Pervasive Adaptive Evolution in Mammalian Fertilization Proteins

Willie J. Swanson,† Rasmus Nielsen,‡ and Qiaofeng Yang†*

*Department of Genome Sciences, University of Washington; †Departments of Biology and Genetics, University of California, Riverside; and ‡Department of Biological Statistics and Computational Biology, Cornell University

Mammalian fertilization exhibits species specificity, and the proteins mediating sperm-egg interactions evolve rapidly between species. In this study, we demonstrate that the evolution of seven genes involved in mammalian fertilization is promoted by positive Darwinian selection by using likelihood ratio tests (LRTs). Several of these proteins are sperm proteins that have been implicated in binding the mammalian egg coat zona pellucida glycoproteins, which were shown previously to be subjected to positive selection. Taken together, these represent the major candidates involved in mammalian fertilization, indicating positive selection is pervasive amongst mammalian reproductive proteins. A new LRT is implemented to determine if the d_N/d_S ratio is significantly greater than one. This is a more refined test of positive selection than the previous LRTs which only identified if there was a class of sites with a d_N/d_S ratio >1 but did not test if that ratio was significantly greater than one.

Introduction

Components of the mammalian egg zona pellucida are subjected to positive Darwinian selection, indicating an adaptive value to diversify the primary amino acid sequence (Swanson et al. 2001). The zona pellucida is an elevated glycoproteineous envelope that surrounds the egg and is involved in the initial stages of sperm-egg interaction (Wassarman 1999). The sperm proteins binding the zona pellucida have been studied in great detail; however, no consensus yet exists as to their binding partners (Wassarman 1999). The observation that the egg coat proteins are subject to positive selection raised the question as to whether the sperm components also are subject to positive selection. The idea is that if the egg proteins evolve at a high rate, then the sperm receptors would have to evolve quickly through compensatory changes to maintain interaction. The selective pressure driving this rapid evolution remains unknown, but it could be the result of a conflict in the reproductive interests between males and females (sexual conflict), sperm competition, or sexual selection (Swanson and Vacquier 2002). In this study, we demonstrate that positive selection is pervasive among reproductive proteins involved in mammalian fertilization. These results suggest that understanding the evolution of fertilization proteins is necessary to determine the molecular basis of fertilization.

We compiled gene alignments from GenBank for several male and female mammalian reproductive proteins (table 1). We analyzed Zonadhesin (*Zen*, Hardy and Garbers 1995), PH20 (*SPAM1*, Lin et al. 1993), Fertilin β (*Adam2*, Zhu, Bansal, and Evans 2000), Fertilin α (*Adam1*, Wong et al. 2001), CD9 (Miyado et al. 2000), Acrosin (*Acr*, Baba et al. 1994), Sperm protein 17 (*SP17*, Richardson, Yamasaki, and O'rand 1994), and β-galactosyltransferase (*gt*, Miller, Macek, and Shur 1992). Some of these genes were extremely divergent, and we deleted regions that did not align reliably. The number of sequences we analyzed ranged from 5 to 10 per gene. All alignments are available from the authors on request, and the protein alignments used are available on the *Molecular Biology and Evolution* journal website as supplementary material (http://www.molbiolevol.org).

We tested for positive selection by comparing the number of nonsynonymous substitutions per nonsynonymous site (d_N; amino acid replacement changes) to the number of synonymous substitutions per synonymous site (d_S; silent changes). Because these numbers are normalized for the number of sites, if selection were neutral (i.e., a pseudogene), $\omega = d_N/d_S$ would equal one. A convincing sign of positive selection is when the ω ratio significantly exceeds one (e.g., Yang and Bielawski 2000), indicating a functional benefit to diversify the amino acid sequence.

Materials and Methods

We used likelihood ratio tests (LRTs) to determine if any codon positions were associated with ω significantly >1 and hence possibly subjected to positive Darwinian selection (Nielsen and Yang 1998; Yang et al. 2000). The power of these tests increases with increased sequence diversity and number. Simulation studies show that the tests are robust when the tree length is approximately one substitution per codon. All data sets analyzed had tree length greater than one substitution per codon (table 1). However, it should be noted that the low number of species may reduce the power and accuracy of these analyses (Anisimova, Bielawski, and Yang 2001, 2002). A neutral model (M7) with ω assumed to be beta-distributed was compared with a selection model (M8) with two additional parameters: p_s, the proportion of codons with $d_N/d_S > 1$, and ω_s, the value of ω in these sites. Positive selection is inferred if the estimate of ω_s is larger than one if an LRT is significant. The LRT is performed by taking the negative of twice the log-likelihood difference between the nested models (M7 and M8) and comparing this to the χ^2 distribution with degrees of freedom equal to the difference in the number of parameters between the models (test I; table 1). For the M7 versus M8 comparison, there are two degrees of freedom. However, as noted in Yang et al. (2000) under the null hypothesis, one of the parameters is on the boundary of the parameter space and

Key words: positive Darwinian selection, sexual conflict, fertilization, likelihood ratio test, speciation.

E-mail: wswanson@gs.washington.edu.

Mol. Biol. Evol. 20(1):18–20. 2003

Table 1
Adaptive Evolution in Mammalian Reproductive Proteins

Protein, Sex in Which it Functions	Putative, Function	*N*	*S*	d_N/d_S	Selection Parameters	*P* Value Test I	*P* Value Test II
PH20, Male	ZP, binding	6	4.1	0.77	$p_s = 0.09$, $\omega_S = 2.8$	**<0.001**	**<0.001**
Fertilin β, male	Sperm-egg, adhesion	6	3.1	0.40	$p_s = 0.10$, $\omega_S = 3.1$	**<0.001**	**<0.001**
Fertilin α, Male	Sperm-egg, adhesion	6	3.4	0.55	$p_s = 0.03$, $\omega_S = 3.9$	**0.001**	**0.001**
CD9, female	Sperm-egg, fusion	5	1.4	0.24	$p_s = 0.07$, $\omega_S = 2.5$	**0.01**	**0.01**
Zonadhesin, male	ZP, adhesion	5	2.7	0.26	$p_s = 0.02$, $\omega_S = 3.6$	**<0.001**	**0.01**
Acrosin, male	Acrosomal, content dispersion	10	2.8	0.28	$p_s = 0.17$, $\omega_S = 1.3$	**0.005**	0.29
SP17, male	ZP, binding	10	1.8	0.27	$p_s = 0.08$, $\omega_S = 3.54$	**<0.001**	**<0.001**
β-Galactosyltransferase, Male	ZP, binding	5	1.3	0.21	$p_s = 0.00$, $\omega_S =$ NA	0.84	1.0

N = number of taxa; *S* = tree length in substitutions per codon; d_N/d_S = ratio averaged across all sites and lineages; p_s = proportion of sites estimated to be under positive selection with $\omega_S > 1$; NA = not applicable. All significant *P* values are in bold.

another parameter is not estimable. The use of two degrees of freedom is therefore an approximation that results in a conservative test. Another problem is that test I may result in a high proportion of significant tests even when there is no positive selection if the beta-distribution provides a poor fit to the true distribution of ω in the interval (0, 1). For example, if much of the probability mass is located around $\omega = 0.5$ and $\omega = 1.0$, M8 may provide a significantly better fit to data than M7, with an estimate of $\omega_s > 1$ with probability 0.5, although no positive selection occurs.

We, therefore, implemented a new version of the LRT which is robust to the assumptions regarding the distribution of ω in (0, 1). It has the additional advantage that the asymptotic distribution of the test statistic follows from standard theory in contrast to test I. The test is performed by adding a category of sites with $\omega_s = 1$ to the null model. The new modified null model M8A then specifies that the distribution of ω follows a mixture between a beta-distribution and a point mass at $\omega = 1$. Model M8A is then compared with a version of model M8, constrained such that $\omega_s \geq 1$, using an LRT. The only difference between the models is that under the null model (M8A) $\omega_s = 1$, whereas in the more general model (M8) $\omega_s \geq 1$. From standard theory (Chernoff 1954), it follows that the log-likelihood ratio statistic is asymptotically distributed as a 50:50 mixture of a point mass at zero and a χ_1^2-distribution.

Test II may in some cases have more power than test I because of the reduction in the degrees of freedom and because the true asymptotic distribution, and not an ad hoc approximation, is used. However, it may in other cases have less power if there exists a category of positively selected sites with a value of ω that is only slightly larger than one.

Results and Discussion

Using test I, there is evidence that most mammalian reproductive genes examined for positive selection have a class of sites with $\omega_s > 1$, although β-galactosyltransferase does not appear to contain any codons under putative positive selection. This indicates that these mammalian reproductive proteins are possibly subjected to positive Darwinian selection. In some cases, when positive selection was identified in a data set, the corresponding ω_s was barely greater than one. We, therefore, used the new version of the LRT described above (test II), which is robust to the distribution of ω in (0, 1). All comparisons except for acrosin remained significant with this more stringent test of selection (test II; table 1). Thus, there is robust evidence that a large number of mammalian reproductive genes are subjected to positive Darwinian selection.

Sites predicted to be the targets of positive Darwinian selection were identified using an empirical Bayes approach (Nielsen and Yang 1998; Yang et al. 2000). In a few of the proteins studied, the active sites implicated to be involved in sperm-egg interaction have been identified. For example, a disintegrin-like domain in fertilin β has been implicated in sperm-egg binding (Zhu, Bansal, and Evans 2000). Remarkably, the majority (>90%) of the sites in our analysis predicted to be under positive Darwinian selection with posterior probabilities greater than 0.90 fall in the C-terminal portion of the molecule containing this putative sperm-egg–binding domain (see supplementary material: http://www.molbiolevol.org). A similar result was obtained for fertilin α (Wong et al. 2001). Although the selective pressure remains unknown, the observation that the sites predicted to be subjected to positive selection fall in putative sperm-egg–binding domains suggests a selective force relating to male-female interaction, in this case fertilization. These results combined with those of earlier studies (Swanson et al. 2001) suggest that evolutionary analyses such as those described in this study may be a powerful way to identify regions with putative functional significance, which could be tested in functional assays.

The finding of a large number of mammalian reproductive proteins being subjected to positive Darwinian selection may have profound implications on studies of fertility. For example, rapidly evolving reproductive molecules could lead to a mismatch in sperm-egg proteins, which could contribute to infertility. This would be analogous to matches in class I major histocompatibility complex molecules necessary for successful skin grafts or matching blood-type groups for blood transfusions. In sea urchins, there is a significant effect of the genotype of reproductive proteins and the success of individual crosses within a population of one species

(Palumbi 1999). The sites predicted to be subjected to positive Darwinian selection in this study could be involved directly in sperm-egg interaction and thus may be good targets to develop nonhormonal means of contraception aimed at disrupting sperm-egg interaction. Finally, inclusion of evolutionary diversification may help clarify some of the controversies in mammalian fertilization that have arisen due to diverse experimental observations (Wassarman 1999). Our results suggest that analyses for the coevolution of rapidly evolving male-female reproductive proteins may provide evidence that is consistent with interaction between some of the proposed, yet controversial, sperm-egg–binding pairs. Currently, sufficient data are not available to perform these coevolution analyses.

Acknowledgments

We thank Drs. J. Calkins and R. Cardullo and two anonymous reviewers for their comments on the manuscript. This work was supported by NSF grant DEB-0111613 to W.J.S. and NSF grant DEB-0089487 to R.N.

Literature Cited

Anisimova, M., J. P. Bielawski, and Z. Yang. 2001. Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. Mol. Biol. Evol. **18**:1585–1592.

———. 2002. Accuracy and power of Bayes prediction of amino acid sites under positive selection. Mol. Biol. Evol. **19**:950–958.

Baba, T., S. Azuma, S.-I. Kashiwabara, and Y. Toyoda. 1994. Sperm from mice carrying a target mutation of the acrosin gene can penetrate the oocyte zona pellucida and effect fertilization. J. Biol. Chem. **269**:31845–31849.

Chernoff, H. 1954. On the distribution of the likelihood ratio. Ann. Math. Stat. **25**:573–578.

Hardy, D. M., and D. L. Garbers. 1995. A sperm membrane protein that binds in a species-specific manner to the egg extracellular matrix is homologous to von Willebrand factor. J. Biol. Chem. **270**:26025–26028.

Lin, Y., L. H. Kimmel, D. G. Myles, and P. Primakoff. 1993. Molecular cloning of the human and monkey sperm surface protein PH-20. Proc. Natl. Acad. Sci. USA **90**:10071–10075.

Miller, D. J., M. B. Macek, and B. D. Shur. 1992. Complementarity between sperm surface beta-1,4-galactosyltransferase and egg-coat ZP3 mediates sperm-egg binding. Nature **357**:589–593.

Miyado, K., G. Yamada, S. Yamada et al. (12 co-authors). 2000. Requirement of CD9 on the egg plasma membrane for fertilization. Science (Washington DC) **287**:321–324.

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

Palumbi, S. R. 1999. All males are not created equal: fertility differences depend on gamete recognition polymorphisms in sea urchins. Proc. Natl. Acad. Sci. USA **96**:12632–12637.

Richardson, R. T., N. Yamasaki, and M. G. O'rand. 1994. Sequence of a rabbit sperm zona pellucida binding protein and localization during the acrosome reaction. Dev. Biol. **165**: 688–701.

Swanson, W. J., and V. D. Vacquier. 2002. Rapid evolution of reproductive proteins. Nat. Rev. Genet. **3**:137–144.

Swanson, W. J., Z. Yang, M. F. Wolfner, and C. F. Aquadro. 2001. Positive Darwinian selection drives the evolution of several female reproductive proteins in mammals. Proc. Natl. Acad. Sci. USA **98**:2509–2514.

Wassarman, P. M. 1999. Mammalian fertilization: molecular aspects of gamete adhesion, exocytosis, and fusion. Cell **96**: 175–183.

Wong, G. E., X. Zhu, C. E. Prater, E. Oh, and J. P. Evans. 2001. Analysis of fertilin alpha (ADAM1)-mediated sperm-egg cell adhesion during fertilization and identification of an adhesion-mediating sequence in the disintegrin-like domain. J. Biol. Chem. **276**:24937–24945.

Yang, Z., and J. P. Bielawski. 2000. Statistical methods for detecting molecular adaptation. Trends Ecol. Evol. **15**:496–503.

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

Zhu, X., N. P. Bansal, and J. P. Evans. 2000. Identification of key functional amino acids of the mouse fertilin beta (ADAM2) disintegrin loop for cell-cell adhesion during fertilization. J. Biol. Chem. **275**:7677–7683.

Edward Holmes, Associate Editor

Accepted August 22, 2002