Evolution of reproductive proteins from animals and plants

Nathaniel L Clark, Jan E Aagaard and Willie J Swanson

Department of Genome Sciences, University of Washington, Box 357730, Seattle, Washington 8195-7730, USA

Correspondence should be addressed to W J Swanson; Email: wswanson@gs.washington.edu

All authors contributed equally to this review

Abstract

Sexual reproduction is a fundamental biological process common among eukaryotes. Because of the significance of reproductive proteins to fitness, the diversity and rapid divergence of proteins acting at many stages of reproduction is surprising and suggests a role of adaptive diversification in reproductive protein evolution. Here we review the evolution of reproductive proteins acting at different stages of reproduction among animals and plants, emphasizing common patterns. Although we are just beginning to understand these patterns, by making comparisons among stages of reproduction for diverse organisms we can begin to understand the selective forces driving reproductive protein diversity and the functional consequences of reproductive protein evolution.

Introduction

Traits that influence reproductive success and contribute to reproductive isolation of animal and plant species have been a central focus of evolutionary biology since Darwin (Darwin 1859). Recently, specific genes have been identified which act at several stages of reproduction to regulate the interaction among male and female gametes (Vacquier 1998). Comparative sequencing studies among taxonomic groups have led to the discovery that reproductive proteins evolve more rapidly than other genes. For example, variation among proteins from reproductive tissues of Drosophila is 2-fold greater than among proteins of non-reproductive tissues (Civetta & Singh 1995). An enormous number of alleles for gamete recognition proteins can be found within naturally occurring plant populations (Wang et al. 2001). Evidence for the rapid divergence of reproductive proteins is also evident among several taxonomic groups. Genome-wide comparisons have shown that reproductive proteins are among the most highly diverged mammalian genes (Makalowski & Boguski 1998) and evolve more rapidly than proteins expressed in other tissues (Torgerson et al. 2002). Sperm proteins from free-spawning marine gastropods are among the most rapidly evolving proteins known, accruing amino acid substitutions at rates several fold higher than the most rapidly evolving mammalian proteins (Metz et al. 1998).

Because reproductive proteins regulate essential processes that fundamentally influence fitness, such high levels of diversity and divergence are remarkable and suggest reproductive proteins may frequently evolve as the result of adaptive evolution. A variety of statistical tests show a signature of adaptive evolution among reproductive proteins (Swanson & Vacquier 2002). This emerging pattern of the adaptive significance of reproductive proteins leads to two intriguing questions. First: what are the selective forces driving reproductive protein evolution? Do selective pressures result from exogenous sources (e.g. microbial attack) or endogenous sources mediated by gametes during fertilization (e.g. conflicts between male and female fitness)? Studies are beginning to suggest co-evolution between interacting pairs of male and female proteins during reproduction may be a major force driving the adaptive evolution of reproductive proteins (Swanson & Vacquier 2002). Secondly: what are the functional consequences of reproductive protein diversification? Experimental studies show functional domains of reproductive proteins evolving under adaptive evolution are sufficient to result in reproductive isolation among closely related species (Lyon & Vacquier 1999, Sainudiin et al. 2005). This has direct implications for both mechanisms and rates of speciation (Coyne & Orr 2004). Addressing questions such as these not only enriches the field of evolutionary biology, but also has significant practical applications in plant and animal breeding and in human health.

In this review, we provide a brief discussion of statistical approaches for identifying proteins under positive selection. We then outline characteristic stages of reproduction for animals and plants from post-copulation (or deposition of pollen on the stigma for plants) through fertilization. Our emphasis is on reviewing the literature that provides
evidence of positive selection for specific proteins acting at these reproductive stages.

**Statistical tests for positive selection**

There are two general classes of statistical tests used to assess whether genes evolve under positive selection (Table 1). The first class includes methods that rely on polymorphism data within species, whereas the second class focuses primarily on nucleotide divergence between species. The strongest support for positive selection results from tests based on both polymorphism and divergence data. Although a thorough discussion is beyond the scope of this review (see additional reviews by Kreitman (2000) and Yang & Bielawski (2000)), we provide a brief description of several tests including their strengths and limitations as a background for the discussion of positive selection on reproductive proteins.

Tests of positive selection employing polymorphism data rely on the expected behavior of selection to alter allele frequencies relative to neutrally evolving loci. When positive selection acts at a locus, variation at linked sites is reduced (a selective sweep). Following the selective sweep, mutation generates polymorphism at linked sites, but alleles initially occur at low frequency. Thus a frequency spectrum showing an excess of low-frequency alleles relative to the expectation for neutral loci is indicative of positive selection. A frequency spectrum skewed towards an excess of intermediate frequency alleles relative to the neutral expectation is indicative of selection acting to favor the maintenance of multiple genotypes at a locus and is referred to as balancing selection. Several test statistics including Tajima’s D (Tajima 1989), Fu and Li’s D (Fu & Li 1993) or Fay and Wu’s H (Fay et al. 2001) measure the effects of selection on the frequency spectrum employing different estimators. A major limitation of these tests is that demographic factors such as genetic bottlenecks and population subdivision can strongly influence the test statistic, and tests may be differentially affected (see Fu 1997). In order to control for demographic effects, approaches such as the Hudson–Kreitman–Aguadé (HKA) test (Hudson et al. 1987) employ statistics comparing polymorphism at two or more unlinked loci. The HKA test is also an example of an approach that utilizes an outgroup (closely related or sister species) to measure the effect of selection. Because levels of polymorphism within and between species should be correlated for neutrally evolving loci, significant deviation from this expectation measured using the HKA test is evidence of adaptive evolution.

The second class of statistical tests focuses primarily on divergence among species, including several methods employing the codon-substitution models developed by Goldman & Yang (1994) designed to test for adaptive divergence in the amino acid sequence of proteins. These tests are intuitively appealing in their use of the non-synonymous (amino acid changing, $d_N$) to synonymous ($d_S$) nucleotide substitution ratio to define the type of selection. Because $d_S$ provides an approximation of the neutral rate of substitution, $d_N/d_S = 1$ indicates amino acid changing substitutions are neutral. In contrast, $d_N/d_S > 1$ indicates a selective advantage to amino acid substitutions in a protein consistent with adaptive divergence among species. Several different implementations of these models allow for variation in $d_N/d_S$ among branches of a phylogeny (Yang 1998), among amino acid sites within a protein (Nielsen & Yang 1998, Yang et al. 2000b), or among sites on particular lineages (Yang & Nielsen 2002). Tests for adaptive divergence among species using these models have become popular in part because of their power to detect selection, which is expected to frequently act only along discrete lineages or at specific sites within a protein (Yang & Nielsen 2002) and because of their potential utility in predicting which sites are the targets of positive selection (Anisimova et al. 2002).

**Stages of reproduction: animals**

Here, we present the post-mating stages of animal reproduction as appropriate for discussion of observations of positive selection among vertebrates, echinoderms, mollusks and insects. In internally fertilizing species, male and female gametes enter and traverse the female reproductive tract from opposite extremes. Along this migration,
the gametes mature and are protected by male- and female-contributed factors. When insemination and ovulation times do not coincide, sperm may be stored in the female tract for long periods until ovulation. Finally, when sperm and egg meet, sperm must traverse several egg barriers for fertilization to occur. These egg vestments are formed by diverse structural components, and sperm have evolved biochemical strategies to cross them.

The oviductal environment around the time of ovulation stimulates the maturation of both spermatozoa and eggs in preparation for fertilization and development. The steps and factors regulating maturation of gametes are largely unknown; furthermore, the steps are likely to be very different between divergent taxonomic groups. Maturation is crucial for fertilization; for example, mammalian sperm are incapable of fertilization upon insemination and must pass through steps of capacitation to gain this capability (Yanagimachi 1994). It has been argued that capacitation is under the synergistic influence of adjoining portions of the female reproductive tract culminating in an appropriate amount of capacitated sperm meeting the newly ovulated egg (Hunter & Rodriguez-Martinez 2004). If events are so synergistically controlled, then a sophisticated set of interactions between gametes and the oviductal microenvironment awaits discovery. In mammals, three known effectors of sperm motility and capacitation are beta-amino acids, bicarbonate ions and progesterone (Boatman 1997). Maturation of post-ovulatory eggs is also required for fertilization. One factor in the oviduct known to affect mammalian eggs is oviductin (OGP), which binds the egg and sperm surfaces and facilitates gamete recognition (Boatman & Magnoni 1995).

Efficient navigation of the female tract through chemotaxis would bring large benefits to sperm charged with the daunting task of finding an egg. Among animals, human sperm have been shown to chemotax to follicular fluid (Kalt et al. 1991), and sperm from the abalone (Haliotis rufescens), a free-spawning marine invertebrate, chemotax to an egg-released amino acid, L-tryptophan (Riffell et al. 2004). If events are so synergistically controlled, then a sophisticated set of interactions between gametes and the oviductal microenvironment awaits discovery. In mammals, three known effectors of sperm motility and capacitation are beta-amino acids, bicarbonate ions and progesterone (Boatman 1997). Maturation of post-ovulatory eggs is also required for fertilization. One factor in the oviduct known to affect mammalian eggs is oviductin (OGP), which binds the egg and sperm surfaces and facilitates gamete recognition (Boatman & Magnoni 1995).

Efficient navigation of the female tract through chemotaxis would bring large benefits to sperm charged with the daunting task of finding an egg. Among animals, human sperm have been shown to chemotax to follicular fluid (Kalt et al. 1991), and sperm from the abalone (Haliotis rufescens), a free-spawning marine invertebrate, chemotax to an egg-released amino acid, L-tryptophan (Riffell et al. 2002). This attractant works species-specifically as demonstrated in experiments with sperm from closely related abalone species (Riffell et al. 2004). It is not known if chemotaxis guides sperm to other important areas of the female reproductive tract, such as sperm storage sites.

Mating often occurs before ovulation, making sperm storage necessary. Sperm storage has been reported among insects, mollusks, annelids, mammals, birds, reptiles and sharks (Neubaum & Wollner 1999a, Ferraguti et al. 2002). A great variety of storage systems exist, and specialized sperm storage organs have evolved in several taxonomic groups, including several independent instances in reptiles, birds and insects (Burke et al. 1972, Neubaum & Wollner 1999b, Sever & Hamlett 2002). These organs are often modified outpocketings or tubules, frequently called spermathecae. Drosophilid flies have two separate sperm storage sites, both a seminal vesicle for storage of ejaculate and spermathecae. Sperm retention times vary widely between species and can steadily supply gametes over days or even over years (Neubaum & Wollner 1999b). In mammals, copulation is usually timed to be a peri-ovulatory event, and sperm is stored for a relatively short period of a few days in a specialized region of the Fallopian tube (Hunter & Rodriguez-Martinez 2004). Yet mechanisms to prolong sperm storage do exist in mammals; for example, female bats are able to store sperm for months (Racey 1979). Little is known about which male or female factors are responsible for proper channeling, storage and protection of sperm during these periods. Progress in identifying these factors has been made in Drosophila species, in which seminal proteins are seen to affect sperm storage (e.g. Acp36DE, Acp29AB) (Wollner 2002).

Sperm encounter several threats in the female reproductive tract, including foreign pathogens and the female immune system (Austin 1975). Drosophila transfers anti-bacterial proteins in seminal plasma (Lung et al. 2001), and several proteins found in human semen (Utleg et al. 2003, Fung et al. 2004) show anti-bacterial activity (PIP, CAMP, lactotransferrin, transferrin, MSMB). Mammalian semen contains prostaglandin E, which locally depresses immune response, perhaps to protect sperm from female immune attack (Kelly & Critchley 1997).

Once sperm and egg meet, the sperm must pass several barriers for fertilization to occur. An outermost, gelatinous layer often surrounds the egg and is passed in several species by sperm hypermotility. Beneath the gelatinous layer a substantial egg coat forms a formidable barrier. In several taxa this egg coat is composed of cross-linked glycoprotein, as in the mammalian zona pellucida (Wasserman et al. 2001) and in the abalone vitelline envelope (Swanson & Vacquier 1997). The acrosomal vesicle at the fore of the sperm head contains proteins which, upon release, open a hole in the egg coat by either non-enzymatic dissolution or proteolysis (Lewis et al. 1982, Wasserman et al. 2001). Triggering the release of acrosomal contents is often mediated by species-specific factors present on the egg coat. These egg factors are biochemically diverse with different organisms employing combinations of several molecular classes including polysaccharides, peptides, glycoproteins and saponins (Vacquier 1998).

Once through the egg coat, the last barrier to fertilization is the egg membrane with which the sperm binds and then fuses. Proteins involved in binding are found on the sperm surface or on an extended sperm appendage, the acrosomal process (Swanson & Vacquier 1995b). Recent progress has been made in understanding mammalian membrane fusion, as both a sperm and an egg protein have been shown necessary for gamete fusion (Kaji et al. 2000, Le Naour et al. 2000, Miyado et al. 2000, Inoue et al. 2005). Once all barriers are passed and fusion of the sperm and egg cell membranes is achieved, the nuclei may finally come together to begin animal development.
Rapidly evolving proteins: animals

In essentially all steps in animal fertilization where the proteins have been identified there is evidence for rapid divergence of the genes encoding the reproductive proteins (Vacquier 1998, Swanson & Vacquier 2002). The steps and genes in Fig. 1 provide an overview of how ubiquitous the presence of rapid evolution is, not only in diversity of presence amongst animal groups but also in the variety of stages during reproduction.

During mating, the male transfers seminal fluid along with sperm during copulation. Timing of seminal fluid transfer varies between animal groups, with some species transferring seminal fluid before sperm and others having simultaneous transfer of sperm and seminal fluid. The functions of seminal fluid have been studied in detail in Drosophila melanogaster (Wolfner 2002), where approximately 83 components have been identified (Swanson et al. 2001b, Mueller et al. 2005). The function of Drosophila seminal fluid proteins ranges from sperm storage, induction of ovulation, reducing the female’s remating rate, and formation of a mating plug (Wolfner 2002, Chapman & Davies 2004). Although at the primary sequence level there is little similarity between Drosophila and mammalian seminal fluid, detailed three-dimensional modeling shows striking similarities (Mueller et al. 2004). The lack of similarity at the primary sequence level may be due to the signal being obscured by rapid evolution. One interesting example of positive selection of a seminal fluid protein is a gene involved in the formation of the semen coagulum. This is particularly interesting in primates, where some primates such as chimpanzees form a solid coagulum and other primates such as humans do not form a firm coagulum. The SEMG2 gene product is a major structural element of the semen coagulum and shows positive selection as indicated by a high $d_{SN}/d_{SE}$ ratio. Interestingly, the rate of evolution of SEMG2 correlates with levels of female promiscuity and firmness of the semen coagulum, indicating the selective pressure may relate to sperm competition (Dorus et al. 2004). Other constituents of the semen coagulum, such as the SEMG1 gene, also show rapid adaptive evolution (Kingan et al. 2003). Additionally, a scan of 161 human seminal proteins found statistically significant signatures of positive selection for seven genes within primates, and greater than 10% had a signature suggestive of positive selection (Clark & Swanson 2005). Seminal fluid genes not exclusively involved in the semen coagulum also show rapid adaptive evolution. In Drosophila, these include the Acp26Aa (Tsaur & Wu 1997) gene that induces ovulation and the Acp36DE gene involved in sperm storage (Begun et al. 2000). In genome-wide scans, approximately 10% of the seminal fluid genes show the signature of possible positive selection (Swanson et al. 2001b); however, further statistical tests are needed to confirm this figure.

Once the sperm enter the female reproductive tract, they need to efficiently pass through the reproductive tract and enter sperm storage prior to fertilization. Sperm motility is an obviously important factor for fertilization success and passage through the female reproductive tract. One might predict ion-channels regulating sperm motility would be conserved; yet surprisingly even these genes show rapid divergence. The CatSper1 gene is required for depolarization-evoked calcium entry and hyperactivated
flagellar movement (Carlson et al. 2003). When this ion channel was compared amongst primates or rodents, there was extreme diversity found in the length of the N-terminus of the protein (Podlaha & Zhang 2003, Podlaha et al. 2005). The functional outcome of the length variation remains unknown, but comparisons with neutral rates of insertion/deletion events suggest that the length diversity was promoted by positive selection. It has been suggested that the diversity of N-terminal length of CatSper1 could be involved in the rate of channel inactivation. Components of the female reproductive tract are less well studied. One interesting example is the OGP gene, found in the estrous oviductal fluid. The role of OGP remains unclear, but it appears to play a role in species-specific fertilization and has been demonstrated to be subjected to positive selection (Swanson et al. 2003).

Once the sperm reaches the site of fertilization, there are multiple examples of both sperm and egg proteins that show extraordinary rates of evolution between closely related species. One of the first steps of sperm–egg interaction is binding of the sperm to the egg coat and induction of the sperm acrosome reaction. In sea urchins, there is significant diversity in the structure of the egg jelly coat glycoproteins (Vilela-Silva et al. 2002), which induces the acrosome reaction. Functional assays show that this diversity results in species-specific acrosome reaction induction. Since there are no methods to study the evolution of carbohydrates, the evolutionary forces generating the diversity remain unknown. In mammals, the acrosome reaction is induced by the zona pellucida, an elevated glycoproteinaceous coat (Wasserman et al. 2004). Previously, ZP3 was characterized as the primary receptor for sperm that induces the acrosome reaction. However, recent evidence from multiple knockout mice suggests that the supermolecular structure of the ZP glycoproteins may induce the acrosome reaction (Rankin et al. 2003). At least two of the major components of the zona pellucida show rapid divergence between species. In fact, the ZP2 and ZP3 glycoproteins are among the 10% most different proteins between rodents and humans. By analyses of site-specific \( d_S/d_I \) ratio, it has been shown that ZP3 sites implicated in the species-specific induction of the acrosome reaction have been the target of positive selection (Swanson et al. 2001a).

Once the sperm has undergone the acrosome reaction, it must create a hole in the egg coat through which it will pass. The dissolution of the egg envelope has been intensively studied in abalone, a free-spawning marine invertebrate. Abalone lysin has been shown to be species-specifically and non-enzymatically dissolve a hole in the egg vitelline envelope (Lewis et al. 1982). Lysin shows rapid evolution, with exons evolving up to 15 times faster than introns (Metz et al. 1998). By analyses of site-specific \( d_S/d_I \) ratios, particular amino acids have been shown to be the target of positive selection (Yang et al. 2000a). When chimeric lysins of these residues undergoing positive selection are made between species by site-directed mutagenesis, the expected switch in specificity is obtained, indicating a link between the positive selection and functional changes (Lyon & Vacquier 1999). In mammals, it is not clear how the sperm penetrates the egg envelope. The protease acrosin (ACR) was thought to be involved, although knockouts of ACR are still fertile, indicating a redundant function (Baba et al. 1994). Interestingly, ACR does show statistically significant signatures of positive selection consistent with some beneficial function for reproduction (Swanson et al. 2003). After the sperm passes the egg coat, the final step of fertilization is fusion between the two gametes (Vacquier 1998). Even the molecules involved in the fusion step show extreme diversity. In mammals, the egg receptor regulating fusion appears to be CD9. When CD9 is knocked out, sperm are unable to undergo fusion (Miyado et al. 2000). Positive selection is observed in CD9, and when sites subjected to positive selection are mapped onto the topology the majority fall on the extracellular loops (Swanson et al. 2003). In marine invertebrates, the sea urchin sperm protein bindin has been implicated in sperm–egg binding and fusion (Vacquier & Moy 1977). Bindin is extremely polymorphic within species, and shows signatures of positive selection between species. One beautiful study of functional differentiation of the alleles showed that sperm with the same bindin genotype as the egg being fertilized had higher fertilization success compared with sperm of another genotype (Palumbi 1999b). Bindin’s egg receptor also shows extensive divergence between species (Kamei et al. 2000). Finally, the abalone sperm protein sp18 has been implicated in sperm–egg fusion (Swanson & Vacquier 1995a), and is perhaps the most rapidly evolving metazoan protein discovered (Swanson & Vacquier 1995b).

**Stages of plant reproduction**

The reproductive structures of angiosperms (flowers) consist of several specialized organs within which the essential stages of plant reproduction take place. Although some flowers are unisexual, typical flowers have both male (stamens) and female organs (pistils). These organs are arrayed adjacent to each other within the innermost whorls of a flower, a proximity effect that is in marked contrast to animals and can result in a high proportion of selfed seed in the absence of barriers to self-fertilization. Plant reproduction involves a novel form of alternation between diploid sporophytic and haploid gametophytic generations. The sporophytic stage predominates the lifecycle of flowering plants, and sporophytic tissues include stamens and pistils. The gametophytic stage is highly reduced and consists of pollen grains and the embryo sac within ovules (Wilson & Yang 2004). This distinction between sporophyte and gametophyte is significant because some of the best-known examples of positive selection of plant reproductive genes involve proteins that regulate the interactions between female sporophytic tis-
sues and the male gametophyte (pollen). We briefly outline several characteristic stages of plant reproduction below, emphasizing the interactive nature of pollination between pollen and tissues of the pistil as well as the female gametophyte. For more detailed reviews, see Higashiyama et al. (2003), Lord & Russell (2002), Swanson et al. (2004) and Wheeler et al. (2001).

**Pollen–stigma interactions**

Pollen grains are multicellular structures surrounded by a protective multilayered cell wall that maintains the male gametophyte in a desiccated, inert state. In the earliest stages of pollination (Fig. 1B), pollen grains are deposited on the terminal structures of the pistil called the stigma. Structural characteristics of stigmas vary widely, but can be broadly classified into two categories including those coated with lipid-rich exudates or dry stigma types. Among taxa with wet stigmas, pollen adheres and hydrates indiscriminately. In contrast, species with dry stigmas show a high degree of selectivity. Initial adherence of pollen to dry stigmas appears to result from the structural complexity of the outer cell wall (exine) allowing for species-specific discrimination (Zinkl et al. 1999). The protein- and lipid-rich pollen coat then mobilizes to the site of pollen–stigma contact, facilitating binding between specific stigmatic and pollen coat proteins (Heizmann et al. 2000) and effectively cross-linking pollen to the stigma surface. Following cross-linking, pollen hydrates via conduits derived from lipids originating in both the pollen coat and stigmatic cells. Hydration of pollen provides both liquid and nutrients necessary to activate metabolism and begin pollen tube growth. There is strong evidence that both cross-linking and hydration of pollen are selectively mediated in taxa with dry-type stigmas, allowing discrimination among species as well as self-recognition (reviewed in Swanson et al. 2004).

Once hydrated, pollen germinates forming a tube that will grow and extend through the stigma and style (Fig. 1B) requiring guidance cues as well as energy and nutrients. Most components of these processes are poorly understood. Initial orientation of the pollen tube at the stigma surface probably involves water gradients (Lush et al. 2000) and chemotaxis via diffusible or substrate-bound factors on the stigma (Kim et al. 2003). Pollen tubes migrate superficially (if a hollow style) or invade through the extracellular spaces of stigmatic tissues requiring digestive enzymes. Stylar tissues then provide a tract for pollen tube guidance from which nutrients and other molecules are absorbed. There is some evidence guidance cues may mediate stylar selectivity, allowing for discrimination among pollen tubes (Shimizu & Okada 2000). In addition, molecules produced by the style and translocated to the pollen tube provide selectivity, including well-studied mechanisms allowing for recognition of self-pollen (reviewed in Kao & Tsukamoto (2004) and McClure (2004)).

At the base of the style lies the ovary, often divided by septa into compartments containing one or more ovules enclosing the egg sac (Fig. 1B). The final guidance cues provided by the pistil direct pollen tubes toward the opening leading to the egg sac (Palanivelu et al. 2003). At this point, a transition occurs between guidance provided by the sporophytic tissues of the pistil and gametophytic tissues including the synergids that flank the egg cell (Higashiyama et al. 2003). After the pollen tube penetrates the egg sac both sperm cells are discharged, one fusing with the egg cell to produce the embryo and the other fusing with the diploid central cell to produce triploid endosperm.

**Rapidly evolving proteins: plants**

In contrast to animals, there are relatively few reproductive proteins from flowering plants known to be under positive selection. This is in part due to the infancy of molecular screens aimed at identifying plant reproductive proteins, despite a long history of interest in such genes as classic examples of adaptive polymorphism (Wright 1939). Below we review the evidence for known cases of positive selection of plant reproductive genes relative to the stage of pollination at which they act, including proteins recently identified in a proteomic screen of pollen coat proteins (Mayfield et al. 2001). Additional screens are underway (Johnson et al. 2004, Marton et al. 2005), and will allow for a more comprehensive investigation as comparative sequence data accumulate.

The initial interactions between pollen and stigma resulting in germination of the male gametophyte represent a primary and crucial point of contact during pollination. The genes contributing to these interactions have been studied extensively and provide several examples of positive selection of plant reproductive proteins including those that contribute to sporophytic self-incompatibility (SI) in the Brassicaceae. In sporophytic SI, a single S-locus of suppressed recombination codes for several SI proteins expressed in the pistil and pollen. These include an S-locus receptor kinase (SRK) (Stein et al. 1991), an S-locus glycoprotein (SLG) (Nasrallah et al. 1991), and an S-locus cysteine-rich protein (SCR) (Schoof et al. 1999, Suzuki et al. 1999). How these genes function to facilitate self-recognition and rejection of self pollen has been reviewed extensively elsewhere (Kachroo et al. 2001, Nasrallah 2002). Self alleles of the pollen coat protein SCR are directly bound by the stigmatic SRK protein, resulting in impaired pollen hydration and germination. Although the stigmatic protein SLG is not necessary for SI, it enhances the response of self-pollen rejection. SRK, SLG and SCR are all highly polymorphic among populations and taxa of Brassicaceae (Nasrallah & Nasrallah 1993, Mable et al. 2003), and alleles are ancient (>20–40 million years (Uyenoyama 1997), consistent with balancing or frequency-dependent selection for recognition loci (Takahata...
& Nei 1990). Adaptive diversification is also evident among all three genes based on $d_{\alpha}/d_{S} > 1$ (Sato et al. 2002, Takebayashi et al. 2003). Thus positive selection drives the adaptive diversification of SRK, SLG and SCR.

In addition to SCR, there are other Brassicaceae pollen coat proteins that show evidence of adaptive diversification. In a comprehensive study of the major pollen coat proteins from Arabidopsis thaliana, Mayfield et al. (2001) identified six lipid-binding oleosin genes. The N-terminus lipid-binding domains (Ting et al. 1998) share only about 50% amino acid identity and evolve more rapidly than adjacent proteins (Mayfield et al. 2001, Fiebig et al. 2004, Schein et al. 2004). The C-terminus of oleosin genes comprising the pollen coat proteins also evolve rapidly due to duplication and deletion of glycin-rich repetitive domains (Fiebig et al. 2004). Although the repetitive nature of the C-terminus precludes robust tests of positive selection based on comparisons such as $d_{\alpha}/d_{S}$, repetitive domains are common features of reproductive proteins and have been proposed as a driving force of positive selection between interacting male and female proteins (Swanson & Vacquier 2002).

Another well-studied example of positive selection in plants involves a second type of self-recognition known as gametophytic SI, which shares a common origin among eudicot families including Rosaceae, Solonaceae and Scrophulariaceae (Igic & Kohn 2001). Gametophytic SI in these plant families has long been known to involve a stylar-expressed extracellular S-locus protein (Anderson et al. 1986) that has RNase activity (S-RNase) (McClure et al. 1989). The pollen component of gametophytic SI has only recently been identified (Sijacic et al. 2004) as an S-locus F-box protein tightly linked to the S-RNase gene (SLF) (Lai et al. 2002). The predominant model of gametophytic SI is that SLF directly binds to S-RNAses of the same haplotype through recognition of variable domains, protecting the cytotoxic S-RNase in these SI crosses from inactivation via ubiquitination, as is believed to occur for non-self pollen haplotypes (Kao & Tsukamoto 2004, McClure 2004; but see Sonneveld et al. 2005 for a different interpretation). As with the proteins mediating sporophytic SI, both S-RNase and SLF show high levels of ancient polymorphism (Loerger et al. 1990, Entani et al. 2003) consistent with balancing or frequency-dependent selection. Similarly, positive selection acts on S-RNase and SLF as $d_{\alpha}/d_{S} > 1$ is evident among functional domains important in recognition of the cognate binding partner (Takebayashi et al. 2003, Ikeda et al. 2004).

Selective forces acting on reproductive proteins

There are several proposed driving forces for the positive selection seen at reproductive loci. While the evolution of plant SI loci is generally understood (i.e. negative frequency-dependent selection), no single hypothesis has been causally linked to the evolution of the remainder of reproductive proteins. Here we present several of these hypotheses of both endogenous and exogenous origin. Predictions made by these hypotheses are discussed, and when possible pertinent cases from material above are highlighted.

Sperm competition is described as post-mating competition between sperm from different males for fertilization of a female’s eggs. Cases of multiple-male mating (polyandry) have been observed in which one male sires a disproportionate amount of eggs (Robinson et al. 1994, Birkhead 1998). Sperm competition predicts a continuous, adaptive ‘arms race’, whose selective intensity should be comparable with the degree of polyandry. Recent work in primate SEMG2 is consistent with this prediction, showing a correlation between degree of positive selection and degree of polyandry (Dorus et al. 2004). Sperm competition could also drive adaptation of male proteins involved in locating, reaching, binding, penetrating and fusing with the egg. Competition between males may even drive adaptation in inseminated proteins which affect sperm storage in the female tract or which affect female behavior, as seen in Drosophila accessory gland proteins.

Sperm competition may direct evolution toward conditions optimal for the male, but female fitness may be optimal under entirely different conditions. Sexual conflict over adaptive optima is thought to lead females and males to counter-adapt, creating a characteristic co-evolutionary chase between male and female characters (Rice & Holland 1997, Gavrilets 2000). In one scenario of sexual conflict, sperm competition leads to fast rates of fertilization, yet females may benefit from a more moderate rate in order to prevent polyspermic fertilization. The larger energy investment put into female gametes makes polyspermy more detrimental to female than male fitness. Consequently, female gamete proteins would evolve to lower the fertilization rate, while sperm proteins would continually attempt to raise it in a context of competition. There may also be sexual conflict operating in Drosophila accessory gland proteins, which manipulate female behavior and sperm storage; these interactions probably have differing optima for females and males.

Sexual selection is widely invoked to explain mating behavior and display, and it may also be operating at the level of gametes (Eberhard 1996). If an egg demonstrates a preference for a certain sperm protein allele, assortative mating results. The fact that sea urchin eggs show affinity to sperm of the same bindin genotype suggests sexual selection as a driving force for divergence (Palumbi 1999a).

Reinforcement is the evolution of reproductive barriers to prevent hybrids. In a case where hybridization between allopatric populations results in offspring of reduced fitness, reinforcement can explain divergence among gamete recognition proteins (Howard 1993). Importantly, reinforcement cannot explain divergence seen in isolated populations, so that the contrast between predictions for allopatric and sympatric populations provides a framework to test reinforcement as a driving force. Particular
test cases for reinforcement include gamete recognition proteins, such as lysin and VERL in abalone and bindin and EBR1 in the sea urchin.

The positive selection of SI loci in plants is thought to principally result from selection to avoid inbreeding. Inbreeding depression is common among natural populations of plants as well as those in horticulture (Cmrak & Roff 1999). If depression is sufficiently strong, it can result in selection for genetic modifiers to avoid inbreeding (Maynard Smith 1971). Once they are established, Wright (1939) showed these loci are subject to negative frequency-dependent selection whereby rare pollen alleles are rejected by pistils at lower rates than common alleles resulting in high allelic diversity within populations. Thus the high levels of polymorphism exhibited by sporophytic SI proteins (SRK, SCR, SLG) as well as gametophytic SI proteins (S-RNase, SLF) reflect the expected outcome of negative frequency-dependent selection acting on genetic loci for avoidance of inbreeding.

Wright's (1939) classic model of negative frequency-dependent selection on SI loci also partially explains adaptive divergence among these SI proteins. Under Wright's model, novel pollen alleles resulting from mutation escape loss due to random genetic drift and are rapidly swept to an equilibrium frequency by selection. If inbreeding depression remains strong, Uyenoyama et al. (2001) showed positive selection can act on compensatory mutations in pistil components of recognition that restore SI, although under complete linkage the mutational pathway for generation of novel functional SI haplotypes in such a two-gene system is complex (Charlesworth 2000, Uyenoyama & Newbiggin 2000). Uyenoyama et al. (2001) showed this process of co-evolution between pollen and pistil components of SI can progress even if the initial mutation at the pollen locus incurs substantial inbreeding depression. In sum, disparate selective forces may drive pollen (increased access to mates) and pistil (avoidance of inbreeding depression) SI proteins despite sharing a common evolutionary history due to linkage at the SI locus (Uyenoyama et al. 2001).

The potential forces discussed above result from endogenous forces of the species' reproductive system. In contrast, pathogen resistance is an exogenous force that may be driving divergence at these loci. Microbial attack may impose a constant pressure for gamete surface proteins to change to elude attackers (Vacquier et al. 1997). Consequently, proteins that recognize these gametes would need to continually adapt to the new surface. Certainly among broadcast spawning invertebrates gametes encounter several microorganisms, and in internal fertilizers sexually transmitted pathogens may pose a threat to gametes.

Several competing hypotheses have been proposed to explain rapid divergence of reproductive characters. It is important to note that several of these hypotheses have overlapping predictions, making their discernment difficult. We expect that diverse approaches to various mating systems can provide clues necessary to explain positive selection of reproductive proteins.

**Significance of reproductive protein evolution**

Throughout this review we have stressed the recurrent observation of rapid evolution, both among different stages of reproduction and across wide taxonomic groups. Rapid evolution is likely to result in functional differences both within and between species, which has significant implications for both evolutionary biologists and human health. For example, evolutionary biologists are interested that rapidly evolving regions of abalone sperm lysin between species have been demonstrated to regulate specificity (Lyon & Vacquier 1999, Yang et al. 2000a), and may be important for reproductive isolation and speciation. Rapidly evolving reproductive proteins could also have implications for human fertility and health. Within humans, approximately 10% of attempted in vitro fertilization trials result in failure, with no known cause assigned (Liu et al. 2001). We hypothesize that these cases of unexplained infertility could be the result of incompatible sperm–egg recognition molecules that arise due to the rapid evolution of reproductive loci. This would be analogous to the need to match MHC or blood type for donations. Similarly, there are implications for crop breeding and agriculture. In plants, the rapid evolution of reproductive proteins might help identify functionally important regions of the molecule that could perhaps be exploited to control the spread of transgenic crops. Thus, an important next step in this field will be to correlate sequence divergence with functional diversification for genes shown to be rapidly evolving. This will be particularly interesting to perform within species; such results may provide clues to understanding the molecular basis of reproductive incompatibilities, which has implications for understanding speciation and infertility.

**Future directions**

It is clear that several classes of reproductive proteins are evolving rapidly across divergent taxonomic groups. To explain this phenomenon researchers must continue to determine its extent by identifying factors in specific reproductive stages (Fig. 1) and conducting comparative studies between orthologs. Several genomic approaches could be utilized to identify new factors, such as mass spectrometric analyses of pollen (Mayfield et al. 2001), semen (Utleg et al. 2003, Fung et al. 2004) or oviductal secretions, expressed sequence tag (EST) sequencing (Swanson et al. 2001b, 2004, Nelson et al. 2002), and microarray analyses (Schlecht & Primig 2003, Wrobel & Primig 2005). It is possible to incorporate evolutionary information into these analyses. For example, comparisons of ESTs from closely related species with one with a completed genome provides a method to identify rapidly evol-
vling proteins (Swanson et al. 2001b, 2004, Barrier et al. 2003). Once putative male and female reproductive genes are identified by these genomic approaches, large-scale proteomic approaches to identifying interacting proteins such as co-affinity purification assays (Li et al. 2004) or pair-wise yeast two-hybrid analyses could be performed (Uetz et al. 2000, Miller et al. 2005) which are more robust than traditional library screening methods. A system-wide picture of selective pressures could be gained by studying the evolution of interacting proteins rather than single factors.

A common goal is to determine the driving forces behind the rapid evolution of these proteins. An important approach to distinguishing between the many hypotheses is to create computational models that provide predictions for comparison with observations from natural or experimental systems. Ultimately, comparison of empirical data with theoretical models will be necessary in order to distinguish the mechanisms driving the rapid evolution of reproductive proteins. We must also describe the consequences of this rapid evolution. For example: does directional selection on gamete-recognition proteins contribute to speciation? Can we measure fitness benefits associated with divergent alleles? Answering such questions may reveal the implications of rapid evolution of reproductive proteins on stock management, agriculture and reproductive health.

Acknowledgements

We thank two reviewers for thorough and insightful comments. W J S was supported by NIH grant HD42563 and NSF grants DEB-0410112 and DEB-0213171. N L C and J E A were supported on training grants from the NIH. The authors declare that there is no conflict of interest that would prejudice the impartiality of this scientific work.

References


Baba T, Azuma S, Kashiwabara S-I & Toyoda Y 1994 Sperm from mice carrying a target mutation of the acrosin gene can penetrate the oocyte zona pellucida and effect fertilization. Journal of Biological Chemistry 269 31845–31849.


Palumbi SR 1999a All males are not created equal: fertility differences depend on gamete recognition polymorphisms in sea urchins. PNAS 96 12632–12637.

Palumbi SR 1999b All males are not created equal: fertility differences depend on gamete recognition polymorphisms in sea urchins. PNAS 96 12632–12637.


Rice WR & Holland B 1997 The enemies within: intergenic conflict, interlocus contest evolution (ICE), and the intraspecific Red Queen. Behavioral Ecology and Sociobiology 41 1–10.


Swanson WJ & Vacquier VD 1995a Liposome fusion induced by a Mr(18000 protein localized to the acrosomal region of acrosome-reacted abalone spermatozoon. Biochemistry 34 14202–14208.


Swanson WJ & Vacquier VD 1997 The abalone egg vitelline envelope receptor for sperm lysin is a giant multivalent molecule. PNAS 94 6724–6729.


Received 20 June 2005
First decision 14 July 2005
Revised manuscript received 1 September 2005
Accepted 13 September 2005