslin, C., Mugnier, S., Callebaut, I., Laurin, M., Pascal, G., Poupon, A., Goudet, G., & Monget, P. (2012). Evolution of genes involved in gamete interaction: Evidence for positive selection, duplications and losses in vertebrates. *PLOS One*, 7(9), e44548. https://doi.org/10.1371/journal.pone.0044548

Monne, M., Han, L., & Jovine, L. (2006). Tracking down the ZP domain: From the mammalian zona pellucida to the molluscan vitelline envelope. *Seminars in Reproductive Medicine*, 24(4), 204–216. https://doi.org/10.1055/s-2006-948550

Monne, M., Han, L., Schwend, T., Burendahl, S., & Jovine, L. (2008). Crystal structure of the ZP-N domain of ZP3 reveals the core fold of animal egg coats. *Nature*, 456(7222), 653–657. https://doi.org/10.1038/nature07599

Monne, M., & Jovine, L. (2011). A structural view of egg coat architecture and function in fertilization. *Biology of Reproduction*, 85(4), 661–669. https://doi.org/10.1095/biolreprod.111.092098

Murata, K., Conte, F. S., McInnis, E., Fong, T. H., & Cherr, G. N. (2014). Identification of the origin and localization of chorion (egg envelope) proteins in an ancient fish, the white sturgeon, Acipenser transmontanus. *Biology of Reproduction*, 90(6), 132. https://doi.org/10.1095/biolreprod.113.116194

Nielsen, R. (2005). Molecular signatures of natural selection. *Annual Review of Genetics*, 39, 197–218. https://doi.org/10.1146/annurev.genet.39.073003.112420

Nixon, B., Aitken, R. J., & McLaughlin, E. A. (2007). New insights into the molecular mechanisms of sperm-egg interaction. *Cellular and Molecular Life Sciences*, 64(14), 1805–1823. https://doi.org/10.1007/s00018-007-6552-x

North, S. J., Hitchen, P. G., Haslam, S. M., & Dell, A. (2009). Mass spectrometry in the analysis of N-linked and O-linked glycans. *Current Opinion in Structural Biology*, 19(5), 498–506. https://doi.org/10.1016/j.sbi.2009.05.005

Nosil, P. (2012). *Ecological speciation*. Oxford University Press.

Okumura, H., Aoki, N., Sato, C., Nadano, D., & Matsuda, T. (2007). 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. *Biology of Reproduction*, 76(1), 9–18. https://doi.org/10.1095/biolreprod.106.056267

Oppen-Berntsen, D. O., Helvik, J. V., & Walther, B. T. (1990). The major structural proteins of cod (*Gadus morhua*) eggshells and protein crosslinking during teleost egg hardening. *Developmental Biology*, 137(2), 258–265. https://doi.org/10.1016/0012-1606(90)90252-E

Palmer, M. R., McDowall, M. H., Stewart, L., Ouaddi, A., MacCoss, M. J., & Swanson, W. J. (2013). Mass spectrometry and next-generation sequencing reveal an abundant and rapidly evolving abalone sperm protein. *Molecular Reproduction and Development*, 80(6), 460–465. https://doi.org/10.1002/mrd.22182

Palumbi, S. R. (2009). Speciation and the evolution of gamete recognition genes: Pattern and process. *Heredity*, 102(1), 66–76. https://doi.org/10.1038/hdy.2008.104

Paoletti, A. C., Parmely, T. J., Tomomori-Sato, C., Sato, S., Zhu, D., Conaway, R. C., Conaway, J. W., Florens, L., & Washburn, M. P. (2006). Quantitative proteomic analysis of distinct mammalian Mediator complexes using normalized spectral abundance factors. *Proceedings of the National Academy of Sciences of the United States of America*, 103(50), 18928–18933. https://doi.org/10.1073/pnas.0606379103

Peichel, C. L. (2005). Fishing for the secrets of vertebrate evolution in threespine sticklebacks. *Developmental Dynamics: An Official Publication of the American Association of Anatomists*, 234(4), 815–823. https://doi.org/10.1002/dvdy.20564

Perez-Riverol, Y., Csordas, A., Bai, J., Bernal-Llinares, M., Hewapathirana, S., Kundu, D. J., Inuganti, A., Griss, J., Mayer, G., Eisenacher, M., Pérez, E., Uszkoreit, J., Pfeuffer, J., Sachsenberg, T., Yilmaz, S., Tiwary, S., Cox, J., Audain, E., Walzer, M., ... Vizcaíno, J. A. (2019). The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Research*, 47(D1), D442–D450. https://doi.org/10.1093/nar/gky1106

Qian, B., Raman, S., Das, R., Bradley, P., McCoy, A. J., Read, R. J., & Baker, D. (2007). High-resolution structure prediction and the crystallographic phase problem. *Nature*, 450(7167), 259–264. https://doi.org/10.1038/nature06249

Rankin, T., Talbot, P., Lee, E., & Dean, J. (1999). Abnormal zonae pellucidae in mice lacking ZP1 result in early embryonic loss. *Development*, 126(17), 3847–3855.

Sano, K., Kawaguchi, M., Katano, K., Tomita, K., Inokuchi, M., Nagasawa, T., Hiroi, J., Kaneko, T., Kitagawa, T., Fujimoto, T., Arai, K., Tanaka, M., & Yasumasu, S. (2017). Comparison of egg envelope thickness in teleosts and its relationship to the sites of ZP protein synthesis. *Journal of Experimental Zoology Part B, Molecular and Developmental Evolution*, 328(3), 240–258. https://doi.org/10.1002/jez.b.22729

Sano, K., Kawaguchi, M., Watanabe, S., Nagakura, Y., Hiraki, T., & Yasumasu, S. (2013). Inferring the evolution of teleostean zp genes based on their sites of expression. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, 320(5), 332–343. https://doi.org/10.1002/jez.b.22507

Sasanami, T., Ohtsuki, M., Ishiguro, T., Matsushima, K., Hiyama, G., Kansaku, N., Doi, Y., & Mori, M. (2006). Zona pellucida domain of ZPB1 controls specific binding of ZPB1 and ZPC in Japanese quail (*Coturnix japonica*). *Cells, Tissues, Organs*, 183(1), 41–52. https://doi.org/10.1159/000094905

Schagger, H., & von Jagow, G. (1987). Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. *Analytical Biochemistry*, 166(2), 368–379. https://doi.org/10.1016/0003-2697(87)90587-2

Sevier, C. S., & Kaiser, C. A. (2002). Formation and transfer of disulphide bonds in living cells. *Nature Reviews Molecular Cell Biology*, 3(11), 836–847. https://doi.org/10.1038/nrm954

Shu, L., Suter, M. J.-F., & Räsänen, K. (2015). Evolution of egg coats: linking molecular biology and ecology. *Molecular Ecology*, 24(16), 4052–4073. https://doi.org/10.1111/mec.13283

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D., & Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, 7, 539. https://doi.org/10.1038/msb.2011.75

Smith, J., Paton, I. R., Hughes, D. C., & Burt, D. W. (2005). Isolation and mapping the chicken zona pellucida genes: An insight into the evolution of orthologous genes in different species. *Molecular Reproduction and Development*, 70(2), 133–145. https://doi.org/10.1002/mrd.20197

Spargo, S. C., & Hope, R. M. (2003). Evolution and nomenclature of the zona pellucida gene family. *Biology of Reproduction*, 68(2), 358–362. https://doi.org/10.1095/biolreprod.102.008086

Swanson, W., & Vacquier, V. (2002). Reproductive protein evolution. *Annual Review of Ecology and Systematics*, 33(1), 161–179. https://doi.org/10.1146/annurev.ecolsys.33.010802.150439

Swanson, W. J., Aagaard, J. E., Vacquier, V. D., Monne, M., Sadat Al Hosseini, H., & Jovine, L. (2011). The molecular basis of sex: Linking yeast to human. *Molecular Biology and Evolution*, 28(7), 1963–1966. https://doi.org/10.1093/molbev/msr026

Swanson, W. J., Nielsen, R., & Yang, Q. (2003). Pervasive adaptive evolution in mammalian fertilization proteins. *Molecular Biology and Evolution*, 20(1), 18–20.

Swanson, W. J., & Vacquier, V. D. (1997). The abalone egg vitelline envelope receptor for sperm lysin is a giant multivalent molecule. *Proceedings of the National Academy of Sciences of the United States of America*, 94(13), 6724–6729. https://doi.org/10.1073/pnas.94.13.6724

Swanson, W. J., & Vacquier, V. D. (2002). The rapid evolution of reproductive proteins. *Nature Reviews Genetics*, 3(2), 137–144. https://doi.org/10.1038/nrg733

Swanson, W. J., Yang, Z., Wolfner, M. F., & Aquadro, C. F. (2001). Positive Darwinian selection drives the evolution of several female reproductive proteins in mammals. *Proceedings of the National Academy of Sciences of the United States of America*, 98(5), 2509–2514. https://doi.org/10.1073/pnas.051605998

Turner, L. M., & Hoekstra, H. E. (2006). Adaptive evolution of fertilization proteins within a genus: variation in ZP2 and ZP3 in deer mice (*Peromyscus*). *Molecular Biology and Evolution*, 23(9), 1656–1669. https://doi.org/10.1093/molbev/msl035

Turner, L. M., & Hoekstra, H. E. (2008a). Causes and consequences of the evolution of reproductive proteins. *The International Journal of Developmental Biology*, 52(5-6), 769–780. https://doi.org/10.1387/ijdb.082577lt

Turner, L. M., & Hoekstra, H. E. (2008b). Reproductive protein evolution within and between species: maintenance of divergent ZP3 alleles in *Peromyscus*. *Molecular Ecology*, 17(11), 2616–2628. https://doi.org/10.1111/j.1365-294X.2008.03780.x

Vacquier, V. D., & Swanson, W. J. (2011). Selection in the rapid evolution of gamete recognition proteins in marine invertebrates. *Cold Spring Harbor Perspectives in Biology*, 3(11), a002931. https://doi.org/10.1101/cshperspect.a002931

Varki, A. (2006). Nothing in glycobiology makes sense, except in the light of evolution. *Cell*, 126(5), 841–845. https://doi.org/10.1016/j.cell.2006.08.022

Vieira, A., & Miller, D. J. (2006). Gamete interaction: Is it species-specific? *Molecular Reproduction and Development*, 73(11), 1422–1429. https://doi.org/10.1002/mrd.20542

Vo, L. H., Yen, T., Macher, B. A., & Hedrick, J. L. (2003). Identification of the ZPC oligosaccharide ligand involved in sperm binding and the glycan structures of *Xenopus laevis* vitelline envelope glycoproteins. *Biology of Reproduction*, 69(6), 1822–1830. https://doi.org/10.1095/biolreprod.103.015370

Wang, H., & Gong, Z. (1999). Characterization of two zebrafish cDNA clones encoding egg envelope proteins ZP2 and ZP3. *Biochimica et Biophysica Acta*, 1446(1-2), 156–160. https://doi.org/10.1016/s0167-4781(99)00066-4

Wassarman, P. M., & Litscher, E. S. (2008). Mammalian fertilization: The egg's multifunctional zona pellucida. *The International Journal of Developmental Biology*, 52(5-6), 665–676. https://doi.org/10.1387/ijdb.072524pw

Wassarman, P. M., & Litscher, E. S. (2016). Chapter thirty-one—A bespoke coat for eggs: Getting ready for fertilization. *Current Topics in Developmental Biology*, 117, 539–552. https://doi.org/10.1016/bs.ctdb.2015.10.018

Wassarman, P. M., & Litscher, E. S. (2018). Chapter ten—The mouse egg's zona pellucida. *Current Topics in Developmental Biology*, 130, 331–356. https://doi.org/10.1016/bs.ctdb.2018.01.003

Wilburn, D. B., Bowen, K. E., Doty, K. A., Arumugam, S., Lane, A. N., Feldhoff, P. W., & Feldhoff, R. C. (2014). 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*, 9(5), e96975. https://doi.org/10.1371/journal.pone.0096975

Wilburn, D. B., Bowen, K. E., Feldhoff, P. W., & Feldhoff, R. C. (2014). Proteomic analyses of courtship pheromones in the redback salamander, *Plethodon cinereus*. *Journal of Chemical Ecology*, 40(8), 928–939. https://doi.org/10.1007/s10886-014-0489-y

Wilburn, D. B., & Swanson, W. J. (2016). From molecules to mating: Rapid evolution and biochemical studies of reproductive proteins. *Journal of Proteomics*, 135, 12–25. https://doi.org/10.1016/j.jprot.2015.06.007

Wilburn, D. B., & Swanson, W. J. (2017). The "ZP domain" is not one, but likely two independent domains. *Molecular Reproduction and Development*, 84(4), 284–285. https://doi.org/10.1002/mrd.22781

Wilburn, D. B., Tuttle, L. M., Klevit, R. E., & Swanson, W. J. (2018). 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*, 115(6), 1310–1315. https://doi.org/10.1073/pnas.1709061115

Wong, J. L., & Wessel, G. M. (2006). Defending the zygote: Search for the ancestral animal block to polyspermy. *Current Topics in Developmental Biology*, 72, 1–151. https://doi.org/10.1016/S0070-2153(05)72001-9

Xu, H., & Freitas, M. A. (2007). A mass accuracy sensitive probability based scoring algorithm for database searching of tandem mass spectrometry data. *BMC Bioinformatics*, 8, 133. https://doi.org/10.1186/1471-2105-8-133

Xu, H., Yang, L., & Freitas, M. A. (2008). 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*, 9, 347. https://doi.org/10.1186/1471-2105-9-347

Xu, H., & Freitas, M. A. (2008). Monte carlo simulation-based algorithms for analysis of shotgun proteomic data. *Journal of Proteome Research*, 7(7), 2605–2615. https://doi.org/10.1021/pr800002u

Xu, H., & Freitas, M. A. (2009). MassMatrix: A database search program for rapid characterization of proteins and peptides from tandem mass spectrometry data. *Proteomics*, 9(6), 1548–1555. https://doi.org/10.1002/pmic.200700322

Xu, H., Hsu, P. H., Zhang, L., Tsai, M. D., & Freitas, M. A. (2010). Database search algorithm for identification of intact cross-links in proteins and peptides using tandem mass spectrometry. *Journal of Proteome Research*, 9(7), 3384–3393. https://doi.org/10.1021/pr100369y

Xu, H., Zhang, L., & Freitas, M. A. (2008). Identification and characterization of disulfide bonds in proteins and peptides from tandem MS data by use of the MassMatrix MS/MS search engine. *Journal of Proteome Research*, 7(1), 138–144. https://doi.org/10.1021/pr070363z

Yamagami, K., Hamazaki, T. S., Yasumasu, S., Masuda, K., & Iuchi, I. (1992). Molecular and cellular basis of formation, hardening, and breakdown of the egg envelope in fish. *International Review of Cytology*, 136, 51–92. https://doi.org/10.1016/s0074-7696(08)62050-1

Yanagimachi, R., Cherr, G., Matsubara, T., Andoh, T., Harumi, T., Vines, C., Pillai, M., Griffin, F., Matsubara, H., Weatherby, T., & Kaneshiro, K.

(2013). Sperm attractant in the micropyle region of fish and insect eggs. *Biology of Reproduction*, *88*(2), 1–11. https://doi.org/10.1095/biolreprod.112.105072

Yang, Z. (1997). PAML: A program package for phylogenetic analysis by maximum likelihood. *Bioinformatics*, *13*(5), 555–556. https://doi.org/10.1093/bioinformatics/13.5.555

Yang, Z. (2007). PAML 4: Phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, *24*(8), 1586–1591. https://doi.org/10.1093/molbev/msm088

Yang, Z., Wong, W. S., & Nielsen, R. (2005). Bayes empirical Bayes inference of amino acid sites under positive selection. *Molecular Biology and Evolution*, *22*(4), 1107–1118. https://doi.org/10.1093/molbev/msi097

Zhang, B., Chambers, M. C., & Tabb, D. L. (2007). Proteomic parsimony through bipartite graph analysis improves accuracy and transparency. *Journal of Proteome Research*, *6*(9), 3549–3557. https://doi.org/10.1021/pr070230d

Zigler, K. S., McCartney, M. A., Levitan, D. R., & Lessios, H. A. (2005). Sea urchin bindin divergence predicts gamete compatibility. *Evolution*, *59*(11), 2399–2404. https://doi.org/10.1111/j.0014-3820.2005.tb00949.x

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.