REPRODUCTION
REVIEW

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.

Reproduction (2006) **131** 11–22

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

ISSN 1470–1626 (paper) 1741–7899 (online)
DOI: 10.1530/rep.1.00357
Online version via www.reproduction-online.org

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.* 2000*b*), 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,

Table 1 Statistical tests commonly used to detect positive selection of proteins.

Test	Type of data	Description	Reference
Tajima D	Polymorphism	Tests skew in the allele-frequency spectrum	Tajima (1989)
Fu and Li D	Polymorphism	Compares frequency of recent vs ancient polymorphism, and may use an outgroup	Fu & Li (1993)
Fay and Wu H	Polymorphism	Compares frequency of intermediate and high-frequency derived polymorphism, requires an outgroup	Fay *et al.* (2001)
HKA	Polymorphism	Compares levels of polymorphism within a species to divergence from an outgroup, requires a reference locus	Hudson *et al.* (1987)
Codon models, branch variation	Divergence	Compares variation in d_N/d_S among branches of a phylogeny	Yang & Nielsen (1998)
Codon models, site variation	Divergence	Compares variation in d_N/d_S among sites (codons) of protein coding genes	Yang *et al.* (2000*b*)
Codon models, branch and site variation	Divergence	Compares variation in d_N/d_S among sites for a subset of the branches of a phylogeny	Yang & Nielsen (2002)

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 (Ralt *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 & Wolfner 1999*a*, 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 & Wolfner 1999*b*, 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 & Wolfner 1999*b*). 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) (Wolfner 2002).

Sperm encounter several threats in the female reproductive tract, including foreign pathogens and the female immune system (Austin 1975). *Drosophila* transfers antibacterial 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 (Wassarman *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, Wassarman *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 1995*b*). 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.* 2001*b*, 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_N/d_S 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.* 2001*b*); 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

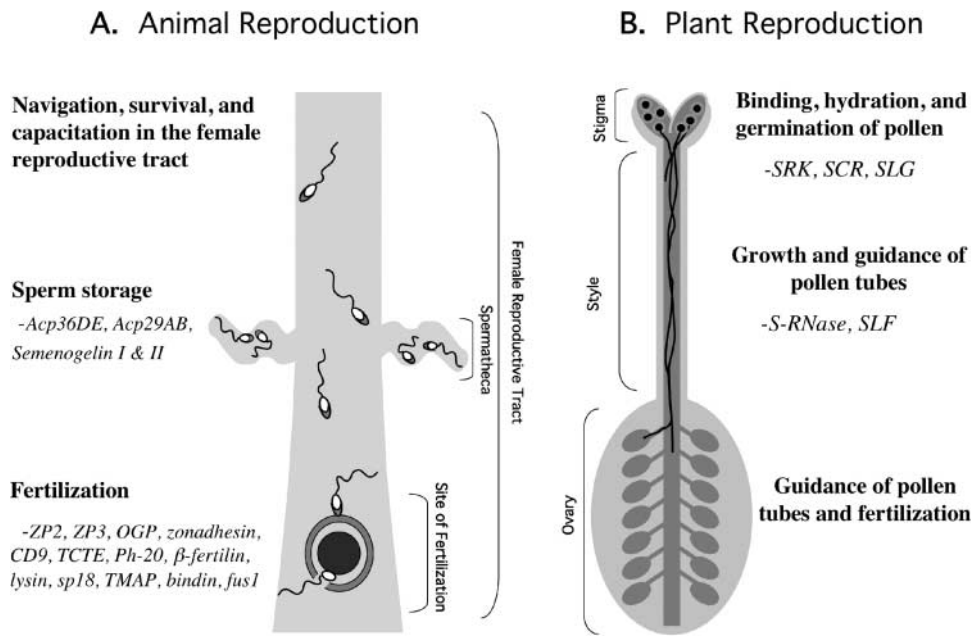

Figure 1 General overview of stages of reproduction for plants and animals. The stages are necessarily oversimplified in order to stress the generality among different taxonomic groups.

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 (Wassarman *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_N/d_S 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.* 2001*a*).

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 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_N/d_S ratios, particular amino acids have been shown to be the target of positive selection (Yang *et al.* 2000*a*). 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 1999*b*). 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 1995*a*), and is perhaps the most rapidly evolving metazoan protein discovered (Swanson & Vacquier 1995*b*).

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 life-cycle 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 synergid cells 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) (Schopfer *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 $d_N/d_S > 1$ (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 glycine-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_N/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 S-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 (Ioerger *et al.* 1990, Entani *et al.* 2003) consistent with balancing or frequency-dependent selection. Similarly, positive selection acts on S-RNAse and SLF as $d_N/d_S > 1$ 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 1999*a*).

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 (Crnokrak & 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 & Newbigin 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.* 2000*a*), 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.* 2001*b*, 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-

ving proteins (Swanson *et al.* 2001*b*, 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 coaffinity 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

Anderson MA, **Cornish EC**, **Mau SL**, **Williams EG**, **Hoggart R**, **Atkinson A**, **Bonig I**, **Grego B**, **Simpson R**, **Roche PJ *et al.*** 1986 Cloning of cDNA for a stylar glycoprotein associated with expression of self-incompatibility in *Nicotiana alata*. *Nature* **321** 38–44.

Anisimova M, **Bielawski JP & Yang Z** 2002 Accuracy and power of Bayes prediction of amino acid sites under positive selection. *Molecular Biology and Evolution* **19** 950–958.

Austin CR 1975 Sperm fertility, viability and persistence in the female tract. *Journal of Reproduction and Fertility Supplement* 75–89.

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.

Barrier M, **Bustamante CD**, **Yu J & Purugganan MD** 2003 Selection on rapidly evolving proteins in the *Arabidopsis genome*. *Genetics* **163** 723–733.

Begun DJ, **Whitley P**, **Todd BL**, **Waldrip-Dail HM & Clark AG** 2000 Molecular population genetics of male accessory gland proteins in *Drosophila*. *Genetics* **156** 1879–1888.

Birkhead TR 1998 Sperm competition in birds. *Reviews in Reproduction* **3** 123–129.

Boatman DE 1997 Responses of gametes to the oviductal environment. *Human Reproduction* **12** 133–149.

Boatman DE & Magnoni GE 1995 Identification of a sperm penetration factor in the oviduct of the golden hamster. *Biology of Reproduction* **52** 199–207.

Burke WH, **Ogasawara FX & Fuqua CL** 1972 A study of the ultrastructure of the uterovaginal sperm-storage glands of the hen, *Gallus domesticus*, in relation to a mechanism for the release of spermatozoa. *Journal of Reproduction and Fertility* **29** 29–36.

Carlson AE, **Westenbroek RE**, **Quill T**, **Ren D**, **Clapham DE**, **Hille B**, **Garbers DL & Babcock DF** 2003 CatSper1 required for evoked Ca^{2+} entry and control of flagellar function in sperm. *PNAS* **100** 14864–14868.

Chapman T & Davies SJ 2004 Functions and analysis of the seminal fluid proteins of male *Drosophila melanogaster* fruit flies. *Peptides* **25** 1477–1490.

Charlesworth D 2000 How can two-gene models of self-incompatibility generate new specificities. *Plant Cell* **12** 309–310.

Civetta A & Singh RS 1995 High divergence of reproductive tract proteins and their association with postzygotic reproductive isolation in *Drosophila melanogaster* and *Drosophila virilis* group species. *Journal of Molecular Evolution* **41** 1085–1095.

Clark NL & Swanson WJ 2005 Pervasive adaptive evolution in primate seminal proteins. *PLoS Genetics* **1** e35.

Coyne JA & Orr HA 2004 *Speciation*. Sunderland: Sinauer Associates.

Crnokrak P & Roff DA 1999 Inbreeding depression in the wild. *Heredity* **83** 260–270.

Darwin C 1859 *On the Origin of Species by Means of Natural Selection, or, the Preservation of Favoured Races in the Struggle for Life*. London: J Murray.

Dorus S, **Evans PD**, **Wyckoff GJ**, **Choi SS & Lahn BT** 2004 Rate of molecular evolution of the seminal protein gene SEMG2 correlates with levels of female promiscuity. *Nature Genetics* **36** 1326–1329.

Eberhard WG 1996 *Female Control: Sexual Selection by Cryptic Female Choice*. Princeton: Princeton University Press.

Entani T, **Iwano M**, **Shiba H**, **Che FS**, **Isogai A & Takayama S** 2003 Comparative analysis of the self-incompatibility (S-) locus region of *Prunus mume*: identification of a pollen-expressed F-box gene with allelic diversity. *Genes to Cells* **8** 203–213.

Fay JC, **Wyckoff GJ & Wu C-I** 2001 Positive and negative selection on the human genome. *Genetics* **158** 1227–1234.

Ferraguti M, **Marotta R & Martin P** 2002 The double sperm line in Isochaetides (Annelida Clitellata, Tubificidae). *Tissue and Cell* **34** 305–314.

Fiebig A, **Kimport R & Preuss D** 2004 Comparisons of pollen coat genes across Brassicaceae species reveal rapid evolution by repeat expansion and diversification. *PNAS* **101** 3286–3291.

Fu YX & Li WH 1993 Statistical tests of neutrality of mutations. *Genetics* **133** 693–709.

Fu YX 1997 Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* **147** 915–925.

Fung KY, **Glode LM**, **Green S & Duncan MW** 2004 A comprehensive characterization of the peptide and protein constituents of human seminal fluid. *Prostate* **61** 171–181.

Gavrilets S 2000 Rapid evolution of reproductive barriers driven by sexual conflict. *Nature* **403** 886–889.

Goldman N & Yang Z 1994 A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Molecular Biology and Evolution* **11** 725–736.

Heizmann P, **Luu DT & Dumas C** 2000 Pollen-stigma adhesion in the Brassicaceae. *Annals of Botany* **85** 23–27.

Higashiyama T, **Kuroiwa H & Kuroiwa T** 2003 Pollen-tube guidance: beacons from the female gametophyte. *Current Opinions in Plant Biology* **6** 36–41.

Howard D 1993 *Reinforcement: Origin, Dynamics, and Fate of an Evolutionary Hypothesis*. In *Hybrid Zones and the Evolutionary Process*. Ed RG Harrison. New York: Oxford University Press.

Hudson RR, **Kreitman M & Aguade M** 1987 A test of neutral molecular evolution based on nucleotide data. *Genetics* **116** 153–159.

Hunter RH & Rodriguez-Martinez H 2004 Capacitation of mammalian spermatozoa *in vivo*, with a specific focus on events in the Fallopian tubes. *Molecular Reproduction and Development* **67** 243–250.

Igic B & Kohn JR 2001 Evolutionary relationships among self-incompatibility RNases. *PNAS* **98** 13167–13171.

Ikeda K, **Igic B**, **Ushijima K**, **Yamane H**, **Hauck NR**, **Nakano R**, **Sassa H**, **Iezzoni AF**, **Kohn JR & Tao R** 2004 Primary structural features of the S haplotype-specific F-box protein, SFB, in *Prunus*. *Sexual Plant Reproduction* **16** 235–243.

Inoue N, **Ikawa M**, **Isotani A & Okabe M** 2005 The immunoglobulin superfamily protein Izumo is required for sperm to fuse with eggs. *Nature* **434** 234–238.

Ioerger TR, **Clark AG & Kao TH** 1990 Polymorphism at the self-incompatibility locus in Solanaceae predates speciation. *PNAS* **87** 9732–9735.

Johnson MA, **von Besser K**, **Zhou Q**, **Smith E**, **Aux G**, **Patton D**, **Levin JZ & Preuss D** 2004 *Arabidopsis* hapless mutations define essential gametophytic functions. *Genetics* **168** 971–982.

Kachroo A, **Schopfer CR**, **Nasrallah ME & Nasrallah JB** 2001 Allele-specific receptor-ligand interactions in *Brassica* self-incompatibility. *Science* **293** 1824–1826.

Kaji K, **Oda S**, **Shikano T**, **Ohnuki T**, **Uematsu Y**, **Sakagami J**, **Tada N**, **Miyazaki S & Kudo A** 2000 The gamete fusion process is defective in eggs of CD9-deficient mice. *Nature Genetics* **24** 279–282.

Kamei N, **Swanson WJ & Glabe CG** 2000 A rapidly diverging EGF protein regulates species-specific signal transduction in early sea urchin development. *Developmental Biology* **225** 267–276.

Kao TH & Tsukamoto T 2004 The molecular and genetic bases of S-RNase-based self-incompatibility. *Plant Cell* **16** (Suppl) S72–S83.

Kelly RW & Critchley HO 1997 Immunomodulation by human seminal plasma: a benefit for spermatozoon and pathogen? *Human Reproduction* **12** 2200–2207.

Kim S, **Mollet JC**, **Dong J**, **Zhang K**, **Park SY & Lord EM** 2003 Chemocyanin, a small basic protein from the lily stigma, induces pollen tube chemotropism. *PNAS* **100** 16125–16130.

Kingan SB, **Tatar M & Rand DM** 2003 Reduced polymorphism in the chimpanzee semen coagulating protein semenogelin I. *Journal of Molecular Evolution* **57** 159–169.

Kreitman M 2000 Methods to detect selection in populations with applications to the human. *Annual Review of Genomics and Human Genetics* **1** 539–559.

Lai Z, **Ma W**, **Han B**, **Liang L**, **Zhang Y**, **Hong G & Xue Y** 2002 An F-box gene linked to the self-incompatibility (S) locus of *Antirrhinum* is expressed specifically in pollen and tapetum. *Plant Molecular Biology* **50** 29–42.

Le Naour F, **Rubinstein E**, **Jasmin C**, **Prenant M & Boucheix C** 2000 Severely reduced female fertility in CD9-deficient mice. *Science* **287** 319–321.

Lewis CA, **Talbot CF & Vacquier VD** 1982 A protein from abalone sperm dissolves the egg vitelline layer by a nonenzymatic mechanism. *Developmental Biology* **92** 227–239.

Li S, **Armstrong CM**, **Bertin N**, **Ge H**, **Milstein S**, **Boxem M**, **Vidalain PO**, **Han JD**, **Chesneau A & Hao T *et al.*** 2004 A map of the interactome network of the metazoan *C. elegans*. *Science* **303** 540–543.

Liu DY, **Clarke GN**, **Martic M**, **Garrett C & Baker HW** 2001 Frequency of disordered zona pellucida (ZP)-induced acrosome reaction in infertile men with normal semen analysis and normal spermatozoa-ZP binding. *Human Reproduction* **16** 1185–1190.

Lord EM & Russell SD 2002 The mechanisms of pollination and fertilization in plants. *Annual Review of Cell and Developmental Biology* **18** 81–105.

Lung O, **Kuo L & Wolfner MF** 2001 *Drosophila* males transfer antibacterial proteins from their accessory gland and ejaculatory duct to their mates. *Journal of Insect Physiology* **47** 617–622.

Lush WM, **Spurck T & Joosten R** 2000 Pollen tube guidance by the pistil of a solanaceous plant. *Annals of Botany* **85** 39–47.

Lyon JD & Vacquier VD 1999 Interspecies chimeric sperm lysins identify regions mediating species-specific recognition of the abalone egg vitelline envelope. *Developmental Biology* **214** 151–159.

Mable BK, **Schierup MH & Charlesworth D** 2003 Estimating the number, frequency, and dominance of S-alleles in a natural population of *Arabidopsis lyrata* (Brassicaceae) with sporophytic control of self-incompatibility. *Heredity* **90** 422–431.

Makalowski W & Boguski MS 1998 Evolutionary parameters of the transcribed mammalian genome: an analysis of 2820 orthologous rodent and human sequences. *PNAS* **95** 9407–9412.

Marton ML, **Cordts S**, **Broadhvest J & Dresselhaus T** 2005 Micropylar pollen tube guidance by egg apparatus 1 of maize. *Science* **307** 573–576.

Mayfield JA, **Fiebig A**, **Johnstone SE & Preuss D** 2001 Gene families from the *Arabidopsis thaliana* pollen coat proteome. *Science* **292** 2482–2485.

Maynard Smith J 1971 The origin and maintainance of sex. In *Group Selection*, pp 163–175. Ed GC Williams. Chicago: Aldine-Artherton.

McClure B 2004 S-RNase and SLF determine S-haplotype-specific pollen recognition and rejection. *Plant Cell* **16** 2840–2847.

McClure BA, **Haring V**, **Ebert PR**, **Anderson MA**, **Simpson RJ**, **Sakiyama F & Clarke AE** 1989 Style self-incompatibility gene products of *Nicotiana alata* are ribonucleases. *Nature* **342** 955–957.

Metz EC, **Robles-Sikisaka R & Vacquier VD** 1998 Nonsynonymous substitution in abalone sperm fertilization genes exceeds substitution in introns and mitochondrial DNA. *PNAS* **95** 10676–10681.

Miller JP, **Lo RS**, **Ben-Hur A**, **Desmarais C**, **Stagljar I**, **Noble WS & Fields S** 2005 Large-scale identification of yeast integral membrane protein interactions. *PNAS* **102** 12123–12128.

Miyado K, **Yamada G**, **Yamada S**, **Hasuwa H**, **Nakamura Y**, **Ryu F**, **Suzuki K**, **Kosai K**, **Inoue K & Ogura A *et al.*** 2000 Requirement of CD9 on the egg plasma membrane for fertilization. *Science* **287** 321–324.

Mueller JL, **Ripoll DR**, **Aquadro CF & Wolfner MF** 2004 Comparative structural modeling and inference of conserved protein classes in *Drosophila* seminal fluid. *PNAS* **101** 13542–13547.

Mueller JL, Ravi Ram K, McGraw LA, Bloch Qazi MC, Siggia ED, Clark AG, Aquadro CF & Wolfner MF 2005 Cross-species comparison of *Drosophila* male accessory gland protein genes. *Genetics* **171** 131–143.

Nasrallah JB 2002 Recognition and rejection of self in plant reproduction. *Science* **296** 305–308.

Nasrallah JB & Nasrallah ME 1993 Pollen-stigma signaling in the sporophytic self-incompatibility response. *Plant Cell* **5** 1325–1335.

Nasrallah JB, **Nishio T & Nasrallah ME** 1991 The self-incompatibility genes of *Brassica* – expression and use in genetic ablation of floral tissues. *Annual Review of Plant Physiology and Plant Molecular Biology* **42** 393–422.

Nelson PS, **Pritchard C**, **Abbott D & Clegg N** 2002 The human (PEDB) and mouse (mPEDB) prostate expression databases. *Nucleic Acids Research* **30** 218–220.

Neubaum DM & Wolfner MF 1999a*a* Mated *Drosophila melanogaster* females require a seminal fluid protein, Acp36DE, to store sperm efficiently. *Genetics* **153** 845–857.

Neubaum DM & Wolfner MF 1999b*b* Wise, winsome, or weird? Mechanisms of sperm storage in female animals. *Current Topics in Developmental Biology* **41** 67–97.

Nielsen R & Yang Z 1998 Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics* **148** 929–936.

Palanivelu R, **Brass L**, **Edlund AF & Preuss D** 2003 Pollen tube growth and guidance is regulated by POP2, an *Arabidopsis* gene that controls GABA levels. *Cell* **114** 47–59.

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

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

Podlaha O & Zhang J 2003 Positive selection on protein-length in the evolution of a primate sperm ion channel. *PNAS* **100** 12241–12246.

Podlaha O, **Webb DM**, **Tucker PK & Zhang J** 2005 Positive selection for indel substitutions in the rodent sperm protein Catsper1. *Molecular Biology and Evolution* **22** 1845–1852.

Racey PA 1979 The prolonged storage and survival of spermatozoa in Chiroptera. *Journal of Reproduction and Fertility* **56** 391–402.

Ralt D, **Goldenberg M**, **Fetterolf P**, **Thompson D**, **Dor J**, **Mashiach S**, **Garbers DL & Eisenbach M** 1991 Sperm attraction to a follicular factor(s) correlates with human egg fertilizability. *PNAS* **88** 2840–2844.

Rankin TL, **Coleman JS**, **Epifano O**, **Hoodbhoy T**, **Turner SG**, **Castle PE**, **Lee E**, **Gore-Langton R & Dean J** 2003 Fertility and taxon-specific sperm binding persist after replacement of mouse sperm receptors with human homologs. *Developmental Cell* **5** 33–43.

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

Riffell JA, **Krug PJ & Zimmer RK** 2002 Fertilization in the sea: the chemical identity of an abalone sperm attractant. *Journal of Experi mental Biology* **205** 1439–1450.

Riffell JA, **Krug PJ & Zimmer RK** 2004 The ecological and evolutionary consequences of sperm chemoattraction. *PNAS* **101** 4501–4506.

Robinson T, **Johnson NA & Wade MJ** 1994 Postcopulatory, prezygotic isolation: intraspecific and interspecific sperm precedence in *Tribolium* spp., flour beetles. *Heredity* **73** 155–159.

Sainudiin R, **Wong WS**, **Yogeeswaran K**, **Nasrallah JB**, **Yang Z & Nielsen R** 2005 Detecting site-specific physicochemical selective pressures: applications to the class I HLA of the human major histocompatibility complex and the SRK of the plant sporophytic self-incompatibility system. *Journal of Molecular Evolution* **60** 315–326.

Sato K, **Nishio T**, **Kimura R**, **Kusaba M**, **Suzuki T**, **Hatakeyama K**, **Ockendon DJ & Satta Y** 2002 Coevolution of the S-locus genes SRK, SLG and SP11/SCR in *Brassica oleracea* and *B. rapa*. *Genetics* **162** 931–940.

Schein M, **Yang ZH**, **Mitchell-Olds T & Schmid KJ** 2004 Rapid evolution of a pollen-specific oleosin-like gene family from *Arabidopsis thaliana* and closely related species. *Molecular Biology and Evolution* **21** 659–669.

Schlecht U & Primig M 2003 Mining meiosis and gametogenesis with DNA microarrays. *Reproduction* **125** 447–456.

Schopfer CR, **Nasrallah ME & Nasrallah JB** 1999 The male determinant of self-incompatibility in *Brassica*. *Science* **286** 1697–1700.

Sever DM & Hamlett WC 2002 Female sperm storage in reptiles. *Journal of Experimental Zoology* **292** 187–199.

Shimizu KK & Okada K 2000 Attractive and repulsive interactions between female and male gametophytes in *Arabidopsis* pollen tube guidance. *Development* **127** 4511–4518.

Sijacic P, **Wang X**, **Skirpan AL**, **Wang Y**, **Dowd PE**, **McCubbin AG**, **Huang S & Kao TH** 2004 Identification of the pollen determinant of S-RNase-mediated self-incompatibility. *Nature* **429** 302–305.

Sonneveld T, **Tobutt KR**, **Vaughan SP & Robbins TP** 2005 Loss of pollen-S function in two self-compatible selections of *Prunus avium* is associated with deletion/mutation of an S haplotype-specific F-box gene. *Plant Cell* **17** 37–51.

Stein JC, **Howlett B**, **Boyes DC**, **Nasrallah ME & Nasrallah JB** 1991 Molecular cloning of a putative receptor protein kinase gene encoded at the self-incompatibility locus of *Brassica oleracea*. *PNAS* **88** 8816–8820.

Suzuki G, **Kai N**, **Hirose T**, **Fukui K**, **Nishio T**, **Takayama S**, **Isogai A**, **Watanabe M & Hinata K** 1999 Genomic organization of the S locus: identification and characterization of genes in SLG/SRK region of S(9) haplotype of *Brassica campestris* (syn. *rapa*). *Genetics* **153** 391–400.

Swanson R, **Edlund AF & Preuss D** 2004 Species specificity in pollen-pistil interactions. *Annual Review of Genetics* **38** 793–818.

Swanson WJ & Vacquier VD 1995*a* Liposome fusion induced by a M(r) 18 000 protein localized to the acrosomal region of acrosome-reacted abalone spermatozoa. *Biochemistry* **34** 14202–14208.

Swanson WJ & Vacquier VD 1995*b* Extraordinary divergence and positive Darwinian selection in a fusagenic protein coating the acrosomal process of abalone spermatozoa. *PNAS* **92** 4957–4961.

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

Swanson WJ & Vacquier VD 2002 Rapid evolution of reproductive proteins. *Nature Reviews. Genetics* **3** 137–144.

Swanson WJ, **Yang Z**, **Wolfner MF & Aquadro CF** 2001*a* Positive Darwinian selection drives the evolution of several female reproductive proteins in mammals. *PNAS* **98** 2509–2514.

Swanson WJ, **Clark AG**, **Waldrip-Dail HM**, **Wolfner MF & Aquadro CF** 2001*b* Evolutionary EST analysis identifies rapidly evolving male reproductive proteins in *Drosophila*. *PNAS* **98** 7375–7379.

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

Swanson WJ, **Wong A**, **Wolfner MF & Aquadro CF** 2004 Evolutionary expressed sequence tag analysis of *Drosophila* female reproductive tracts identifies genes subjected to positive selection. *Genetics* **168** 1457–1465.

Tajima F 1989 Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123** 585–595.

Takahata N & Nei M 1990 Allelic genealogy under overdominant and frequency-dependent selection and polymorphism of major histocompatibility complex loci. *Genetics* **124** 967–978.

Takebayashi N, **Brewer PB**, **Newbigin E & Uyenoyama MK** 2003 Patterns of variation within self-incompatibility loci. *Molecular Biology and Evolution* **20** 1778–1794.

Ting JTL, **Wu SSH**, **Ratnayake C & Huang AHC** 1998 Constituents of the tapetosomes and elaioplasts in *Brassica campestris* tapetum and their degradation and retention during microsporogenesis. *Plant Journal* **16** 541–551.

Torgerson DG, **Kulathinal RJ & Singh RS** 2002 Mammalian sperm proteins are rapidly evolving: evidence of positive selection in functionally diverse genes. *Molecular Biology and Evolution* **19** 1973–1980.

Tsaur SC & Wu CI 1997 Positive selection and the molecular evolution of a gene of male reproduction, Acp26Aa of *Drosophila*. *Molecular Biology and Evolution* **14** 544–549.

Uetz P, **Giot L**, **Cagney G**, **Mansfield TA**, **Judson RS**, **Knight JR**, **Lockshon D**, **Narayan V**, **Srinivasan M**, **Pochart P *et al.*** 2000 A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae* [see comments]. *Nature* **403** 623–627.

Utleg AG, **Yi EC**, **Xie T**, **Shannon P**, **White JT**, **Goodlett DR**, **Hood L & Lin B** 2003 Proteomic analysis of human prostasomes. *Prostate* **56** 150–161.

Uyenoyama MK 1997 Genealogical structure among alleles regulating self-incompatibility in natural populations of flowering plants. *Genetics* **147** 1389–1400.

Uyenoyama MK & Newbigin E 2000 Evolutionary dynamics of dual specificity self-incompatibility alleles. *Plant Cell* **12** 310–312.

Uyenoyama MK, **Zhang Y & Newbigin E** 2001 On the origin of self-incompatibility haplotypes: transition through self-compatible intermediates. *Genetics* **157** 1805–1817.

Vacquier VD 1998 Evolution of gamete recognition proteins. *Science* **281** 1995–1998.

Vacquier VD & Moy GW 1977 Isolation of bindin: the protein responsible for adhesion of sperm to sea urchin eggs. *PNAS* **74** 2456–2460.

Vacquier VD, **Swanson WJ & Lee YH** 1997 Positive Darwinian selection on two homologous fertilization proteins: what is the selective pressure driving their divergence? *Journal of Molecular Evolution* **44** S15–S22.

Vilela-Silva AC, **Castro MO**, **Valente AP**, **Biermann CH & Mourao PA** 2002 Sulfated fucans from the egg jellies of the closely related sea urchins *Strongylocentrotus droebachiensis* and *Strongylocentrotus pallidus* ensure species-specific fertilization. *Journal of Biological Chemistry* **277** 379–387.

Wang X, **Hughes AL**, **Tsukamoto T**, **Ando T & Kao T** 2001 Evidence that intragenic recombination contributes to allelic diversity of the S-RNase gene at the self-incompatibility (S) locus in *Petunia inflata*. *Plant Physiology* **125** 1012–1022.

Wassarman PM, **Jovine L & Litscher ES** 2001 A profile of fertilization in mammals. *Nature Cell Biology* **3** 59–64.

Wassarman PM, **Jovine L**, **Litscher ES**, **Qi H & Williams Z** 2004 Egg-sperm interactions at fertilization in mammals. *European Journal of Obstetrics, Gynecology, and Reproductive Biology* **115** Suppl 1 S57–S60.

Wheeler MJ, **Franklin-Tong VE & Franklin FCH** 2001 The molecular and genetic basis of pollen-pistil interactions. *New Phytologist* **151** 565–584.

Wilson ZA & Yang C 2004 Plant gametogenesis: conservation and contrasts in development. *Reproduction* **128** 483–492.

Wolfner MF 2002 The gifts that keep on giving: physiological functions and evolutionary dynamics of male seminal proteins in *Drosophila*. *Heredity* **88** 85–93.

Wright S 1939 The distribution of self-sterility alleles in populations. *Genetics* **24** 538–552.

Wrobel G & Primig M 2005 Mammalian male germ cells are fertile ground for expression profiling of sexual reproduction. *Reproduction* **129** 1–7.

Yanagimachi R 1994 *Mammalian fertilization*. In *The Physiology of Reproduction*, pp 189–317. Eds E Knobil & JD Neill. New York: Raven Press Ltd.

Yang Z 1998 Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Molecular Biology and Evolution* **15** 568–573.

Yang Z & Nielsen R 1998 Synonymous and non synonymous rate variation in nuclear genes of mammals. *Journal of Molecular Evolution* **46** 409–418.

Yang Z & Bielawski JP 2000 Statistical methods for detecting molecular adaptation. *Trends in Ecology and Evolution* **15** 496–503.

Yang Z & Nielsen R 2002 Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Molecular Biology and Evolution* **19** 908–917.

Yang Z, **Swanson WJ & Vacquier VD** 2000*a* Maximum-likelihood analysis of molecular adaptation in abalone sperm lysin reveals variable selective pressures among lineages and sites. *Molecular Biology and Evolution* **17** 1446–1455.

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

Zinkl GM, **Zwiebel BI**, **Grier DG & Preuss D** 1999 Pollen-stigma adhesion in *Arabidopsis*: a species-specific interaction mediated by lipophilic molecules in the pollen exine. *Development* **126** 5431–5440.

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