Let's talk about sex! Testing sexual selection during the evolution of mammalian sperm proteins.

By Rebecca Bergen

Supervisor: Dr. A. Civetta

A thesis submitted in partial fulfillment of the Honours Thesis (05.4111/6) Course, Department of Biology, The University of Winnipeg 2008

Abstract

Fertilization in mammals is a complex multi-step process. A number of sperm proteins have been identified as playing different roles throughout the process of fertilization. These proteins are needed for sperm competitiveness during fertilization and are involved in male – female molecular interactions. Different molecular evolution studies have identified most of these proteins as evolving under positive (adaptive) selection between species and the assumption is usually made that sexual selection has shaped their evolution. The candidate mammalian sperm proteins in this research project were acrosin, ADAM1, ADAM2, ADAM3, calmegin, GAPDS, PRM1, SEMG2, SPA17, SPACA3, SPAM, and zonadhesin. To test whether these sperm proteins have evolved through sexual selection, sequence data analysis within a phylogenetic context was used in conjunction with information on the species mating systems. DNA and amino acid sequences were retrieved from GenBank (www.ncbi.nlm.nih.gov/). The sequences were then used to search the GenBank database for similar sequences in different species of mammals using a sequence similarity search algorithm. The alignment algorithm CLUSTAL was used to align each gene's DNA and amino acid sequence among species. To test for evidence of positive selection and changes in evolutionary rates among species the data were analyzed with PAML (phylogenetic analysis by maximum likelihood). Of particular interest were those phylogenetic branches that led to promiscuous species. Results show evidence of positive selection for most genes analyzed and there is evidence of significant differences in rates of evolution along lineages. However, no gene shows accelerated evolution linked to sexual promiscuity.

Acknowledgements

I would like to thank my supervisor, Dr. A. Civetta, for all his help and for offering me this opportunity; my committee members, Dr. J. Franck and Dr. M. Silcox, for their guidance and support; and Dr. R. Moodie, the program coordinator, for keeping us informed and on track. I would also like to thank my family and friends for supporting me this last year.

Table of Contents

Abstract.....ii

Acknowledgements.....iii

Table of Contents.....iv

List of Tables.....vi

List of Figures.....vii

Introduction.....1

- Sexual Selection.....1
- Molecular Evolution and Positive Selection.....3
- Fertilization.....4
- Candidate Proteins.....5
- Objective.....8

Methods.....9

Results.....12

- DNA and Protein Accession Numbers.....12
- Acrosin PAML Results and Phylogeny.....15
- ADAM1 PAML Results and Phylogeny.....17
- ADAM2 PAML Results and Phylogeny.....17
- ADAM3 PAML Results and Phylogeny.....19
- Calmegin PAML Results and Phylogeny.....20
- GAPDS PAML Results and Phylogeny.....23
- SEMG2 PAML Results and Phylogeny.....25

SPACA3 PAML Results and Phylogeny.....27
SPAM PAML Results and Phylogeny.....28
Overall Summary of Results.....29

Discussion.....32
PAML Results and Positive Selection.....32
Future Avenues.....33

Conclusions.....37

References.....38

List of Tables

Table 1: Accession Numbers.....12

Table 2: PAML Results for Acrosin.....16

Table 3: ADAM1 PAML Results.....18

Table 4: ADAM2 PAML Results.....20

Table 5: ADAM3 PAML Results.....21

Table 6: CalmeginA PAML Results.....22

Table 7: CalmeginB PAML Results.....23

Table 8: PAML Results for GAPDS.....25

Table 9: SEMG2 PAML Results.....26

Table 10: SPACA3 PAML Results.....28

Table 11: PAML Results for SPAM.....29

Table 12: Summary of PAML Results.....31

List of Figures

Figure 1: Acrosin Alignment.....10

Figure 2: Acrosin Phylogeny.....17

Figure 3: ADAM1 Phylogeny.....19

Figure 4: ADAM2 Phylogeny.....21

Figure 5: ADAM3 Phylogeny.....22

Figure 6: CalmeginA Phylogeny.....24

Figure 7: CalmeginB Phylogeny.....24

Figure 8: GAPDS Phylogeny.....26

Figure 9: SEMG2 Phylogeny.....27

Figure 10: SPACA3 Phylogeny.....29

Figure 11: SPAM Phylogeny.....30

Introduction

Mammalian fertilization is a complex process that requires the sperm crossing a series of barriers before binding to the egg membrane. During this process, a series of sperm and egg expressed gene products interact in complex manners (Primakoff & Myles, 2002). In a number of species rapid molecular evolution has been detected in sex-related genes (Civetta & Singh, 1999; Van Doorn *et al.*, 2001; Civetta 2003a). Does a species mating system influence the rate and mode of evolution of sperm genes that are involved in fertilization? Several candidate genes were selected for this study. As the proteins produced by the candidate genes are involved in male-female interactions of fertilization, the assumption has been previously made that sexual selection through competition between members of the same sex (sperm competition) or through interactions between members of the opposite sex (female cryptic choice) have shaped the evolution of the these proteins. If so, the rate of evolution of these sperm proteins should be affected by the species mating system, with species that display promiscuous or polyandrous mating systems undergoing faster rates of evolution driven by selection. Promiscuous or polyandrous species are those where the female of the species are mating with multiple males.

Sexual Selection

In 1859 Darwin defined sexual selection and its mechanism as “a struggle between males for possession of the females” which would not result in death, but that the unsuccessful competitor would have few or no offspring (O’Donald, 1980). It is in essence the competition for mates, which ultimately extends to competition for sperm-

egg fusion. Today it is recognized that sexual selection can promote reproductive isolation and speciation (Gage *et al.*, 2002). It is thought that sperm surface proteins have evolved rapidly so as to provide a proper interaction with their counterpart on the egg surface (Van Doorn *et al.*, 2001). Mammalian males have a higher potential for fertilization than do the females as they are only limited by the number of ova available, which need to be produced by the female and require a higher investment of energy. Females represent a limiting resource at this stage, as the males' investment into the gametes is small. As a result, males are not picky in their choice of mates. One male can fertilize any number of females, while a female can only be fertilized by one male at a time. This leads females to be choosy about who is going to fertilize their available eggs (O'Donald, 1980). Sexual selection can continue after mating through sperm competition and/or female cryptic choice. Logically then multi-male, multi-female, or polyandrous species would be likely to display both sperm competition and female cryptic choice.

In those species where the female is mating with more than one male, sperm competition is a development that could insure one male's sperm will be more likely to fertilize an egg over another male's sperm (Andersson & Simmons, 2006). The ability of sperm to be competitive can be achieved through a number of ways. The raffle principle operates on the premise that the more sperm a male enters into the competition, the greater the likelihood that one of his sperm will ultimately fertilize the egg. Sperm morphology and motility can also be a factor. The male that produces sperm that are designed to be quicker and that are motile in greater numbers would be more likely to out-compete other males (Stockley, 2004). Finally, variation at specific genes can confer differences in the ability of one males' sperm to out-compete another males' sperm

(Clark *et al.*, 1999; Fiumera *et al.*, 2005; Fiumera *et al.*, 2007). We also know that, at least in *Drosophila*, sperm competitive fertilization ability is influenced by female genotype and involve complex male-female interactions suggesting female cryptic choice (Clark & Begun, 1998l; Clark *et al.*, 1999).

Molecular Evolution and Positive Selection

There are many fertilization proteins that seem to be conserved across the invertebrate and vertebrate phyla. These proteins are those involved in the biochemical reactions of the sperm interior (i.e. activation of sperm ion channels to induce the acrosome reaction, and fusion of acrosomal vesicle with sperm plasma membrane) and the regulation of sperm motility. Amino acids predicted to undergo positive selection fall in the domain of sperm-egg binding, which suggests that the selection force relates to male-female interaction (i.e. fertilization). Genes that have been identified that show a high divergence between species are good candidates to examine the molecular foundations of species diversity, as these genes may have had an important role in the beginning stages of the formation of new species (Torgerson *et al.*, 2002). Those genes that code for proteins in the male reproductive tract have been shown to evolve rapidly. There are a couple of explanations for this phenomenon. Sperm competition between males is driving this rapid divergence of sperm proteins to increase each males' probability of successful fertilization. Another explanation is that the critical reproductive function that sperm proteins have may cause them to possess significant selective constraints which could cause them to display similar divergence patterns as other tissue-specific genes (Torgerson *et al.*, 2002). If the male reproductive genes are evolving fast, then so to must their female counterparts, as for successful fertilization the sperm and egg

must recognize and interact with one another. This male-female coevolution may also be a mechanism of sexual selection. Even if a males' sperm outcompetes another males' sperm in reaching the egg, if it cannot fertilize it, that male would not be sexually fit (Torgerson *et al.*, 2002).

Fertilization

Mammals reproduce, for the most part, via internal fertilization. The sperm has to travel through the female reproductive tract to reach the egg and fertilize it. Sperm and egg remain separate until the time of fertilization but they both must develop and mature to be complementary.

Sperm initially develop in the testis. As the sperm pass through the epididymis, the rest of the male reproductive tract and then further through the female reproductive tract, environmental cues cause the sperm to reach full maturity. The sperm are now said to be capacitated. Being capacitated means the sperm have altered patterns of motility and that they can perform regulated exocytosis (i.e. the acrosome reaction) (Primakoff & Myles, 2007).

Fertilization is complete when the sperm directly interacts with the egg. This process requires that the sperm and egg recognize each other and that they have the ability to form a bond. This is accomplished by gamete surface proteins (Primakoff & Myles, 2002). First, the sperm must penetrate the cumulus, a layer of ovarian follicular cells that surrounds the egg. It is thought that this is accomplished through hyper-motility and a surface hyaluronidase that helps digest a pathway through this extracellular matrix. The sperm then interacts with the zona pellucida. This extracellular matrix is composed

of three glycoproteins, ZP1, ZP2, and ZP3. Sperm adhesion to the zona pellucida is mediated by an O-linked carbohydrate on ZP3 (Primakoff & Myles, 2002). Located on the head of the sperm is the acrosome, a membrane-bound, lysosome-like organelle that through exocytosis of enzymes, allows the sperm to digest through the zona pellucida and adhere to and fuse with the egg plasma membrane (Wassarman, 1987). This ZP3-induced exocytosis occurs through two sperm signalling pathways. ZP3 binds to galactosyltransferase (GalT), a receptor on the sperm, which results in elevated levels of cytoplasmic calcium due to the activation of a heterotrimeric GTP-binding protein and phospholipase C. The binding of ZP3 and GalT also stimulates a transient influx of calcium through T-type channels. These processes allow the acrosome reaction to occur, allowing the sperm to penetrate the zona pellucida (Primakoff & Myles, 2002). Once the sperm has penetrated the zona pellucida it is now free to fuse with the egg plasma membrane. Unlike the binding of the sperm to the zona pellucida, the fusion of sperm and egg is not particularly species specific (Wassarman, 1987). However, the process of internal fertilization tends to negate accidental interspecies fertilization, as could occur in external fertilizers. The natural complexity and multi-step process of reproduction provide many opportunities for protein coevolution between sperm and egg.

Candidate Proteins

A number of sperm proteins have been carefully identified as having a role in the journey sperm follow towards fertilization. Different molecular evolution studies have identified some proteins as evolving under positive (adaptive) selection between species (Torgerson *et al.*, 2002, Civetta, 2003b, Swanson *et al.*, 2003, Dorus *et al.*, 2004, and Glassey & Civetta, 2004).

One protein that has been identified as being necessary for sperm fertility is calmegin. Calmegin is a testis-specific molecular chaperone that enables sperm migration into the oviduct and is essential for sperm adhesion to the zona pellucida, as well as the egg plasma membrane. It is a calcium binding protein that is localized on the endoplasmic reticulum membrane where it functions in chaperoning the formation of dimerized proteins (Ikawa *et al.*, 2001).

Glyceraldehyde-3-phosphate dehydrogenase (GAPDS) is another sperm-specific protein. Located on the fibrous sheath, a cytoskeletal structure that encapsulates the principal piece of the sperm, GAPDS is involved in sperm motility. It is essential for successful fertilization that sperm be motile. The flagellum of the sperm generates motility but requires energy in the form of ATP to support the coordinated movements of the central axoneme and flagellar components. Most of this energy is generated by glycolysis. GAPDS is an enzyme that is necessary for the glycolytic pathway (Miki *et al.*, 2004).

Upon ejaculation, mammalian semen thickens, varying from a consistency of viscous gel in humans to a hard, rubbery copulation plug seen in rats. This thickening provides several functions, including maintaining the presence of sperm in the female reproductive tract to preventing subsequent males from inseminating a female that has recently been mated. Semenogelin2 (SEMG2), a high molecular weight protein, is involved in this semen coagulation. Within the female reproductive tract are proteolytic enzymes that eventually degrade the coagulum so that the sperm are then free to continue towards the egg (Jensen-Seaman & Li, 2003).

To penetrate the cumulus, a substantial cell barrier surrounding the egg, sperm use hyperactivity and a glycosylphosphatidylinositol-anchored surface hyaluronidase, called PH-20, or SPAM (sperm adhesion molecule) (Primakoff & Myles, 2002). It is located on the plasma membrane and the inner acrosomal membrane. It has also been implicated in the binding of sperm to the zona pellucida (Lin *et al.*, 1993). Other proteins thought to be involved in sperm-egg interaction at the zona pellucida and/or the egg plasma membrane are some members of the ADAM gene family, sperm autoantigenic protein 17, (Spa17), acrosin, and zonadhesin (ZAN). The ADAM genes are composed of a disintegrin and metalloprotease and are transmembrane proteins, sharing a conserved multidomain structure. These proteins are expressed predominantly in the testis (Kim *et al.*, 2006). ADAM1 and ADAM2 function in the transport of specific sperm proteins which include ADAM3, from the endoplasmic reticulum of the testicular germ cells onto the cell surface. This movement to the cell surface allow these proteins to function in sperm migration into the oviduct, dispersal of cumulus cells, and aid in binding sperm to the zona pellucida (Nishimura *et al.*, 2004). Spa17, a zona pellucida binding protein, is one of the most adhesive, high-affinity binding molecules (Wen *et al.*, 2001). Acrosin, a multifunctional protein, is abundant in the acrosome. It is released during the acrosome reaction, binds to the zona pellucida, and disperses the acrosomal contents (Gatesy and Swanson, 2007). Another multidomain, transmembrane protein is ZAN. ZAN is released from the sperm during the acrosome reaction but remains in the acrosomal shroud, where it functions to bind the shroud to the zona pellucida (Lea *et al.*, 2001). SPACA3, a lysozyme-like gene, is expressed predominantly in the epididymis. It is an intra-acrosomal sperm protein that is involved in sperm-egg binding (Zhang *et al.*, 2005).

During the final stages of mammalian spermatogenesis, the sperm undergo considerable morphological change, namely, the replacement of histones with protamines during the restructuring of the genome in the nucleus. Protamine1 (Prm1) is a sperm-specific protein that is involved in this chromatin condensation. This is done so sperm can maintain a streamlined shape (Martins *et al.*, 2004).

Objective

It has generally been assumed that those genes that are involved in fertilization and reproduction are evolving under sexual selection. In this study, evidence of selection at the molecular level of sperm proteins and whether or not a species mating system influences the rate and mode of evolution is tested. Studies that clarify the intermolecular relationships of proteins related to fertilization are necessary to further our understanding of fertility in molecular terms, as well as to aid in clinical fertilization programs and provide more contraceptive avenues.

Methods

The experimental design for testing the hypothesis was based primarily on a bioinformatic approach. An initial literature search was conducted to find candidate mammalian sperm proteins. Twelve candidate proteins were found; acrosin, ADAM1, ADAM2, ADAM3, calmegin, GAPDS, PRM1, SEMG2, SPA17, SPACA3, SPAM, and zonadhesin. The DNA and amino acid sequences for these genes from various species were retrieved from GenBank (www.ncbi.nlm.nih.gov/). The sequences found were then used to search the GenBank database for similar sequences in different species of mammals using a sequence similarity search algorithm (BLAST: Basic Local Alignment Search Tool, which finds regions of similarity between biological sequences). Specifically tBLASTn was used, which searches the translated nucleotide database by using a protein query. The number of species considered for each gene varies according to available data.

The mating system for each of the species used in the study was needed, as the objective is dependent on the mating system of each species. Information on the mating system of the species was obtained initially from http://animaldiversity.ummz.umich.edu/site/index.html (Meyers *et al.*, 2006). The information that was found from this site was later confirmed from the literature (Bronson, 1979; Eisenberg, 1981; Nowak, 1999)

The DNA and amino acid sequences were aligned between species for each of the genes using the alignment algorithm CLUSTAL W (Figure 1). CLUSTAL W is a global sequence alignment algorithm (Thompson *et al.*, 1994). The DNA sequences were

aligned on the basis of the amino acid sequences, in order to maximize conservation of the sequence and to keep the reading frame the same. The alignment files were exported and converted to .pml file types by using DAMBE.

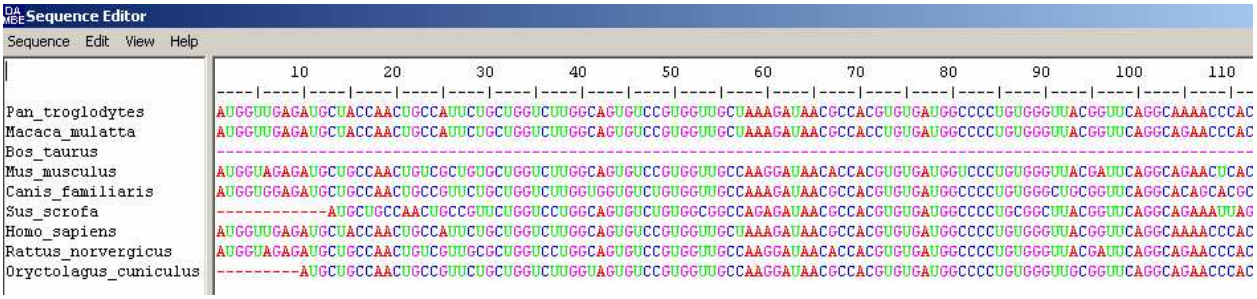


Figure 1: *Acrosin Alignment*. An alignment of acrosin in DAMBE.

The mode and rate of DNA sequence evolution was analyzed within the context of the species phylogeny using MEGA4.0 and phylogenetic analysis by maximum likelihood (PAML). Two phylogenies were reconstructed for each gene, the neighbour-joining tree and the maximum parsimony tree. Both types were done using the bootstrap model, a method that gives information about the sampling distribution by collecting its values from many resamples (Hesterberg *et al.*, 2005), and the poisson correction was applied, which corrects for multiple hits in a sequence over time. The species-tree topologies were used to test for evidence of positive selection and/or changes in evolutionary rates along tree branches at each gene using PAML analysis (Figure 2) (Springer *et al.*, 2004). This chi-square (χ^2) test was done with significance $P< 0.05$ for all models and all genes.

To test for positive selection at the protein level the number of nonsynonymous substitutions per nonsynonymous site was compared to the number of synonymous

substitutions per synonymous site (dn/ds=ω). If ω was significantly greater than one then that would indicate that positive selection was occurring, as the nonsynonymous substitution is offering an advantage in fitness for the protein (Yang *et al.*, 2000). Two types of models were used to analyze the data: codon site models (M7 and M8) and branch models (M0, M1 and M2). M7 is a model where ω values follow a beta distribution of density $f(\omega)=\omega^{p-1}(1-\omega)^{q-1}/\beta(p, q), 0\leq\omega\leq1$, but the model does not account for positively selected sites (notice ω ranges between 0 and 1). M8 (*beta&ω*) adds an extra class of sites to the beta model, where a proportion p_0 of sites have ω determined from the β distribution and the sites that remain (p_1) have ω_1 higher than 1 (positively selected codon sites) (Yang *et al.*, 2000). For the analysis of rates of evolution along branches, model 0 (M0) was used, which assumes one constant ω for all branches; M1, in which ω is freely estimated along each branch; and M2 in which a foreground branch is allowed to freely evolve and a single ω value is fixed for the background. To test for the presence of positively selected codon sites, the likelihood of the M7 and M8 models are compared using a likelihood-ratio test (Anisimova *et al*., 2001). To test for variable rates of evolution along branches in a phylogeny the likelihood of the M1 and M0 models are compared. If the free-ratio (M1) model fits the data better, then the M2 model is used to test for acceleration/deceleration of evolutionary rates along specific branches

Results

DNA and Protein Accession Numbers

After determining from the literature which proteins would be under investigation in this study a search of GenBank was conducted to find the DNA and protein accession numbers. The results of this search can be seen in Table 1. The mating system for each species was included so that those species that displayed promiscuous behaviour could be tested further. These accession numbers could be used to obtain the amino acid sequence for each gene, which could then be inputted into the MEGA4.0 program. This program, by using the alignment algorithm CLUSTAL, aligned a specific gene's sequence between species.

Table 1: *Accession Numbers.* DNA and protein accession numbers for each species and each gene found in GenBank, as well the mating behaviours for each species (www.ncbi.nlm.nih.gov/, 2007; Meyers *et al.*, 2006; Eisenberg, 1981). (Polygynandrous = promiscuous).

Gene	Order	Species	Accession # (DNA)	Accession # (protein)	Mating System
Acrosin	Rodentia	*Mus musculus*	NM_013455	NP_038483	polygynous
		Rattus norvegicus	NM_012490	NP_036622	polygynandrous
	Primates	*Homo sapiens*	NM_001097	NP_001088	variable
		Pan troglodytes	XM_001144705	XP_001144705	polygynandrous/polygynous
		Macaca mulatta	XM_001113191	XP_001113191	polygynandrous
	Artiodactyla	*Bos taurus*	NM_173886	NP_776311	polygynous
		Sus scrofa	NM_214033	NP_999198	polygynous
	Carnivora	*Canis familiaris*	XM_843155	XP_848248	monogamous/polygynandrous
ADAM1	Rodentia	*Mus musculus*	NM_172126	NP_742124	polygynous
		Rattus norvegicus	NM_020078	NP_064463	polygynandrous
	Primates	*Papio hamadryas*	Y15520	CAA75674	polygynous
		Saguinus oedipus	Y15512	CAA75666	monogamous
		Macaca fascicularis	Y09304	CAA70487	polygynandrous/polygynous
		Macaca mulatta	XM_001109414	XP_001109414	polygynandrous
		Pan troglodytes	XR_023088		polygynandrous/polygynous
		Pongo pygmaeus	Y15491	CAA75659	polygynous

	Artiodactyla	*Bos taurus*	XM_866254	XP_871347	polygynous
	Lagomorpha	*Oryctolagus cuniculus*	NM_001082701	NP_001076170	polygynandrous
	Perissodactyla	*Equus caballus*	XM_001490830	XP_001490880	polygynous
ADAM2	Rodentia	*Mus musculus*	NM_009618	NP_0033748	polygynous
		Rattus norvegicus	NM_020077	NP_064462	polygynandrous
	Primates	*Homo sapiens*	NM_001464	NP_001455	variable
		Pan troglodytes	XM_519722	XP_519722	polygynandrous/polygynous
		Macaca mulatta	XM_001095726	XP_001095726	polygynandrous
		Macaca fascicularis	X77653	CAA54733	polygynandrous/polygynous
	Artiodactyla	*Bos taurus*	NM_174228	NP_776653	polygynous
		Sus scrofa	NM_213957	NP_999122	polygynous
	Lagomorpha	*Oryctolagus cuniculus*	NM_001082677	NP_001076146	polygynandrous
	Carnivora	*Canis familiaris*	XM_532795	XP_532795	monogamous/polygynandrous
	Perissodactyla	*Equus caballus*	XM_001490821	XP_001490871	polygynous
	Monotremata	*Ornithorhyncus anatinus*	XM_001517203	XP_00157253	monogamous
ADAM3	Rodentia	*Mus musculus*	NM_009619	NP_033749	polygynous
		Rattus norvegicus	NM_020302	NP_064698	polygynandrous
	Primates	*Homo sapiens*	XR_0011679		Variable
		Macaca mulatta	NR_001569		polygynandrous
	Artiodactyla	*Bos Taurus*	NM_001098166	NP_001091635	polygynous
		Sus scrofa	NM_001098595	NP_001092065	polygynous
			NM_001097513	NP_001090982	
	Carnivora	*Canis familiaris*	XM_844512	XP_849605	monogamous/polygynandrous
Calmegin	Rodentia	*Mus musculus*	NM_009904	NP_034034	polygynous
		Rattus norvegicus	XM_001063487	XP_001063487	polygynandrous
	Primates	*Homo sapiens*	NM_004362	NP_004353	variable
		Macaca mulatta	XM_001089882	XP_001089882	polygynandrous
			XM_001089765	XP_001089765	
		Pan troglodytes	XR_023076		
	Artiodactyla	*Bos taurus*	NM_001034205	NP_001029377	polygynous
	Carnivora	*Canis familiaris*	XM_533285	XP_533285	monogamous/polygynandrous
	Monotremata	*Ornithorhyncus anatinus*	XM_001513550	XP_001513600	monogamous
GAPDS	Rodentia	*Mus musculus*	NM_008085	NP_032111	polygynous
		Rattus norvegicus	NM_023964	NP_076454	polygynandrous
	Primates	*Homo sapiens*	NM_014364	NP_055179	variable
		Pan troglodytes	XM_512590	XP_512590	polygynandrous/polygynous
			XM_001160042	XP_001160042	
		Macaca mulatta	XM_001095332	XP_00109532	polygynandrous
	Artiodactyla	*Bos taurus*	NM_001040552	NP_001035642	polygynous
	Carnivora	*Canis familiaris*	XM_533693	XP_533693	monogamous/polygynandrous
	Perissodactyla	*Equus caballus*	XM_001491988	XP_001492038	polygynous

PRM1	Rodentia	*Mus musculus*	NM_013637	XP_038665	polygynous
		Rattus norvegicus	NM_001002850	NP_001002850	polygynandrous
SEMG2	Primates	*Homo sapiens*	NM_003008	NP_002999	variable
		Pan troglodytes	NM_001009138	NP_001009138	polygynandrous/polygynous
		Gorilla gorilla	AY781387	AAV51945	polygynous
		Pongo pygmaeus	AY781388	AAV51946	polygynous
		Hylobates lar	AY781389	AAV51947	monogamous
		Macaca nemestrina	AY781391	AAV51949	polygynandrous
		Colobus guereza	AY781392	AAV51950	polygynous
		Macaca fascicularis	AY781390	AAV51948	polygynandrous/polygynous
		Macaca mulatta	XM_001109088	XP_001109088	polygynandrous
			XM_001109141	XP_001109141	
			XM_001109200	XP_001109200	
			XM_001109253	XP_001109253	
		Hylobates klossii	AY259291	AAP86629	monogamous
		Ateles geoffrayi	AY781393	AAV51951	polygynandrous
SPA17	Rodentia	*Mus musculus*	NM_011449	NP_035579	polygynous
		Rattus norvegicus	NM_053482	NP_445934	polygynandrous
	Primates	*Homo sapiens*	NM_017425	NP_059121	variable
		Papio hamadryas	U75209	AAB38534	polygynous
		Macaca mulatta	XM_001109798	XP_001109798	polygynandrous
		Macaca fascicularis	AF05551	AAB62888	polygynandrous/polygynous
		Callithrix jacchus	AF134585	AAD30113	polyandrous
		Pan troglodytes	XR_020231		polygynandrous/polygynous
	Artiodactyla	*Bos taurus*	NM_001076440	NP_001069908	polygynous
		Ovis aries	NM_001009738	NP_001009738	polygynous
	Lagomorpha	*Oryctolagus cuniculus*	NM_001082626	NP_001076095	polygynandrous
	Carnivora	*Canis familiaris*	XM_536538	XP_536538	monogamous/polygynandrous
			XM_853706	XP_858799	
	Monotremata	*Ornithorhyncus anatinus*	XM_001510449	XP_001510499	monogamous
	Didelphimorphia	*Monodelphis domestica*	NM_001032992	NP_001028164	polygynandrous
SPACA3	Rodentia	*Mus musculus*	XM_354632	XP_354632	polygynous
			NM_029367	NP_083643	
		Rattus norvegicus	XM_213416	XP_213416	polygynandrous
	Primates	*Homo sapiens*	NM_173847	NP_776246	variable
		Pan troglodytes	XM_001174564	XP_001174564	polgynandrous/polygynous
		Macaca mulatta	XM_001113227	XP_001113227	polygynandrous
			XM_001113262	XP_001113262	
	Artiodactyla	*Bos taurus*	NM_001105007	NP_001098477	polygynous
	Carnivora	*Canis familiaris*	XM_848245	XP_853338	monogamous/polygynandrous
SPAM	Rodentia	*Mus musculus*	NM_009241	NP_033267	polygynous
		Rattus norvegicus	NM_053967	NP_446419	polygynandrous

		Cavia porcellus	X56332	CAA39768	polygynandrous
	Primates	*Homo sapiens*	NM-003117	NP_003108	variable
			NM_153189	NP694859	
		Pan troglodytes	XM_527873	XP_527873	polygynandrous/polygynous
		Macaca mulatta	XM_001086882	XP_001086882	polygynandrous
	Artiodactyla	*Bos taurus*	NM_001008413	NP_0010086882	polygynous
		Sus scrofa	NM_214011	NP_999176	polygynous
	Lagomorpha	*Oryctolagus cuniculus*	NM_001082672	NP_001076141	polygynandrous
	Carnivora	*Vulpes vulpes*	U41412	AAB86588	monogamous
		Canis familiaris	XM_532443	XP_532443	monogamous/polygynandrous
	Perissodactyla	*Equus caballus*	XM_001488064	XP_001488114	polygynous
ZAN	Rodentia	*Mus musculus*	NM_011741	NP_035871	polygynous
		Rattus norvegicus	XM_222025	XP_222025	polygynandrous
	Primates	*Homo sapiens*	NM_003386	NP_003377	variable
			NM_173059	NP_775082	
		Macaca mulatta	XR_014144		polygynandrous
	Artiodactyla	*Sus scrofa*	NM_214383	NP_999548	polygynous
	Lagomorpha	*Oryctolagus cuniculus*	NM_001082081	NP_001075550	polygynandrous
	Carnivora	*Canis familiaris*	XM_546949	XP_546949	monogamous/polygynandrous
	Monotremata	*Ornithorhyncus anatinus*	XM_001519548	XP_001519598	monogamous

Acrosin PAML Results and Phylogeny

The data from the alignment and the phylogenies were run through the PAML program. For each gene several models were run, including M0, M1, M2 (branch tests), M7 and M8 (codon site tests). M2, which was only used if variable rates of evolution were occurring, was testing for sexual selection so the specific branches tested were those that led to promiscuous species, such as the primate lineage as a whole, and then more specifically *Pan troglodytes*, *Macaca mulatta*. As well, *Oryctolagus cuniculus* was also tested under M2. Comparing M7 and M8 to each other indicates positive selection if the result is significant.

The comparison of M7 and M8 shows that the acrosin gene is undergoing positive selection (Table 2). The $2\Delta l$ value is 8.66, which is significant at $P< 0.05$, therefore

providing evidence of positive selection. However, there was no evidence of significant variation in rates of evolution along lineages (M1 *vs*. M0 2Δl = 16.86; see Table 2).

Table 2: *PAML Results for Acrosin***.** PAML results for acrosin at P< 0.05 (significant values in bold) where significant values indicate positive selection is taking place. DF=degrees of freedom; np= number of parameters; L= likelihood.

Location	Model	parameters	np	L	2Δl	DF	P<0.05
Branch	M0 (one)	ω=0.31	18	-5307.77			
	M1 (free)		33	-5299.34	16.86	15	25.00
Site	NSsites =7	p= 0.38 q= 0.71	19	-5212.51			
	NSsites =8	p0= 0.96 p= 0.47 q=1.05 (p1= 0.04)	21	-5207.70	**8.66**	2	5.99
		w= 2.15					

The phylogeny of the species using the d_n values of acrosin was determined from the PAML results (Figure 3). Along each branch the d_n/d_s (ω) value is shown. The branch leading to *Pan troglodytes* and *Homo sapiens* is the only one showing an ω value higher than one. This indicates that there is an increase of the rate of evolution along this branch, possibly due to positive selection. Those branches that have an ω value of ∞ may not actually be significant as this value may be due to a d_s value of 0. Within this paper ∞ values will not be treated as significant.

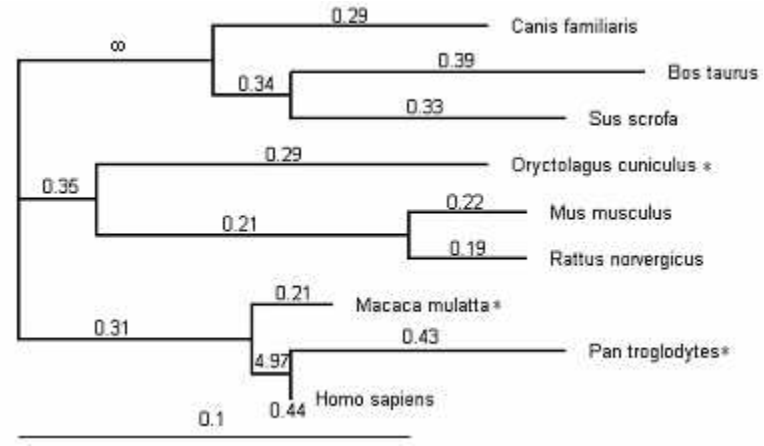


Figure 2: *Acrosin Phylogeny.* Acrosin phylogeny of species plotted using d_n values (ω values shown for each branch) (*=promiscuous species).

ADAM1 PAML Results and Phylogeny

The PAML analysis for ADAM1 reveals that this gene is undergoing positive selection (M8vs.M7) (Table 3). There is also evidence of significant variation in evolutionary rates along lineages. However, none of the branches leading to promiscuous species show significant results. The phylogeny of ADAM1 is seen in Figure 4. The branch leading to primates, rodents and lagomorphs has an ω value that is high relative to the values for all the other branches. This seems to indicate that there is an increase in the rate of evolution at this point of divergence.

ADAM2 PAML Results and Phylogeny

Evidence of positive selection can be seen based on the comparison of M7 and M8, where 2Δl= 97.06 (Table 4). There is also evidence of significant variation in ω across lineages (M1vs.M0) (Table 4). The *Macaca mulatta* lineage, as well as the

primate lineage show evidence of acceleration in ω rates relative to other lineages. This could be used to suggest that there

Table 3: *ADAM1 PAML Results.* ADAM1 PAML results at P< 0.05 (significant values in bold) where significant values indicate positive selection is taking place. DF=degrees of freedom; np= number of parameters; L= likelihood.

Location	Model	parameters	np	L	2Δl	DF	P<0.05
Branch	M0 (one)	ω=0.48	18	-12583.43			
	M1 (free)		33	-12564.89	**37.00**	15	25.00
	M2(macaca)	ω=0.48	19	-12583.11	0.64	1	3.84
	M2(orycto)	ω=0.48	19	-12583.35	0.16	1	3.84
	M2 (prmts)	ω=0.48	19	-12583.43	0	1	3.84
Site	NSsites =7	p= 0.24 q= 0.28	19	-12225.52			
	NSsites =8	p0= 0.88 p= 0.33 q= 0.51 (p1= 0.12) w= 3.17	21	-12157.63	**135.78**	2	5.99

is sexual selection along these lineages, but none of the other promiscuous lineages (i.e. *Oryctolagus, Pan*) show any such significance when M2 and M0 models were compared, indicating that the *Macaca* and primate significance may be due to some other factor other than promiscuity.

When the phylogeny, determined from the PAML results, is analysed there is only one branch that seems to indicate an increased rate of evolution due to positive selection as the ω has a value of 1.72. That branch is the one that leads to *Canis familiaris*, *Equus caballus,* and the artiodactyls. *Canis familiaris* tends to be monogamous, whereas *Equus*

caballus and the artiodactyls tend to be polygynous. As none of these species are promiscuous the reason for this increased rate of evolution would not seem to be due to sexual selection.

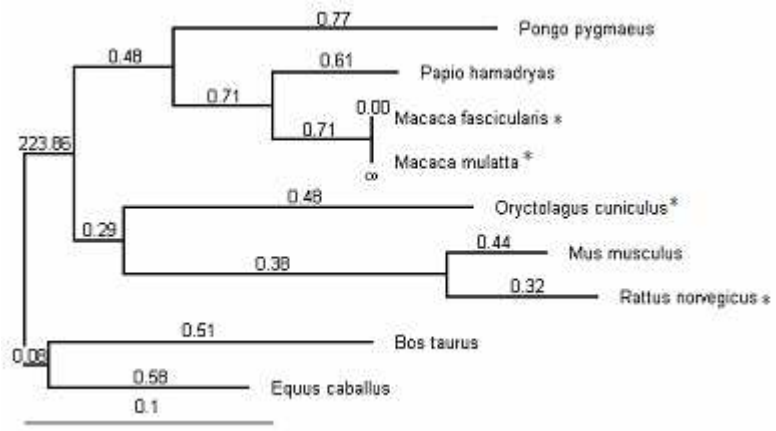


Figure 3: *ADAM1 Phylogeny*. ADAM1 phylogeny plotted using d_n values (ω values shown for each branch) (*=promiscuous species).

ADAM3 PAML Results and Phylogeny

The search in GenBank for ADAM3 accession numbers turned up only a few species, but none of those species that were promiscuous species (see Table 1). This gene shows evidence of positive selection as is seen when M7 is compared to M8 with the resulting 2Δl value of 38.64, which is significant at $P<0.05$ (Table 5). M1 is also significant, at 2Δl= 24.24. There is only one branch in the phylogeny, leading to *Bos taurus* and *Sus scrofa*, that shows evidence of an increased rate in evolution due to positive selection (ω=5.29) (Figure 6). Both of these species are polygynous, which would seem to indicate that this increased value is not due to sexual selection.

Table 4: *ADAM2 PAML Results.* ADAM2 PAML results at P< 0.05 (significant values in bold) where significant values indicate positive selection is taking place. DF=degrees of freedom; np= number of parameters; L= likelihood.

Location	Model	parameters	np	L	2Δl	DF	P<0.05
Branch	M0 (one)	ω=0.37	24	-13337.13			
	M1 (free)		45	-13281.06	**112.14**	21	32.67
	M2(pan)	ω=0.77	25	-13336.71	0.84	1	3.84
	M2(macaca)	ω=0.37	25	-13332.68	**8.90**	1	3.84
	M2(orycto)	ω=0.37	25	-13337.12	0.02	1	3.84
	M2 (prmts)	ω=0.36	25	-13334.66	**4.94**	1	3.84
Site	NSsites =7	p= 0.32 q= 0.47	25	-12830.46			
	NSsites =8	p0= 0.88 p= 0.40	27	-12781.93	**97.06**	2	5.99
		q= 0.79 (p1= 0.12)					
		w= 2.47					

Calmegin PAML Results and Phylogeny

Calmegin was an interesting case as, the sequence for *Pan troglodytes* had several open reading frames (ORFs). Two were quite large compared to the rest and were used separately for analysis, with each run using a different ORF. To distinguish the results, calmeginA or calmeginB were randomly designated to one ORF each.

Neither calmeginA nor calmeginB show significant results when M7 is compared to M8, which seems to indicate that positive selection is not occurring (Table 6 and Table 7). Neither calmeginA nor calmeginB shows significance under any model, or along any branch.

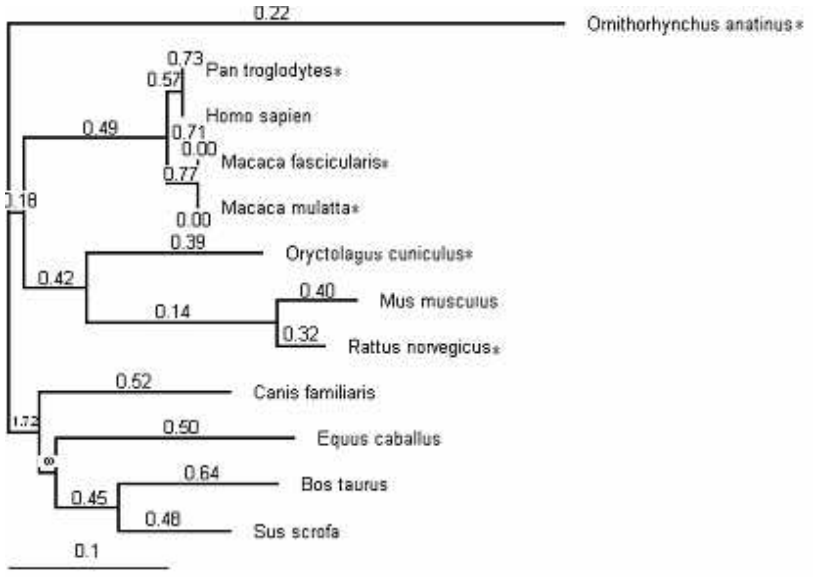


Figure 4: *ADAM2 Phylogeny.* ADAM2 phylogeny of species plotted using d_n values (ω values shown for each branch) (*=promiscuous species).

Table 5: *ADAM3 PAML Results.* ADAM3 PAML results at P< 0.05 (significant values in bold) where significant values indicate positive selection is taking place. DF=degrees of freedom; np= number of parameters; L= likelihood.

Location	Model	parameters	np	L	2Δl	DF	P<0.05
Branch	M0 (one)	ω=0.42	9	-8249.94			
	M1 (free)		15	-8237.82	**24.24**	6	12.59
Site	NSsites =7	p= 0.32 q= 0.37	10	-8117.52			
	NSsites =8	p0= 0.86 p= 0.61	12	-8098.20	**38.64**	2	5.99
		q= 1.15 (p1= 0.14)					
		w= 2.37					

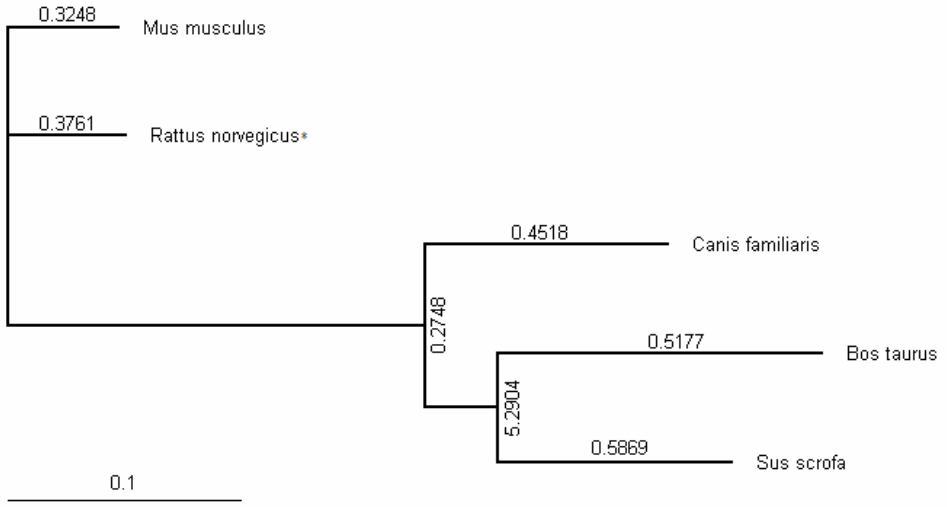


Figure 5: *ADAM3 Phylogeny*. ADAM3 phylogeny of species plotted using d_n values (ω values shown for each branch) (*=promiscuous species).

Table 6: *CalmeginA PAML Results*. CalmeginA PAML results at P< 0.05 (significant values in bold) where significant values indicate positive selection is taking place. DF=degrees of freedom; np= number of parameters; L= likelihood.

Location	Model	parameters	np	L	2Δl	DF	P<0.05
Branch	M0 (one)	ω= 0.12	15	-2779.74			
	M1 (free)		27	-2775.58	8.32	12	21.03
Site	NSsites =7	p= 0.34 q= 1.96	16	-2754.37			
	NSsites =8	p0= 1 p= 0.37	18	-2754.37	0	2	5.99
		q= 1.96 (p1= 0)					
		w= 3.03					

The phylogenies for both calmeginA and calmeginB show similar results, as neither indicates branches that are undergoing increased rates of evolution. There are a couple that have ω values of ∞, but as mentioned previously, these high rates are a convergence of d_s=0 values.

GAPDS PAML Results and Phylogeny

GAPDS is interesting because it does not show positive selection when M7 is compared to M8 in this study but it does in previous studies (Table 8) (Torgerson *et al.*, 2002*)*. Figure 9 shows the phylogeny for GAPDS. Two of the branches have ω values that are relatively high. The branch leading to primates, rodents and lagomorphs has an ω value of 73.74, and the branch leading to *Bos taurus* and *Equus caballus* had an ω value of 1.03.

Table 7: *CalmeginB PAML Results*. CalmeginB PAML results at P< 0.05 (significant values in bold) where significant values indicate positive selection is taking place. DF=degrees of freedom; np= number of parameters; L= likelihood.

Location	Model	parameters	np	L	2Δl	DF	P<0.05
Branch	M0 (one)	ω=0.14	16	-2811.30			
	M1 (free)		29	-2796.83	28.94	13	22.36
Site	NSsites =7	p= 0.35 q= 1.66	17	-2771.06			
	NSsites =8	p0= 0.86 p= 2.05	19	-2769.02	4.08	2	5.99
		q= 24.47 (p1= 0.15)					
		w= 1.0					

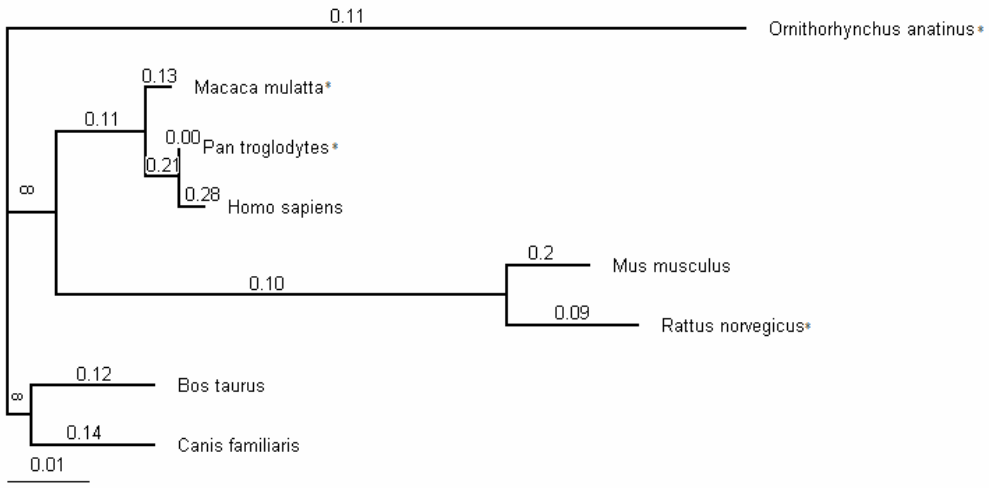


Figure 6: *CalmeginA Phylogeny*. CalmeginA phylogeny of species plotted using d_n values (ω values shown for each branch) (*=promiscuous species).

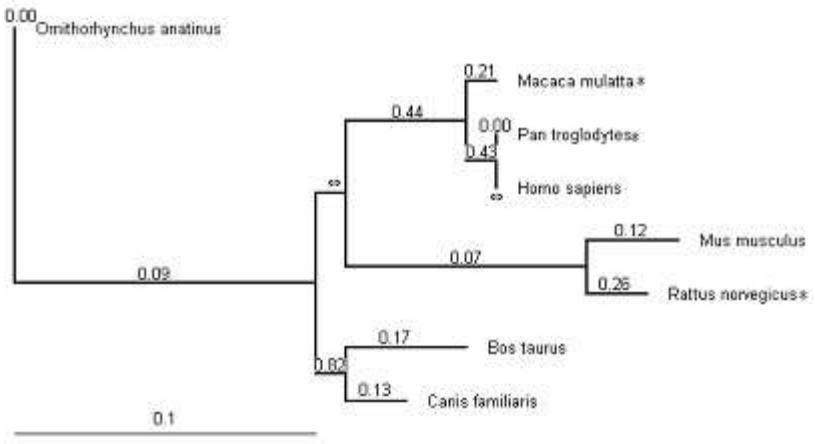


Figure 7: *CalmeginB Phylogeny*. CalmeginB phylogeny of species plotted using d_n values (ω values shown for each branch).

Table 8: *PAML Results for GAPDS.* PAML results for GAPDS at P< 0.05 (significant values in bold) where significant values indicate positive selection is taking place. DF=degrees of freedom; np= number of parameters; L= likelihood.

Location	Model	parameters	np	L	2Δl	DF	P<0.05
Branch	M0 (one)	ω=0.12	18	-4099.77			
	M1 (free)		33	-4089.75	20.04	15	25.00
Site	NSsites =7	p= 0.19 q= 1.03	19	-3992.94			
	NSsites =8	p0= 0.97 p= 0.23 q= 1.71 (p1= 0.03) w= 1.94	21	-3990.74	4.4	2	5.99

SEMG2 PAML Results and Phylogeny

SEMG2 is interesting because the accession numbers that were obtained for it were almost solely comprised from primates. The only other order from which accession numbers were obtained was from rodents, the closest relative to primates in this study. When M7 is compared to M8 the value for 2Δl is significant at 15.98 (Table 9). There is also significance under M1. However, none of the branches, including the primate branch show significance under M2. This is interesting as the phylogeny shows a couple of branches where ω is relatively high, indicating that these branches, which lead to *Ateles geoffrayi*, and the great apes (*Homo sapiens, Pan troglodytes,* and *Gorilla gorilla)*, are undergoing increased rates of evolution but not due to sexual selection and mating behaviour (Figure 10).

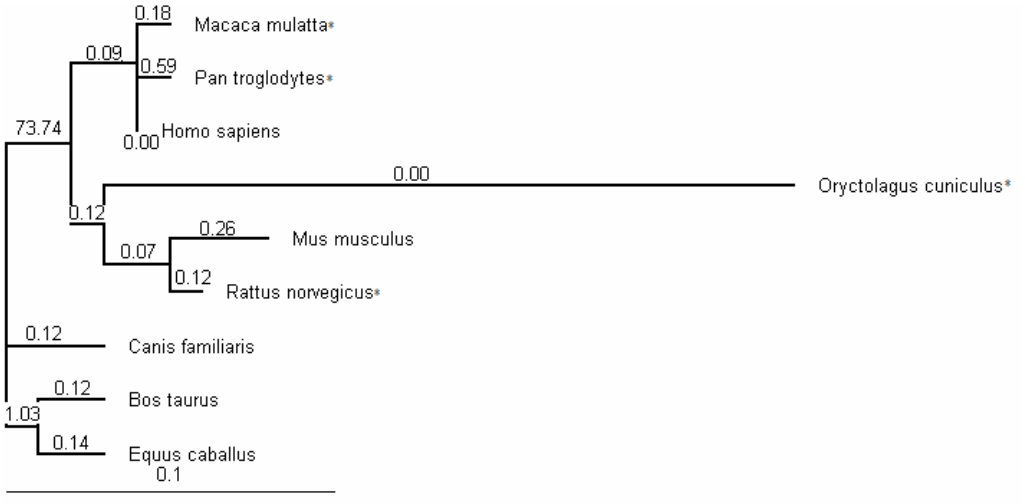


Figure 8: *GAPDS Phylogeny*. GAPDS phylogeny of species plotted using d_n values (ω values shown for each branch) (*=promiscuous species).

Table 9: *SEMG2 PAML Results*. SEMG2 PAML results at P< 0.05 (significant values in bold) where significant values indicate positive selection is taking place. DF=degrees of freedom; np= number of parameters; L= likelihood.

Location	Model	parameters	np	L	2Δl	DF	P<0.05
Branch	M0 (one)	ω=0.75	25	-3582.95			
	M1 (free)		47	-3561.51	**42.88**	22	33.92
	M2(pan)	ω=1.10	26	-3582.89	0.12	1	3.84
	M2(macaca)	ω=0.74	26	-3581.87	2.16	1	3.84
	M2 (prmts)	ω=0.69	26	-3581.66	2.58	1	3.84
Site	NSsites =7	p= 0.68 q= 0.29	26	-3566.54			
	NSsites =8	p0=0.91 p= 0.80	28	-3558.55	**15.98**	2	5.99
		q= 0.40 (p1= 0.09)					
		w= 3.94					

SPACA3 PAML Results and Phylogeny

The PAML results for SPACA3 show no significant 2Δl results (Table 10). This indicates that neither positive selection, nor sexual selection as evidenced by increased rates of evolution along promiscuous branches is occurring. In fact, based on M1vsM0, M0 is a better fit for the data, which indicates that there are constant rates of evolution occurring. The phylogeny is interesting then because the branch leading to *Pan troglodytes* and *Homo sapiens* has an ω value of 1.03 which is high, relative to the rest of the branches. This indicates that this branch is undergoing an increased rate of evolution (Figure 11).

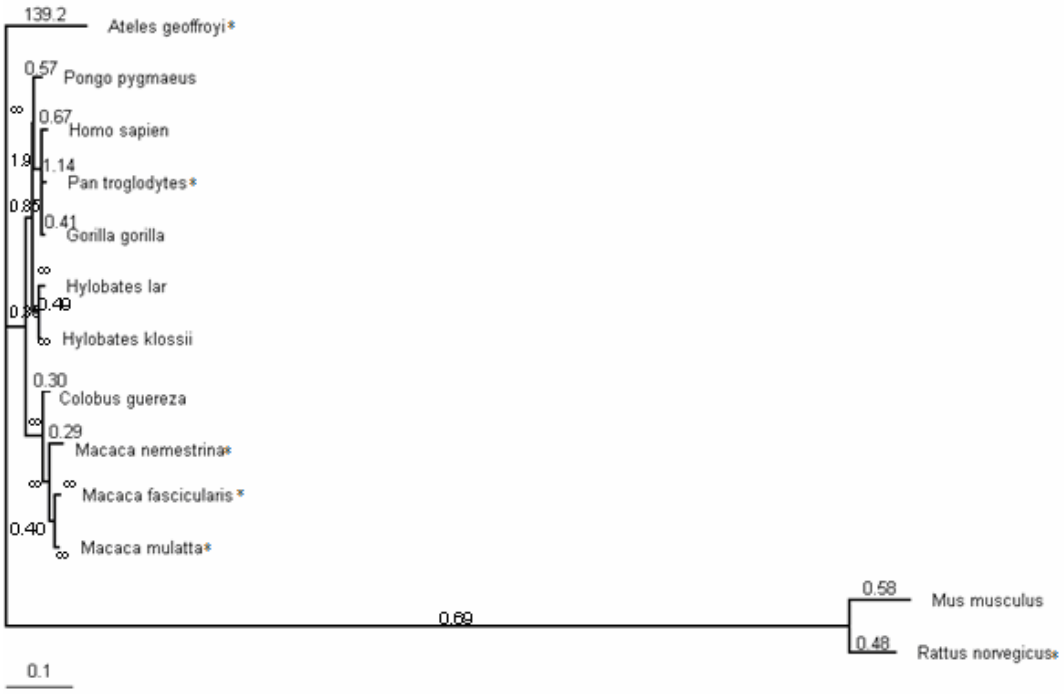


Figure 9: *SEMG2 Phylogeny*. SEMG2 phylogeny of species plotted using d_n values (ω values shown for each branch) (*=promiscuous species).

SPAM PAML Results and Phylogeny

In Table 11 the PAML results for SPAM show that this gene is undergoing positive selection when M7 is compared to M8, as 2Δl=64.92. However, none of the branch models show significant 2Δl values. Again, this is interesting when taken into consideration with the phylogeny which does show high ω values along some branches, indicating that variable and increased rates of evolution are occurring (Figure 12). The branch leading to the primates, rodents and

Table 10: *SPACA3 PAML Results*. SPACA3 PAML results at P< 0.05 (significant values in bold) where significant values indicate positive selection is taking place. DF=degrees of freedom; np= number of parameters; L= likelihood.

Location	Model	parameters	np	L	2Δl	DF	P<0.05
Branch	M0 (one)	ω=0.20	14	-1257.45			
	M1 (free)		25	-1250.07	14.76	11	19.68
Site	NSsites =7	p= 0.52 q= 1.54	15	-1246.80			
	NSsites =8	p0= 1.00 p= 0.52	17	-1246.80	0	2	5.99
		q= 1.54 (p1= 0.00)					
		w= 1.00					

lagomorphs has an ω value of 303.91. The branch leading to *Pan troglodytes* and *Homo sapiens* also has a high ω value, where ω=1.36. As well, the branch leading to *Vulpes vulpes* has a relatively high ω value of 1.36.

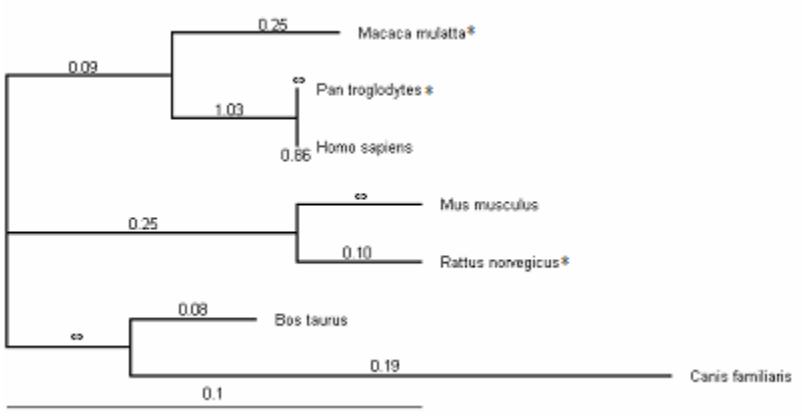


Figure 10: *SPACA3 Phylogeny*. SPACA3 phylogeny of species plotted using d_n values (ω values shown for each branch) (*=promiscuous species).

Table 11: *PAML Results for SPAM*. PAML results for SPAM at P< 0.05 (significant values in bold) where significant values indicate positive selection is taking place. DF=degrees of freedom; np= number of parameters; L= likelihood.

Location	Model	parameters	np	L	2Δl	DF	P<0.05
Branch	M0 (one)	ω=0.57	24	-10095.25			
	M1 (free)		45	-10084.83	20.84	21	32.67
Site	NSsites =7	p= 0.29 q= 0.26	25	-9835.51			
	NSsites =8	p0= 0.87 p= 0.34 q= 0.38 (p1= 0.13) w= 2.34	27	-9803.05	**64.92**	2	5.99

Overall Summary of Results

An overall summary of the findings in PAML can be seen in Table 12. The majority of the genes showed evidence of positive selection. The exceptions are calmegin, GAPDS, and SPACA3. Four genes show evidence of variable rates of

evolution across lineages but for the most part the genes did not show evidence of sexual selection along the promiscuous branches, except for the *Macaca* and primate lineages in ADAM2 (Table 12).

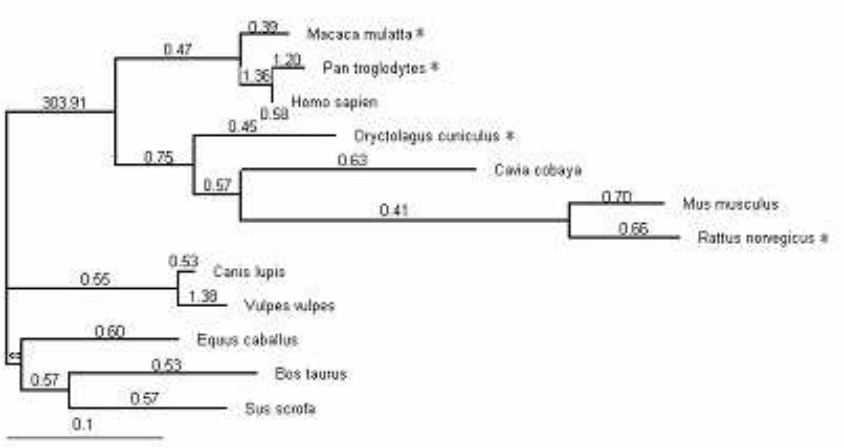


Figure 11: *SPAM Phylogeny*. SPAM phylogeny of species plotted using d_n values (ω values shown for each branch) (*=promiscuous species).

There were a couple of problems with some of the genes that caused them to be excluded from the analysis. PRM1 was excluded as the only available sequence data were for *Mus musculus* and *Rattus norvegicus* at the time the GenBank search was conducted. PAML requires a species tree topology and therefore at least three species. Some of the sequences found for SPA17 had annotation problems. As stop codons appeared in the middle of some sequences causing the open reading fram (ORF) to be disrupted, SPA17 was excluded from the analysis. Zonadhesin (ZAN) was also excluded as the sequences for some of the species had several gaps throughout.

Table 12: *Summary of PAML Results.* PAML results at P< 0.05. X indicates if the PAML results revealed positive selection, variable rates of evolution and if selection (i.e. sexual selection) occurred along any tested branches.

Gene	Positive Selection	Variable Rates	Sexual selection
Acrosin	X		
ADAM1	X	X	
ADAM2	X	X	X
ADAM3	X		
CalmeginA			
CalmeginB			
GAPDS			
SEMG2	X	X	
SPACA3			
SPAM	X		

Discussion

PAML Results and Positive Selection

Alleles of a gene that confer an advantage in terms of fitness are expected to increase in frequency due to positive selection, while sexual selection should increase the frequency of alleles that provide one individual's gametes an advantage in fertilizing or being fertilized over another's gametes. In this study most of the candidate proteins were chosen based on previous studies that indicated that these genes had undergone positive selection (Torgerson *et al.*, 2002, Civetta, 2003b, Swanson *et al.*, 2003, Dorus *et al.*, 2004, and Glassey & Civetta, 2004). However, there were three genes when analysed through PAML that did not show positive selection: calmegin, GAPDS, and SPACA3. Calmegin may not have shown positive selection due to the fact that through the course of gathering data it was found to have several ORFs for *Pan troglodytes*. As this gene was run twice, using each of the two longest ORFs for the trials, it may be that nonsynonymous changes may be localized in the sections of the gene that were not included. However, as this gene had not been previously studied it may be that it is not undergoing positive selection. GAPDS and SPACA3 were previously analysed using PAML to reveal evidence of positive selection. The contradictory results of this study could be due to a larger number of species being used in this study compared to the low numbers that were used in the original studies (Torgerson *et al.*, 2002; Swanson *et al.*, 2003).

The effect of sexual selection at the molecular level can manifest itself by either specific signals of selection at codon sites (M8 vs. M7 model comparisons) and/or by

more subtle acceleration/deceleration of evolutionary rates along lineages. Calmegin, GAPDS and SPACA3 did not show positive selection at specific codon sites suggesting that selection, if operating, might be variable across species so that the methods do not pick specific sites. The phylogenies of GAPDS and SPACA3 both show branches that have high ω values. There is obviously something along these branches for the specific gene that is undergoing change in rate of evolution but it is probably not due to promiscuity, and ultimately not sexual selection.

The PAML analysis revealed that most of the genes were undergoing positive selection. There were some branches in a couple of the genes that showed increased rates of evolution, but for the most part those branches that led to promiscuous species did not show overall an increased rate of evolution. This would indicate that for those branches that did show an increase, it was not due to promiscuity and sexual selection. When analysing the phylogenies for each gene that were created using the results from PAML there were several branches throughout that have high ω values. However, there are no trends based on mating behaviours that can be readily seen along those branches that have high ω values. The original hypothesis then, which stated that the mating behaviour of a species would have an affect on the rate of evolution, has not been supported through the course of the research conducted in this study.

Future Avenues

It was automatically assumed previously that because these genes are associated with reproduction and are undergoing positive selection then they must also be undergoing sexual selection. This is not necessarily true, even if the genes are involved in

reproduction. Even if, as it seems that promiscuous or polyandrous mating behaviours does not affect the rate of evolution of these proteins, there still may be other factors that do. Through the analysis there were branches for several proteins that did show an increase in rate of evolution. Future avenues of research may be to determine whether population size, litter size, breeding season, or possibly differences in parental investment are factors that affect the rate of evolution. In this study, the branches that were chosen to be further examined for varying rates of evolution were *Pan troglodytes*, *Macaca mulatta*, the primate lineage, and *Oryctolagus cuniculus*. However, the mating behaviours of primates vary among species so they cannot be generalized as whole to be promiscuous. If the primate branch is excluded and the remaining three species are considered there is significant variation between species in population size, litter size, breeding season, and parental investment that may affect rates of evolution. *Oryctolagus* has a large population size, which could in turn be due to its large litter size, frequent breeding season, and moderate parental investment (Nowak, 1999). *Pan* and *Macaca*, both primates, tend to have smaller population sizes, smaller litter sizes, a shorter and less frequent breeding season, and greater parental investment. Although it would be difficult to quantify these factors into neat and clearly defined groups, it could be possible that these are factors that would affect the evolution of these sperm proteins and other genes involved in reproduction.

Even though the hypothesis has not been supported in this study, this does not mean that sexual selection is not occurring. Sexual selection can occur through sperm competition and female cryptic choice. It may be that simply testing for sexual selection might have compounded and masked the results. It would be difficult to test male and

female components of sexual selection separately. Also, it may have been an over-simplification to say that sexual selection through the form of sperm competition and female cryptic choice is not occurring in monogamous or polygynous species. It may be that those species that are considered monogamous or polygynous may not actually be truly monogamous or polygynous, but that "cheating" may occur (Palombit, 1994). Further research needs to be done to determine if sexual selection as defined in this thesis is occurring in non-promiscuous groups. It is also possible that the hypothesis was not supported because of the species that were tested. For the most part, each gene was tested using species that varied from being very closely related to being very distantly related. The choice of species was restricted by what was available in GenBank. For future studies the hypothesis could be retested using only a single lineage, such as the primate lineage. Having the species be so distantly related could compound and mask the results, because the species are now so diverse that the diversity observed in the genes cannot be attributed solely to mating behaviours and sexual selection, even if in fact, these are contributing factors. It is also possible that the choice of genes affected the outcome. The specific roles of these sperm proteins may have skewed the results. For the most part the candidate sperm proteins were sperm-surface proteins that were involved in the acrosome reaction and sperm-egg binding, with the exception of SEMG2 (Lin *et al.*, 1993; Ikawa *et al.*, 2001; Lea *et al.*, 2001; Wen *et al.*, 2001; Primakoff & Myles, 2002; Jensen-Seaman & Li, 2003; Martins *et al.*, 2004; Miki *et al.*, 2004; Nishimura *et al.*, 2004; Zhang *et al.*, 2005; Gatesy and Swanson, 2007).

Another aspect to consider would be the evolution of promiscuity and/or polyandry. It may be possible that these species did not show increased rates of evolution

along branches that are considered promiscuous/polyandrous because the individual species may have evolved from an ancestor that was promiscuous/polyandrous. If promiscuity/polyandry is no longer evolving along a branch then it may be that the genes are no longer evolving due to promiscuity/polyandry. Future studies may need to be conducted that take into consideration where promiscuity/polyandry has evolved and if it is still evolving along specific branches.

Conclusions

Through the course of this study several findings were revealed:

1. Most of the candidate proteins were found to undergo positive selection, in concordance with previous research.
2. Evidence of significant differences in rates of evolution along lineages was found.
3. However, none of this evidence of accelerated evolution among the genes seems to be linked to sexual promiscuity, and therefore the hypothesis has not been supported.

References

Andersson, M. and Simmons, L.W. 2006. Sexual selection and mate choice. *TRENDS in Ecology and Evolution* **21**: 296-302.

Anisimova, M., Bielawski, J.P., and Yang, Z. Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. *Molecular Biology and Evolution* **18**: 1585-1592.

Bronson, F.H. 1979. The reproductive ecology of the house mouse. *The Quarterly Review of Biology* **54**: 265-292.

Civetta, A. 2003a. 2003. Shall we dance or shall we fight? Using DNA sequence data to untangle controversies surrounding sexual selection. *Genome* **46**: 925-929.

Civetta, A. 2003b. Positive selection within sperm-egg adhesion domains of fertilin: an ADAM gene with a potential role in fertilization. *Molecular Biology and Evolution* **20**: 21-29.

Civetta, A., and Singh, R.S. 1999. Broad-sense sexual selection, sex gene pool evolution, and speciation. *Genome* **42**: 1033-1041.

Clark, A.G., Begun, D.J., and Prout, T. 1999. Female x male interactions in drosophila sperm competition. *Science* **283**: 217-220.

Clark, A.G., and Begun, D.J. 1998. Female genotypes affect sperm displacement in drosophila. *Genomics* **149**: 1487-1493.

Dorus, S., Evans, P.D., Wyckoff, G.J., Choi, S.S., and Lahn, B.T. 2004. Rate of molecular evolution of the seminal protein gene SEMG2 correlates with levels of female promiscuity. *Nature Genetics* **36**: 1326-1329.

Eisenberg, J.F. 1981. The Mammalian Radiations: An analysis if trends in evolution, adaptation, and behaviour. *The University of Chicago Press*, Chicago.

Fiumera, A.C., DuMont, B.L., and Clark, A.G. 2007. Associations between sperm competition and natural variation in male reproductive genes on the third chromosome of *Drosophila melanogaster*. *Genetics* **176**: 1245-1260.

Fiumera, A.C., DuMont, B.L., and Clark, A.G. 2005. Associations between variation in second chromosome male reproductive proteins and sperm competitive ability in *Drosophila melanogaster*. *Genetics* **169**: 243-257.

Gage, M.J., Parker, G.A., Nylin, S., and Wiklund, C. 2002. Sexual selection and speciation in mammals, butterflies and spiders. *Proceedings of the Royal Society B* **269**: 2309-2316.

Gatesy, J., and Swanson, W.J. 2007. Adaptive evolution and phylogenetic utility of *Acr* (acrosin), a rapidly evolving mammalian fertilization gene. *Journal of Mammalogy* **88**: 32-42.

Glassey, B., and Civetta, A. 2004. Positive selection at reproductive ADAM genes with potential intercellular binding activity. *Molecular Biology and Evolution* **21**: 851-859.

Hesterberg, T.C., Monoghan, M.S., Clipson, A., and Epstein, R. 2005. Bootstrap Methods and Permutation Tests, 2nd ed. *W.H. Freeman,* New York.

Ikawa, M., Nakanishi, T., Yamada, S., Wada, I., Kominami, K., Tanaka, H., Nozaki, M., Nishimune, Y., and Okabe, M. 2001. Calmegin is required for fertilin α/β heterodimerization and sperm fertility. *Developmental Biology* **240**: 254-261.

Jensen-Seaman, M.I., and Li, W-H. 2003. Evolution of the hominoid semenogelin genes, the major proteins of ejaculated semen. *Journal of Molecular Evolution* **57**: 261-270.

Kim, T., Oh, J., Woo, J-M., Choi, E., Im, S.H., Yoo, Y.J., Kim, D.H., Nishimura, H., and Cho, C. 2006. Expression and relationship of male reproductive ADAMs in mouse. *Biology of Reproduction* **74**: 744-750.

Lea, I.A., Sivashanmugam, P., and O'Rand, M.G. 2001. Zonadhesin: characterization, localization, and zona pellucida binding. *Biology of Reproduction* **65**: 1691-1700.

Lin, Y., Kimmel, L.H., Myles, D.G., and Primakoff, P. 1993. Molecular cloning of the human and monkey sperm surface protein PH-20. *Proceedings of the National Academy of Sciences of the United States of America* **90**: 10071-10075.

Martins, R.P., Ostermeier, G.C., and Krawetz, S.A. 2004. Nuclear matrix interactions at the human protamine domain. *The Journal of Biological Chemistry* **279**: 51862-51868.

Meyers, P., Espinosa, R., Parr, C.S., Jones, T., Hammond, G.S. and Dewey, T.A. 2006. The Animal Diversity Web (online). Accessed October 13, 2007 at http://animaldiversity.org.

Miki, K., Qu, W., Goulding, E.H., Willis, W.D., Bunch, D.O., Strader, L.F., Perreault, S.D., Eddy, E.M., O'Brien, D.A., and Yanaimachi, R. 2004. Glyceraldehyde-3-phosphate dehydrogenase-S, a sperm-specific glycolytic enzyme, is required for sperm motility and male fertility. *Proceedings of the National Academy of Sciences of the United States of America* **101**: 16501-16506.

Nishimura, H., Kim, E., Nakanishi, T., and Baba, T. 2004. Possible function of the ADAM1a/ADAM2 fertilin complex in the appearance of ADAM3 on the sperm surface. *The Journal of Biological Chemistry* **279**: 34957-34962.

Nowak, R.M. 1999. Walker's Mammal of the World, 6th ed. *The John Hopkins University Press*, London.

O'Donald, Peter. 1980. Genetic Models of Sexual Selection. *Cambridge University Press*, Cambridge. pp. 1.

Palombit, R.A. 1994. Extra-pair copulations in a monogamous ape. *Animal Behaviour* **47**: 721-723.

Primakoff, P., and Myles, D.G. 2007. Cell-cell membrane fusion during mammalian fertilization. *Federation of European Biochemical Societies Letters* **581**: 2174-2180.

Primakoff, P., and Myles, D.G. 2002. Penetration, adhesion, and fusion in mammalian sperm-egg interaction. *Science* **296**: 2183-2185.

Springer, M.S., Stanhope, M.J., Madsen, O., and de Jong, W.W. 2004. Molecules consolidate the placental mammal tree. *TRENDS in Ecology and Evolution* **19**: 430-438.

Stockley, P. 2004. Sperm competition in mammals. *Human Fertility* **7**: 91-97.

Swanson, W.J., Nielson, R., and Yang, Q. 2003. Pervasive Adaptive Evolution in Mammalian Fertilization Proteins. *Molecular Biology and Evolution* **20**: 18-20.

Thompson, J.D., Higgins, D.G., and Gibson, T.J. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting,

position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* **22**: 4673-4680.

Torgerson, D.G., Kulathinal, R.J., and Singh, R.S. 2002. Mammalian sperm proteins are rapidly evolving: Evidence of positive selection in functionally diverse genes. *Molecular Biology and Evolution* **19**: 1973-1980.

Vacquier, V.D. 1998. Evolution of gamete recognition proteins. *Evolution of Sex* **281**: 1995-1998.

Van Doorn, G.S., Luttikhuizen, P.C., and Weissing, F.J. 2001. Sexual selection at the protein level drives the extraordinary divergence of sex-related genes during sympatric speciation. *Proceedings of the Royal Society B* **268**: 2155-2161.

Wassarman, P.M. 1987. The biology and chemistry of fertilization. *Science* **235**: 553-260.

Wen, Y., Richardson, R.T., Widgren, E.E., and O'Rand, M.G. 2001. Characterization of Sp17: a ubiquitous three domain protein that binds heparin. *Biochemical Journal* **357**: 25-31.

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

Zhang, K., Gao, R., Zhang, H., Cai, X., Shen, C., Wu, C., Zhao, S. And Yu, L. 2005. Molecular cloning and characterization of three novel lysozyme-like genes, predominantly expressed in the male reproductive system of humans, belonging to the c-type lysozyme/alpha-lactalbumin family. *Biology of Reproduction* **73**: 1064-1071.

2007. National Center of Biotechnology Information (online). Accessed September-October 2007 at www.ncbi.nlm.nih.gov/