DOI: 10.1534/genetics.107.085902

Comparative Analysis of Testis Protein Evolution in Rodents

Leslie M. Turner,*,[1] Edward B. Chuong*,[2] and Hopi E. Hoekstra†

**Division of Biological Sciences, University of California, San Diego, California 92093 and †Department of Organismic and Evolutionary Biology and The Museum of Comparative Zoology, Harvard University, Cambridge, Massachusetts 02138*

Manuscript received December 14, 2007
Accepted for publication May 22, 2008

ABSTRACT

Genes expressed in testes are critical to male reproductive success, affecting spermatogenesis, sperm competition, and sperm–egg interaction. Comparing the evolution of testis proteins at different taxonomic levels can reveal which genes and functional classes are targets of natural and sexual selection and whether the same genes are targets among taxa. Here we examine the evolution of testis-expressed proteins at different levels of divergence among three rodents, mouse (*Mus musculus*), rat (*Rattus norvegicus*), and deer mouse (*Peromyscus maniculatus*), to identify rapidly evolving genes. Comparison of expressed sequence tags (ESTs) from testes suggests that proteins with testis-specific expression evolve more rapidly on average than proteins with maximal expression in other tissues. Genes with the highest rates of evolution have a variety of functional roles including signal transduction, DNA binding, and egg–sperm interaction. Most of these rapidly evolving genes have not been identified previously as targets of selection in comparisons among more divergent mammals. To determine if these genes are evolving rapidly among closely related species, we sequenced 11 of these genes in six Peromyscus species and found evidence for positive selection in five of them. Together, these results demonstrate rapid evolution of functionally diverse testis-expressed proteins in rodents, including the identification of amino acids under lineage-specific selection in Peromyscus. Evidence for positive selection among closely related species suggests that changes in these proteins may have consequences for reproductive isolation.

ONE of the most striking patterns in molecular evolution is that reproductive proteins evolve faster than other protein classes, a pattern consistent across diverse taxa (Singh and Kulathinal 2000; Swanson and Vacquier 2002; Clark *et al.* 2006). These rapidly evolving proteins serve diverse functions in both males and females and act at various stages of the fertilization process ranging from navigation of sperm through the female reproductive tract through egg–sperm fusion (Clark *et al.* 2006). Many questions, however, remain unresolved: (1) Do proteins involved in particular biological functions or participating in specific steps of fertilization evolve more rapidly than others?, (2) Are the same proteins and amino acid sites targets of selection in different taxa?, and (3) Does divergence in reproductive proteins contribute to reproductive isolation between closely related taxa?

In mammals, research on reproductive protein evolution has focused primarily on sequence analysis of candidate genes chosen because of their role in fertilization. This approach has identified positive selection (mainly on the basis of relative rates of nonsynonymous *vs.* synonymous change) acting on genes involved in sperm motility, semen coagulation, sperm–egg binding, and sperm–egg fusion (Clark *et al.* 2006). The functions of numerous proteins involved in fertilization, however, are unknown (Jansen *et al.* 2001; Tanphaichitr *et al.* 2007); therefore, candidate gene approaches are likely to miss important targets of selection. In contrast, a genomewide analysis of reproductive proteins can characterize general patterns of evolution as well as identify rapidly evolving genes. Such genomic approaches have been particularly useful in identifying rapidly evolving male accessory gland proteins (Acps) in Drosophila (Swanson *et al.* 2001a) and crickets (Andres *et al.* 2006; Braswell *et al.* 2006), female reproductive tract proteins in Drosophila (Swanson *et al.* 2001b), and seminal proteins in primates (Clark and Swanson 2005).

We use a genomic approach to characterize reproductive protein evolution in three rodents, mouse (*Mus musculus*), rat (*Rattus norvegicus*), and deer mouse (*Peromyscus maniculatus*). Rodents are an excellent system for investigating mammalian reproductive protein evolution. Fertilization is better characterized in Mus than any other mammal, due to its importance as a model in human reproductive health research (Nixon *et al.* 2007). Both Mus and Rattus have complete genome sequences that are well annotated, enabling broadscale

Sequence data from this article have been deposited with the GenBank Data Libraries under accession nos. EU836254–EU836319.

[1]*Corresponding author:* Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Biology, August Thienemann Strasse 2 24306 Ploen, Germany. E-mail: turner@evolbio.mpg.de

[2]*Present address:* Department of Genetics, Stanford University, Stanford, CA 94305-5120.

comparisons and links to protein function. In contrast, Peromyscus has been studied extensively in the wild and is a speciose genus with well-documented diversity in reproductive morphology, physiology, ecology, behavior, and mating system (Kleiman 1977; Wolff 1989; Dewey and Dawson 2001). Because Peromyscus exhibit a range of reproductive isolation among populations, subspecies, sister species, species, and species groups (Dice 1933; Liu 1953; Maddock and Dawson 1974), we can gain a more in-depth understanding of patterns of reproductive protein evolution and their potential consequences for speciation (Coyne and Orr 2004). Moreover, the intensity of sperm competition and sexual conflict, two selective forces that may drive reproductive protein divergence, is expected to differ among species of Peromyscus with different mating systems. Comparisons of reproductive protein evolution between these species can thus inform our understanding of the nature of selection acting on fertilization.

Here we report a comparative genomic analysis of testis-expressed proteins in these rodents. First, we identify rapidly evolving proteins by comparing expressed sequence tags (ESTs) from testes of *M. musculus*, *R. norvegicus*, and *P. maniculatus* to orthologous sequences from the Mus and Rattus genomes. Comparisons of multiple species allow us to test for differences in lineage-specific rates of evolution. Second, we test for evidence of positive selection on these proteins at a timescale relevant to reproductive isolation by sequencing a subset of rapidly evolving genes in six Peromyscus species. Together, these analyses identify a large number of rapidly evolving proteins, many of which have not been implicated previously as targets of selection and specific amino acid sites that may play a role in reproductive isolation among rodents.

MATERIALS AND METHODS

***P. maniculatus* testis cDNA library construction and EST sequencing:** We had a cDNA library prepared by Amplicon Express (Pullman, WA) using *P. maniculatus* testis tissue from a single adult male with the ZAP-cDNA synthesis kit (Stratagene, La Jolla, CA). EST sequencing is described in Glenn *et al.* (2008); GenBank accession nos. EV469380–472065). Briefly, the cDNA library was amplified, and phagemids excised according to the manufacturer's protocol. Resulting colonies were grown overnight in Luria-Bertani/ampicillin broth in deep-well plates. PCR-amplified inserts from bacterial cultures of 4800 clones were sequenced from the 5′ end using BigDye terminator (v. 3.1, Applied Biosystems, Valencia, CA) and run on an ABI automated sequencer (3100, 3730, 3730XL, Applied Biosystems). We made base calls (with embedded PHRED, Ewing and Green 1998; Ewing *et al.* 1998) and trimmed both vector sequences and sequence ends from ESTs to reduce error rate to <0.05 (PHRED quality value >13) using the program ALIGNER (CodonCode, Dedham, MA). We discarded sequences <90 bp in length and changed remaining bases with quality values <13 to unknowns. We assembled ESTs into contigs using the CAP3 sequence assembly program (Huang and Madan 1999).

Evolutionary EST analysis: We obtained testis cDNA sequences from *M. musculus* (Stratagene mouse testis library, 6068 sequences; RIKEN full-length enriched mouse testis cDNA library, 14,000 sequences) and *R. norvegicus* (NIH_MGC_238 library, 13,046 sequences) from the NCBI dbEST database. We identified orthologs by pairwise comparison of ESTs to transcript libraries from the NCBI Reference Sequence (RefSeq) database (nuclear chromosomal cDNA only, downloaded December, 2006) using FASTX (v. 3.3, default settings, Pearson 1990). We made the following four comparisons (EST *vs.* RefSeq): *P. maniculatus vs. M. musculus* (PM), *P. maniculatus vs. R. norvegicus* (PR), *M. musculus vs. R. norvegicus* (MR), and *R. norvegicus vs. M. musculus* (RM). We made the reciprocal MR and RM comparisons to increase the sample size of ortholog pairs and to assess the effects of differences between EST and genome sequence sources on evolutionary rate estimates. We defined orthologs as sequence pairs that have a minimum of 40% sequence identity for $>20\%$ of EST length. If multiple RefSeqs met these criteria, the most likely ortholog was determined as either: (1) the sequence with the greatest amino acid identity (% sequence identity $\times$ alignment length) or (2) the sequence with the lowest divergence at synonymous sites (d_S). There were few discrepancies between these criteria, and most of these were matches to alternate isoforms. In these few cases, we used the first criterion, amino acid identity, because it is more conservative; estimates of rate of evolution (*i.e.*, ω, defined below) for orthologous pairs on the basis of amino acid identity were the same or lower than estimates determined for best match on the basis of d_S. We concatenated nonoverlapping ESTs matching the same RefSeq.

For each orthologous pair, we estimated the rate of evolution as the ratio of nonsynonymous substitution rate to synonymous substitution rate ($d_N/d_S = \omega$). ω-Values for neutrally evolving genes are expected to equal one whereas ω-values <1 indicate purifying selection and $\omega > 1$ is considered strong evidence for positive selection. This test is stringent, as pairwise ω-values are averaged across all amino acid sites. A literature survey of studies that used a maximum likelihood (ML) approach (Yang *et al.* 2000) to detect selection showed that most genes with overall $\omega > 0.5$ show evidence for positive selection acting on a subset of amino acid sites (Swanson *et al.* 2004). We therefore classify all genes identified here with $\omega > 0.5$ as "rapidly evolving."

We estimated ω using ML as implemented in the CODEML program from the PAML package (runmode -2, v 3.14; Yang 2000). We excluded ortholog pairs with $d_S > 1.5$ from further analysis as these are unreliable due to estimation errors (Castillo-Davis *et al.* 2004). For pairs with estimated ω-values >1, we ran an additional model in PAML with ω fixed at 1. To determine whether the estimated value of ω was significantly >1, we compared the estimated ω-model to the fixed ω (neutral) model using a likelihood ratio test (LRT). The test statistic for the LRT is the negative of twice the difference in log likelihoods between models ($-2\Delta\ln L$), and is χ^2 distributed with degrees of freedom equal to the difference in number of estimated parameters (in this case 1). Alignment of ESTs, identification of orthologs, and implementation of models in PAML were automated using perl, Bioperl (v. 1.5; Stajich *et al.* 2002), and PHP scripts.

For rapidly evolving genes with orthologs identified in all three rodents, we performed a three-species comparison to identify lineage-specific increases in the rate of amino acid change. We estimated lineage-specific ω-values using the free ratios model in CODEML and performed a LRT comparing the free ratios model to the single-ratio model to test whether there is significant evidence of rate variation across lineages (test statistic $= -2\Delta\ln L$, χ^2, d.f. $= 2$).

We obtained expression information for *M. musculus* RefSeqs from the Genomics Institute of the Novartis Research Foundation (GNF) gcRMA-condensed data set (Wu and Irizarry 2005). We classified expression class solely on the basis of the Mus data because testis expression data are not available for Rattus or Peromyscus; we thus assumed that maximal tissue of expression is the same for all three lineages. We classified tissue specificity following Winter *et al.* (2004); tissue specificity (T_S) is defined as the expression of a given gene in one tissue relative to total expression of that gene in all tissues. Genes with maximum T_S ($maxT_S$) < 0.08 are considered "housekeeping" (H) genes. We classified the remaining ESTs with $maxT_S$ for testis as testis-specific (TS) and those with $maxT_S$ for another tissue are classified as nontestis-specific (NTS). Expression patterns from three genes with different values of $maxT_S$ are provided as examples in supplemental Figure S1.

We then compared rates of evolution between H, TS, and NTS ESTs. Because tissue-specific genes evolve more rapidly than genes with broader expression, likely due to reduced pleiotropy (Duret and Mouchiroud 2000; Winter *et al.* 2004), we compared the ω-distribution of ESTs in each expression class using an analysis of covariance (ANCOVA) with level of tissue specificity ($maxT_S$) as a covariate. We transformed ω and $maxT_S$ values toward normality and equal variances between groups; ω-values were natural log-transformed and $maxT_S$ values were arcsine-transformed. To equalize variances among groups, we excluded ESTs with $\omega = 0$. A greater proportion of NTS/H ESTs have $\omega = 0$, thus their exclusion results in a conservative test.

We identified putative secreted proteins (containing signal peptides) using Signal P (v. 3.0; Nielsen and Krogh 1998; Bendtsen *et al.* 2004) and cell-surface proteins (with predicted transmembrane domains) using TMHMM (Sonnhammer *et al.* 1998; Krogh *et al.* 2001). We tested whether average ω differed significantly between extracellular (contain a signal peptide and/or a transmembrane domain) and intracellular protein genes using permutation tests (10,000 permutations).

We determined the function of Mus EST homologs using the PANTHER classification system (Thomas *et al.* 2003). We tested whether any particular biological process or molecular function was overrepresented among rapidly evolving genes by comparing the proportion in the rapidly evolving group relative to expected on the basis of representation among total EST homologs for each species. We tested the significance of overrepresentation using the binomial test (Cho and Campbell 2000) with Bonferroni correction for multiple comparisons.

Identification of protein domains in Peromyscus EST sequences: We searched for all unique Peromyscus testis sequences (unigenes) in the InterPro-combined protein database (Mulder *et al.* 2007) using InterProScan (Zdobnov and Apweiler 2001), which uses a variety of search algorithms to identify homology between six-frame translations of input nucleotide sequences and known protein domains. This method allows for identification of domains in all ESTs, including those that do not have orthologs identified in Mus or Rattus.

Additional sequencing in Peromyscus: We obtained testis tissue from a single male from each of six Peromyscus species (*P. aztecus*, *P. californicus*, *P. eremicus*, *P. leucopus*, *P. maniculatus*, and *P. polionotus*) from the Peromyscus Genetic Stock Center (supplemental Table S1). Tissue was excised from freshly sacrificed adult males and stored in RNAlater solution (Sigma, St. Louis). We extracted RNA using the RNeasy kit (QIAGEN, Valencia, CA) and synthesized cDNA using a Superscript III RT kit (Invitrogen, Carlsbad, CA). We amplified all genes under standard PCR conditions, using primers designed by aligning *P. maniculatus* EST sequences to GenBank sequences from Mus and Rattus. To determine species relationships, we sequenced a 1213-bp region of the mitochondrial genome (including *COIII* and *ND3*) from one individual from each of the six Peromyscus species (supplemental Table S1) using published primers (Hoekstra *et al.* 2004).

We directly sequenced or cloned (TOPO-TA, Invitrogen) PCR products. We performed cycle sequencing with BigDye terminator (v. 3.1, Applied Biosystems) and ran products on an ABI 3100 automated sequencer (Applied Biosystems). We checked base calls by eye, assembled contigs, and aligned sequences in SEQUENCHER (Gene Codes, Ann Arbor, MI). We used MUSCLE (default parameter settings; Edgar 2004) when sequence alignments were ambiguous. A large repetitive region from one of the genes (*Gm1276*, see results) was excluded because reliable alignment was not possible.

Analysis of Peromyscus testis-expressed gene sequences: To determine species relationships, we constructed Bayesian and ML phylogenies of the six species, on the basis of the mitochondrial sequences and 1201 bp of the nuclear genes *Mc1r* and *Lcat* (Turner and Hoekstra 2006). We identified the most appropriate substitution model (GTR + Γ) using MODELTEST v. 3.7 (Posada and Crandall 1998). A partition homogeneity test implemented in PAUP* (Swofford 2002) was not significant, indicating no conflicts between data partitions. We performed Bayesian analysis in MRBAYES (v. 3.1; Huelsenbeck and Ronquist 2001), with data partitioned by gene and codon position. We performed two runs for 10 million generations and discarded the first million generations as burn-in. The 99% credible set for the Bayesian analysis contains a single tree, identical in topology to the ML tree: (((*P. polionotus*, *P. maniculatus*), *P. leucopus*), *P. aztecus*, (*P. eremicus*, *P. californicus*)). This topology is consistent with published species trees of Peromyscus (Turner and Hoekstra 2006; Bradley *et al.* 2007).

Using this ML/Bayesian tree, we implemented the codon-based ML method (Nielsen and Yang 1998; Yang *et al.* 2000) to detect positive selection in Peromyscus testis genes. This method employs a LRT to compare a neutral model, where ω for all sites is constrained to be <1, to a selection model where a subset of sites has $\omega > 1$ (test statistic $= -2\Delta\ln L$, χ^2). We performed the following model comparisons (neutral *vs.* selection): M1a *vs.* M2a, M7 *vs.* M8, and M8A *vs.* M8. M1a has two site classes, the first with $0 < \omega < 1$ and the second with $\omega = 1$. M2a adds an additional "selection" class with $\omega \geq 1$. In M7, ω varies as a beta distribution between 0 and 1, and M8 adds a selection class with $\omega \geq 1$. M8A is a modified version of M8 where ω for the selection class is constrained to equal one. The M8A *vs.* M8 comparison tests whether ω is significantly >1, providing a control for false positives resulting from a poor fit of the data to the beta distribution. For this comparison, the test statistic is distributed as a 50:50 mixture of a point mass at zero and a χ^2 distribution with one degree of freedom (Swanson *et al.* 2003). We implemented codon models in CODEML. Specific amino acid sites subject to positive selection were identified using the Bayes empirical Bayes (BEB) procedure (Yang *et al.* 2005).

In addition, we applied codon models to determine whether genes positively selected within Peromyscus have evidence for positive selection among divergent species of mammals. For the five genes with evidence for positive selection within Peromyscus (see results), we identified homologs from other mammals in GenBank using BLAST (see supplemental Table S2 for species and accession numbers). To avoid significant results due to positive selection within Peromyscus, we included sequence from only a single species, *P. maniculatus*, in these analyses. We aligned amino acid sequences using default parameter settings in MUSCLE, adjusted the corresponding nucleotide alignments in MEGA (Kumar *et al.* 2001), and excluded sites with alignment gaps. We constructed neighbor-joining trees in PAUP* using model

parameters determined in MODELTEST. We ran codon models in CODEML, as above.

RESULTS

EST sequencing: Sequencing of 4800 ESTs from the *P. maniculatus* testis cDNA library resulted in 3840 quality sequences >90 bp in length. After removal of redundant sequences and assembly of overlapping sequences into contigs there was a total of 2364 unigenes (446 contigs, 1918 singlets).

Evolutionary EST analysis: To identify the most rapidly evolving testis proteins, we compared orthologous genes in Peromyscus, Mus, and Rattus. We found orthologs in Mus and/or Rattus for ~43% of unique *P. maniculatus* EST sequences (Table 1), resulting in 1014 (PM) and 993 (PR) orthologous pairs. The 20,068 Mus EST sequences included 11,203 unigenes; we identified orthologs in Rattus for 37% of unigenes, for a total of 4171 pairs. Thirteen thousand forty-six Rattus ESTs collapsed into 7448 unigenes and we found Mus orthologs for 56% of these, for a total of 4207 orthologous pairs. The lower proportion of Mus ESTs with identified Rattus orthologs is not surprising because the Rattus genome sequence was completed more recently than the Mus genome, is therefore less well annotated, and has lower sequencing coverage (Waterston *et al.* 2002; Gibbs *et al.* 2004).

For each EST–RefSeq comparison, estimates of ω for the vast majority of ortholog pairs are $\ll 1$, consistent with the action of purifying selection (Table 1). A small percentage of pairs (1.3–2.4%), however, have $\omega > 1$ (a signature of positive selection) and three of these pairs have ω-values significantly >1. These three genes are all from the MR comparison and include a hypothetical protein of unknown function (*LOC691850*) and two microtubule-associated proteins: a signaling protein involved in spermatogenesis (Mast2, Walden and Cowan 1993) and a protein with microtubule motor activity and a lipid-binding domain whose function in testis is unknown (Stard9, Kanno *et al.* 2007). An additional 7.5–12.2% of ortholog pairs have $0.5 < \omega < 1$. All rapidly evolving genes ($\omega > 0.5$) are listed in supplemental Table S3.

A representative plot of d_N *vs.* d_S values for all pairs from the PM comparison is presented in Figure 1A. Proportions of ESTs in three ω-classes ($\omega < 0.5$, $0.5 < \omega < 1$, $\omega > 1$) are not significantly different among three of the EST–RefSeq comparisons (PM, PR, RM, $P = 0.96$; Pearson's χ^2, d.f. = 4). The MR comparison has the largest proportion of ESTs in both rapidly evolving classes (12.2%, $0.5 < \omega < 1$; 2.4%, $\omega > 1$), resulting in a significant effect of species comparison on ω-class when MR is included ($P < 0.001$; Pearson's χ^2, d.f. = 6). Some genes are rapidly evolving in all species comparisons, whereas other genes appear to be rapidly evolving in just one or two lineages (Figure 2B). Overall, 44% (72/163) of rapidly evolving genes with orthologous pairs identified in multiple species comparisons are rapidly evolving in more than one comparison and 18% (29/163) are rapidly evolving in all comparisons. The genes that have $\omega > 0.5$ in all comparisons (PM; PR; RM and/or MR) are listed in Table 2.

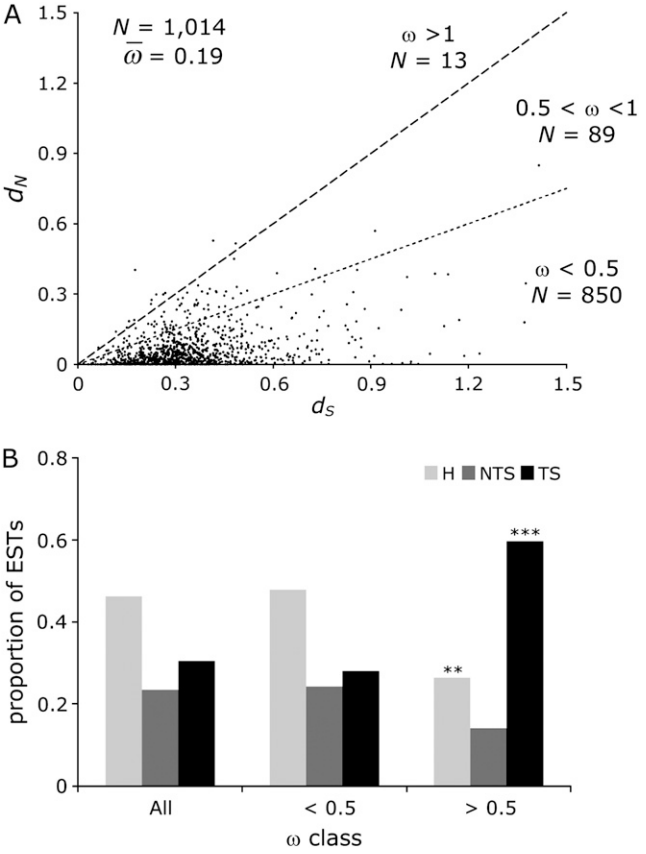

Figure 1.—Evolutionary rates of testis-expressed ESTs. (A) d_N *vs.* d_S estimated in PAML (Yang 2000), each point represents the respective substitution rate for a given *Peromyscus maniculatus* testis EST *vs.* its *Mus musculus* homolog. (B) Proportion of ESTs in each expression class among all ESTs and among ESTs grouped by ω-value; H, housekeeping; NTS, nontestis-specific; TS, testis-specific; $^{**}P < 0.01$ and $^{***}P < 0.001$ in a two-tailed binomial test for under- or overrepresentation of ESTs of an expression type in the given ω-class.

Testis-specific genes evolve rapidly: Genes expressed in testes may also be expressed in other tissues; we used expression data from Mus to determine which genes are testis specific and which have broader patterns of expression. Using these data, we tested if rates of protein evolution are correlated with expression pattern. Expression data were available for 62–73% of Mus orthologs. For all species comparisons, mean ω-values for testis-specific genes were higher than overall means and means for other expression classes, indicating that testis-specific genes evolve more rapidly on average than nontestis-specific and housekeeping genes (Table 1). This analysis shows that there is a highly significant

TABLE 1

Evolutionary rates of testis-expressed genes

Comparison	N	Homologs	$\overline{L_C}$	$\overline{d_N}$	$\overline{d_S}$	$\overline{\omega}$	Rapidly evolving $0.5 < \omega < 1.0$	Rapidly evolving $\omega > 1.0$	$\overline{\omega}$ IC	EC	H	NTS	TS
Peromyscus *vs.* Mus	2,364	1014	157	0.06	0.35	0.19 (0.17–0.20)	76 (7.5%)	13 (0) (1.3%)	0.18 (0.17–0.20) (908)	0.23* (0.19–0.27) (106)	0.13 (0.12–0.16) (343)	0.15 (0.13–0.18) (174)	0.26*** (0.23–0.29) (226)
Peromyscus *vs.* Rattus	2,364	993	157	0.07	0.36	0.19 (0.18–0.21)	81 (8.2%)	14 (0) (1.4%)	0.18 (0.17–0.20) (858)	0.20 (0.17–0.24) (103)	0.13 (0.11–0.15) (325)	0.16 (0.14–0.19) (165)	0.27*** (0.23–0.30) (209)
Mus *vs.* Rattus	11,203	4171	147	0.07	0.25	0.28 (0.26–0.30)	509 (12.2%)	99 (3) (2.4%)	0.27 (0.25–0.29) (3474)	0.33* (0.31–0.35) (697)	0.23 (0.21–0.24) (1225)	0.23 (0.21–0.25) (710)	0.33*** (0.30–0.35) (667)
Rattus *vs.* Mus	7,448	4207	200	0.06	0.24	0.23 (0.22–0.23)	342 (8.1%)	60 (0) (1.4%)	0.22 (0.22–0.23) (3558)	0.25** (0.23–0.27) (649)	0.19 (0.18–0.20) (1530)	0.19 (0.18–0.20) (936)	0.29*** (0.28–0.31) (586)

N, number of unique ESTs compared to Refseqs. Homologs, number of homolog pairs identified, as defined in MATERIALS AND METHODS. $\overline{L_C}$, mean length of aligned region in codons. $\overline{d_N}$, mean nonsynonymous substitution rate; $\overline{d_S}$, mean synonymous substitution rate; $\overline{\omega}$, mean ω for all ESTs. Number and percentage of rapidly evolving ESTs are given for two ω-classes: $0.5 < \omega < 1.0$, and $\omega > 1.0$. For the $\omega > 1.0$ class, the number of ESTs with ω significantly >1.0 is given in parentheses (as determined by a likelihood ratio test comparing estimated ω-models *vs.* models with ω fixed at 1.0, implemented in PAML). Mean ω-values and 95% confidence intervals (on the basis of 1000 bootstrap replicates, in parentheses) are given for IC, intracellular protein genes; EC, extracellular protein genes (with a predicted signal sequence and/or transmembrane domain); H, housekeeping genes; NTS, nontestis-specific genes; TS, testis-specific genes. The number of ESTs in each category is given in parentheses below. We compared $\overline{\omega}$ for EC *vs.* IC genes using permutation tests (10,000 permutations; $^*P < 0.05$; $^{**}P < 0.01$); and H, NTS, and TS genes using an ANCOVA determining effects of expression class on ω, controlling for degree of tissue specificity of expression ($^{***}P < 0.0001$).

effect of testis-specific expression on ω in all species comparisons (ANCOVA, $P < 0.0001$).

In addition, we compared the proportion of genes from each expression class between groups of ESTs with different ω-values ($\omega < 0.5$ *vs.* $\omega > 0.5$). The results provide an intuitive demonstration of the elevated evolutionary rate of testis-specific genes. For all four EST–RefSeq comparisons, there is a significant relationship between expression class and ω-class (Pearson's χ^2, d.f. = 2, $P < 0.0001$). Specifically, the proportion of testis-specific genes was significantly higher among rapidly evolving genes than expected on the basis of the proportion of all genes that are testis specific (two-tailed binomial test, $P < 0.0001$). Overrepresentation of testis-specific genes among rapidly evolving genes from one comparison (PM) is depicted in Figure 1B.

Rapidly evolving genes are functionally diverse: To determine if particular functional classes of genes tend to be rapidly evolving, we used the PANTHER classification system to assign genes to functional categories. In all three pairwise species comparisons, genes unclassified for both biological process and molecular function are overrepresented in the rapidly evolving class ($P < 0.002$). In addition, defense and immunity proteins ($P < 0.002$) and KRAB box transcription factors ($P < 0.005$) are overrepresented in the Mus–Rattus comparison (MR/RM).

The list of rapidly evolving genes (Table 2) includes genes with $\omega > 0.5$ in all three pairwise species comparisons (PM; PR; RM and/or MR), and 11 genes chosen for sequencing in additional Peromyscus species because they had the highest pairwise ω-values in the PM comparison (excluding hypothetical proteins) in initial EST comparisons. Three genes from the latter category (*Gsg1*, *H1fnt*, and *Smcp*) had high ω-estimates in preliminary screens, but much lower ω-estimates in the final screen, subsequent to corrections of alignments or changes in Mus RefSeqs. The amount of functional information available for these rapidly evolving genes varies. Some genes have known roles in sperm–egg interaction (*Acr*, Howes *et al.* 2001; *Spa17*, Richardson *et al.* 1994; *Spag8*, Cheng *et al.* 2007) or spermatogenesis (*Hils1*, Yan *et al.* 2003). Another set of genes has inferred function on the basis of domain homology; a wide variety of functions are represented including receptor activity, DNA binding, and protein binding. However, the majority of genes have no available functional information.

We also compared values of ω for secreted (contain a signal peptide) and cell-surface proteins (contain a transmembrane domain) to ω-values of intracellular proteins. Extracellular proteins are more likely to interact with foreign proteins, such as those of pathogens, the female reproductive tract and gametes, or sperm/seminal proteins from another male and thus may experience stronger selection than intracellular proteins (Swanson *et al.* 2001a; Clark and Swanson 2005). We find that extracellular proteins have higher ω-values on average than intracellular proteins for all species comparisons (Table 1), and this difference is statistically significant for the PM ($P = 0.037$), RM ($P = 0.003$), and MR ($P = 0.017$) comparisons.

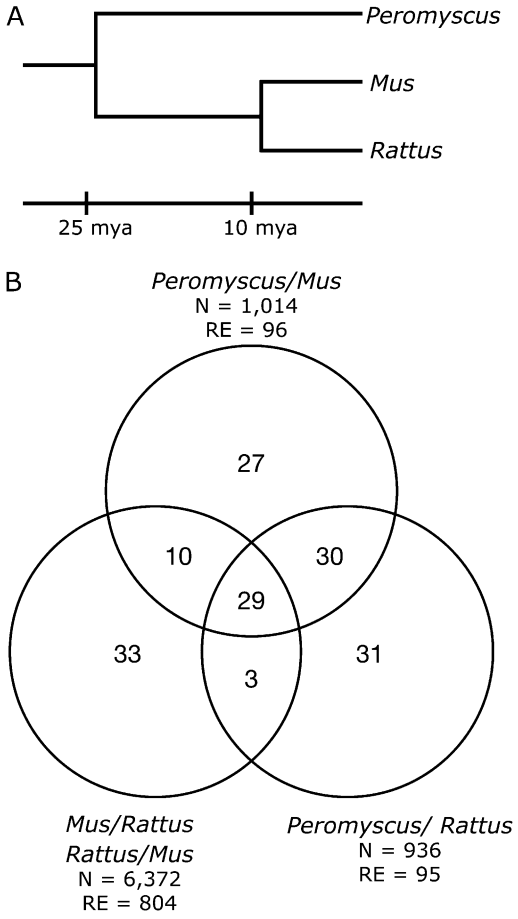

Figure 2.—Rapidly evolving testis genes in three rodent lineages. (A) Relationships and divergence times between Peromyscus, Mus, and Rattus (Steppan *et al.* 2004). (B) Rapidly evolving ($\omega > 0.5$) testis genes with homologs identified in >1 species comparison are shown, including genes identified as rapidly evolving in one or multiple comparisons. Total number of EST pairs (N) and total number of rapidly evolving genes (RE) are given for each species comparison.

We looked for protein domain homology in all Peromyscus ESTs using InterProScan to obtain information about possible functions of genes when we could not identify orthologs in Mus or Rattus (nonmatches). However, this analysis was not informative; very few nonmatches had any indication of homology. Seventeen nonmatches have predicted signal protein sequences or transmembrane domains, indicating they are extracellular and may interact with foreign proteins.

Evidence for adaptive evolution of testis genes in Peromyscus: For the 11 most rapidly evolving genes identified in the preliminary genomic analysis, we sequenced most or all of the coding regions in six Peromyscus species to determine whether there is evidence of adaptive change between closely related taxa. We specifically

TABLE 2

Rapidly evolving testis-expressed genes

Gene	Symbol	Chromosome (Mus)	Gene ontology	Function	ω_{PM}	ω_{PR}	ω_{MR}	ω_{RM}
Hypothetical protein LOC70900	*4921517D22Rik*[b]	13	Unknown	Unknown	2.28	1.20	0.73	0.63
Hypothetical protein LOC238663	*4932411G14Rik*[b]	13	Unknown	Unkown	1.24	1.27	ND	0.72
Hypothetical protein LOC242838	*4932412H11Rik*[b]	5	Protein binding	Unknown	1.21	1.37	1.44	ND
Leucine-rich region containing 50	*Lrrc50*[a,b]	8	Protein binding	Unknown	1.17	1.32	1.16	0.11
Histone H1-like protein in spermatids 1	*Hils1*[a,b]	11	DNA binding; histone binding	DNA condensation during spermiogenesis (Yan *et al.* 2003)	1.13	0.89	1.24	ND
Hypothetical protein LOC210940	*4931408C20Rik*[b]	1	Unknown	Unknown	0.91	1.01	ND	0.87
Gene model 1276	*Gm1276*[a,b]	19	Receptor activity; signal transduction	Unknown	0.88	0.82	ND	0.82
PHD finger protein 8	*Phf8*[a]	X	DNA binding; metal ion binding; protein binding; zinc ion binding	Unknown	0.81	0.72	0.44	ND
Chemokine-like factor isoform 1	*Cklf*[b]	8	Cytokine activity; chemotaxis	Unknown	0.81	0.54	0.55	ND
Preproacrosin	*Acr*[a]	15	Acrosin activity; amidase activity; fucose binding; hydrolase activity; mannose binding; peptidase activity; protein binding; serine-type endo-peptidase activity	Secondary binding to zona pellucida (ZP2); dispersal of acrosomal contents	0.80	0.40	0.23	0.17
Hypothetical protein LOC71831 isoform 3	*1700007B14Rik*[b]	8	Unknown	Unknown	0.80	0.70	ND	0.63
Coiled-coil-helix-coiled-coil-helix domain containing 6	*Chchd6*[b]	6	Unknown	Unknown	0.77	0.60	0.69	ND
Hypothetical protein LOC75275	*4930563P21Rik*[b]	2	Unknown	Unknown	0.77	0.67	0.51	ND
Hypothetical protein LOC78174	*4930503B16Rik*[b]	5	Cytochrome-c oxidase activity; electron transport; mitochondrial respiratory chain	Unknown	0.75	0.53	0.53	ND
Sperm-associated antigen 8	*Spag8*[a,b]	4	Unknown	Unknown	0.72	0.71	ND	0.63
Hypothetical protein LOC381816	*4922502D21Rik*[b]	6	Sugar binding	Unknown	0.72	0.95	1.18	1.07
Acrosome formation associated factor	*Afaf*[b]	4	Unknown	Acrosome formation during spermiogenesis (Li *et al.* 2006)	0.71	0.90	0.52	ND
CKLF-like MARVEL transmembrane domain containing 2A	*Cmtm2a*[b]	8	Cytokine activity; protein binding; transcription corepressor activity; chemotaxis; negative regulation of transcription (DNA dependent)	Androgen receptor corepressor involved in regulation of transcription (Jeong *et al.* 2004)	0.68	0.61	0.67	0.67

(*continued*)

TABLE 2

(Continued)

Gene	Symbol	Chromosome (Mus)	Gene ontology	Function	ω_{PM}	ω_{PR}	ω_{MR}	ω_{RM}
Sperm autoantigenic protein 17	*Spa17*[a,b]	9	cAMP-dependent protein kinase regulator activity	Zona pellucida binding	0.67	0.66	0.52	ND
Similar to protein C14orf32 homolog	*C130032J12Rik*[b]	14	Unknown	Unknown	0.67	0.52	ND	0.95
Hypothetical protein LOC239036	*4930596D02Rik*[b]	14	Calcium ion binding; guanyl-nucleotide exchange factor activity; regulation of small GTPase-mediated signal transduction; small GTPase-mediated signal transduction	Unknown	0.64	0.63	0.51	0.75
Hypothetical protein LOC271036 (CatSper$^{\beta}$)	*4932415G16Rik*[b]	12	Unknown	Part of CatSper1 ion channel protein complex, required for sperm hyperactivation (Liu *et al.* 2007)	0.63	0.81	0.57	ND
Gene model 884	*Gm884*[b]	11	Unknown	Unknown	0.62	0.56	0.64	ND
Hypothetical protein LOC73309	*1700047L15Rik*[b]	12	Unknown	Unknown	0.59	0.63	0.59	ND
Hypothetical protein LOC67687 isoform 2	*1700011L22Rik*[b]	8	Unknown	Unknown	0.59	0.59	0.75	0.82
Lysosomal-associated membrane protein 1	*Lamp1*[b]	8	Unknown	Release of spermatozoa from epithelium during spermatogenesis (Guttman *et al.* 2004)	0.58	0.53	ND	0.54
Testis-specific protein Ddc8	*Ddc8*[a,b]	11	Unknown	Unknown	0.56	0.68	0.54	0.34
Hypothetical protein LOC70980	*4931431F19Rik*[b]	7	Unknown	Unknown	0.55	0.54	0.61	0.56
Spermatogenesis-associated 3	*Spata3*[b]	1	Apoptosis; spermatogenesis	Unknown	0.54	0.72	1.16	ND
Similar to kinesin-like motor protein C20orf23	*C20orf23*[b]	2	Unknown	Unknown	0.54	0.70	1.82	ND
Germ cell-specific gene 1	*Gsg1*[a,c]	6	Unknown	Unknown	0.44	0.30	ND	0.36
Histone H1 variant	*H1fnt*[a,c]	15	DNA binding	DNA condensation during spermiogenesis, essential for proper nuclear morphology (Martianov *et al.* 2005; Tanaka *et al.* 2005)	0.18	0.14	0.74	0.52
Sperm mitochondria-associated cysteine-rich protein	*Smcp*[a,c]	3	Selenium binding	Sperm motility (Nayernia *et al.* 2002)	0.14	ND	ND	ND

Gene names, symbols, and gene ontology (GO) terms are indicated for Mus homologs for the most rapidly evolving proteins. All ω-values were estimated in PAML (Yang 2000). P, *P. maniculatus*; M, *M. musculus*; R, *R. norvegicus*. ω indicates pairwise ω between EST and RefSeq for the species pair, *e.g.*, ω_{PM} is between *P. maniculatus* EST and *M. musculus* RefSeq.

[a] Candidate genes sequenced in additional Peromyscus species.

[b] $\omega > 0.5$ in all comparisons.

[c] Included because initial screen showed high ω-values.

TABLE 3

Adaptive evolution of testis-expressed genes in Peromyscus

Gene	L_C	ω_{PM}	ω *vs.* Mus	ω Peromyscus	Variable aa sites (%)	M8 *vs.* M8A	ω_s	p_s
Lrrc50	622	1.17	0.30	0.41	65 (10.4)	0.039*	3.90	0.05
Hils1	162	1.13	1.16	0.42	15 (9.3)	0.416	NA	NA
Gm1276	830	0.88	0.94	0.69	74 (8.9)	0.005**	10.12	0.02
Phf8	447	0.81	0.27	0.47	118 (26.4)	0.127	NA	NA
Acr	428	0.80	0.42	0.22	18 (4.2)	0.012*	7.55	0.01
Spag8	263	0.72	0.64	0.57	54 (20.5)	0.500	NA	NA
Spa17	147	0.67	0.59	0.16	4 (2.7)	0.283	NA	NA
Ddc8	539	0.56	0.47	0.55	50 (9.3)	0.042*	2.19	0.28
Gsg1	364	0.44	0.48	0.34	29 (8.0)	0.022*	2.18	0.17
H1fnt	304	0.18	0.56	0.28	21 (6.9)	0.225	NA	NA
Smcp	136	0.14	0.13	0.11	7 (5.1)	0.500	NA	NA

L_C, length of sequence analyzed in codons; ω_{PM}, ω of the Peromyscus EST *vs.* Mus homolog; ω *vs.* Mus, pairwise ω for the entire *P. maniculatus* sequence *vs.* the Mus homolog; ω Peromyscus, ω in Peromyscus sample, averaged across all sites and lineages [estimated with PAML, M0 (Yang 2000)]; M8 *vs.* M8A, *P*-value of likelihood ratio test; ω_s, ω estimate for "$\omega > 1$" class; p_s, proportion of sites in the "$\omega > 1$" class for M8. *$P < 0.05$, **$P < 0.01$.

looked for evidence of rapid amino acid change as indicated by high values of ω, evidence for positive selection on a subset of amino acid sites, and changes in protein length.

Estimates of pairwise ω for the entire coding region sequenced in *P. maniculatus vs.* Mus homologs (Table 3) were consistent with ω-values for the shorter ESTs in some instances (*e.g.*, *Hils1* and *Gm1276*) and inconsistent in others (*e.g.*, *Lrrc50* and *Acr*). Thus, the EST screen identified both genes with high rates of evolution across their entire length as well as genes with rapidly evolving regions. However, ω within Peromyscus is not significantly correlated with pairwise ω-estimates of *P. maniculatus vs.* Mus ($P = 0.16$), although there is a nonsignificant positive trend ($R^2 = 0.20$). We performed a three-species analysis of *P. maniculatus*, Mus, and Rattus sequences for each gene in PAML to estimate lineage-specific values of ω. We tested the hypothesis that the estimate of ω for the Peromyscus lineage is a better predictor of ω among Peromyscus species, as this value is not affected by evolution along the Mus lineage. However, there is significant evidence of variation in ω among lineages for only 3 of the 11 genes (*H1fnt*, *Smcp*, and *Gsg1*), and the lineage-specific estimates were actually poorer predictors ($P = 0.60$, $R^2 = 0.03$) of ω within Peromyscus than the pairwise *P. maniculatus*–Mus estimates. Three of the 11 genes (*Gm1276*, *Spag8*, and *Ddc8*) have $\omega > 0.5$ within Peromyscus, suggesting they are rapidly evolving and may be subject to positive selection.

Using a ML approach, we tested whether any of the 11 genes have statistical evidence for positive selection within Peromyscus. Results from the ML codon models indicate that 5 of the 11 genes show evidence for positive selection in the Peromyscus genus (Table 3). For each of these genes, comparisons of M8 to M8A are significant, indicating that a proportion of sites are subject to selection and ω for the selected class is significantly >1. In addition, comparisons of M2 *vs.* M1 as well as M8 *vs.* M7 are also significant for one of these genes (*Gm1276*).

For each gene sequenced in Peromyscus, we determined several estimates of evolutionary rate: pairwise ω of EST sequences *vs.* Mus and *vs.* Rattus RefSeqs, lineage-specific ω in Peromyscus determined through three-species analysis and overall ω determined through comparison of the full *P. maniculatus* sequence to Mus. Surprisingly, none of these measures was a good predictor of which genes have evidence for positive selection within Peromyscus. For example, of the five positively selected genes, one gene has the highest estimate of ω *vs.* Mus from the EST screen (*Lrrc50*), whereas another has an ω-estimate <0.5 (*Gsg1*). Overall measures of ω within Peromyscus for positively selected genes ranged from 0.22 (*Acr*) to 0.69 (*Gm1276*). Further, none of these genes have remarkably high levels of amino acid variation (range 4.2–10.4% variable aa sites), highlighting how genes can be subject to positive selection even when overall variability is low.

Amino acid alignments of these five genes are given in Figure 3, and amino acid sites assigned to the positive selection class using the BEB procedure are indicated. With the exception of two sites in ACR (397 and 412), the posterior probabilities of assignment of the sites to the positive selection class are <0.95. These results provide a preliminary indication of the spatial distribution of target sites along the protein. For example, in ACR and LRRC50, target sites are clustered, whereas in GM1276 and DDC8, sites are scattered.

In addition to changes in amino acid sequence, changes in protein length in response to selection have been described in reproductive proteins (Podlaha and Zhang 2003; Podlaha *et al.* 2005; Hawthorne *et al.* 2006). We examined length variation to determine if there is evidence for this type of change in Peromyscus.

One gene, *Phf8*, has premature stop codons, some of which evolved along the lineage between Peromyscus

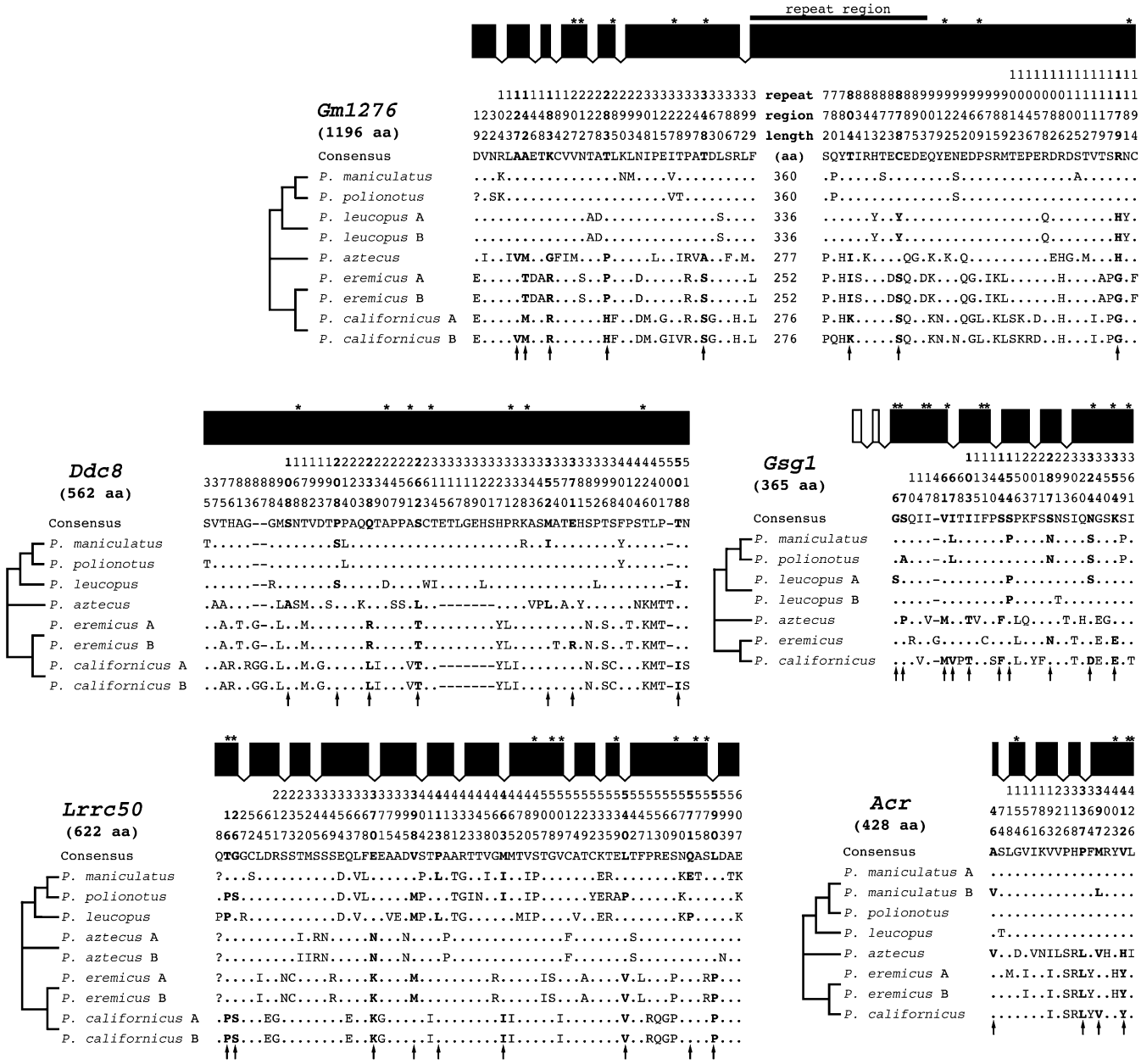

FIGURE 3.—Positively selected testis genes in Peromyscus. Alignments of variable amino acid sites are shown for each gene. Dots indicate identity with the consensus sequence. Amino acid sites identified as positively selected with BEB analysis are indicated with arrows and are in boldface type. Relationships between species are shown to the left of each alignment (on the basis of maximum likelihood analysis of 1213 bp of the mitochondrial genome and 1201 bp of the nuclear genes *Mc1r* and *Lcat*). Exon/intron structure of each gene is shown above the alignment. Boxes indicate exons and are drawn to scale within each gene; open boxes indicate noncoding exons and solid boxes, coding exons. Asterisks at the top indicate the positions of the positively selected sites.

and the Murids (Mus and Rattus) and some evolved within Peromyscus. In *P. aztecus*, there are numerous frameshifts and stop codons; the predicted protein product would be only 66 aa long (full-length protein is 908 aa in Mus), and thus unlikely to be functional. We therefore excluded the *P. aztecus* sequence from further analysis. The remaining species have premature stops 180 (*P. polionotus*) or 62 (*P. californicus*, *P. eremicus*, *P. leucopus*, and *P. maniculatus*) codons upstream of the Mus stop codon. PHF8 has an exceptionally high proportion of variable amino acid sites (26%) relative to the other proteins (Table 3), and ω is relatively high within Peromyscus (0.47) but there is no evidence for positive selection. Thus, high ω in *Phf8* may result from relaxed purifying selection rather than positive selection.

The gene with the highest ω-value, *Gm1276*, contains a large repeat region that varies in length among Peromyscus species from 252–360 aa (Figure 3). Homologous repeat regions are present in all available mammalian sequences. In Mus and Rattus, the regions are similar in

length to the longer sequences found in Peromyscus species (360 and 330 aa, respectively), but the region is much shorter (≤125 aa) in other mammals.

Smcp has a repeat region that varies both in length and in motif sequence in diverse mammals, including *P. maniculatus*, Mus, and Rattus (Hawthorne *et al.* 2006). Length variation among Peromyscus species is limited; there are a few amino acid indels within repeats, but all six species have the same number of repeats. For *Smcp*, estimates of pairwise ω based on final alignments are low, and there is no evidence for positive selection in Peromyscus either in terms of amino acid sequence or length variation (Tables 2 and 3).

Rapid evolution of testis genes can be lineage specific: Most evidence for the rapid evolution of reproductive proteins in mammals rests on comparisons of highly divergent taxa. We tested if genes with evidence for positive selection in Peromyscus (*Ddc8*, *Gsg1*, *Lrrc50*, *Acr*, and *Gm1276*) also show evidence for positive selection in comparisons among divergent mammals. For each of these five genes, we identified homologs in 5–10 additional mammals. For *Ddc8*, *Gsg1*, and *Lrrc50*, there is no evidence for positive selection in higher-level comparisons (see supplemental Table S4). Both the M8 *vs.* M7 and M8A *vs.* M8 comparisons identify a subset of sites in *Acr* that are positively selected ($P \leq 0.01$; $p_s = 0.07$, $\omega_s = 1.82$); however, the M2 *vs.* M1 comparison is not significant ($P = 0.22$). Such mixed results are similar to previous analysis of *Acr* sequences from a smaller sample of mammals (Swanson *et al.* 2003). For *Gm1276*, all model comparisons provide strong evidence ($P < 0.0001$) for positive selection (M8: $p_s = 0.25$, $\omega_s = 2.12$). The 13 sites with high probability (>0.9) of being in the positively selected class are distributed along the length of the protein, similar to the pattern we observe within Peromyscus. One site (142) is positively selected at both taxonomic levels.

DISCUSSION

Here, we identify a functionally diverse set of genes that are evolving rapidly in rodents, most of which have not been identified previously as targets of selection or functionally characterized. Evolutionary analysis of the same genes at different taxonomic depths often yields different patterns; some genes have evidence for positive selection across divergent mammalian taxa whereas rapid evolution of other genes is lineage specific. These results show that there is not a common set of targets of positive selection among mammalian reproductive proteins, and instead studies of gametic isolation will require species-specific examination. In addition, we find evidence for positive selection acting on five genes within Peromyscus, raising the possibility that these genes may contribute to reduced fertilization potential between these closely related species. This study contributes to a growing body of evidence documenting a remarkable pattern of rapid evolution of reproductive proteins in animals.

Functional roles of rapidly evolving genes: Identifying the functions of rapidly evolving genes may reveal whether selection targets a particular biological process or fertilization step which in turn may suggest evolutionary forces (*e.g.*, sperm competition, sexual conflict, pathogen defense) responsible for the rapid evolution of testis-expressed proteins. We find that extracellular proteins are evolving more rapidly, on average, than intracellular proteins (Table 1), suggesting that interaction with foreign proteins (such as female reproductive proteins or proteins from other males) may result in stronger selection. Some of the most rapidly evolving genes in our analysis have well-described roles in fertilization (Table 2). One gene, *Hils1*, is involved in DNA condensation during spermatogenesis, specifically repackaging DNA onto testis-specific histones to produce the densely packed chromatin of sperm. Evidence for positive selection has been reported previously for three other DNA packaging sperm proteins (*Prm1*, *Prm2*, *Tnp2*; Retief *et al.* 1993; Queralt *et al.* 1995; Wyckoff *et al.* 2000; Torgerson *et al.* 2002; but see Clark and Civetta 2000; Rooney *et al.* 2000), suggesting the process of DNA condensation may be a common target of selection.

Two other genes, *Acr* and *Spa17*, are zona pellucida (egg coat) binding proteins. Much research effort has focused on proteins involved in egg and sperm binding, as this interaction is critical to species-specific fertilization (Wassarman *et al.* 2001). Numerous gamete-binding proteins from both egg and sperm evolve rapidly and show evidence for positive selection at various levels of divergence in mammals (Swanson *et al.* 2001b; Jansa *et al.* 2003; Swanson *et al.* 2003; Glassey and Civetta 2004; Good and Nachman 2005; Gasper and Swanson 2006; Podlaha *et al.* 2006; Turner and Hoekstra 2006, 2008a; Hamm *et al.* 2007). *Acr* is of particular interest because knockouts have partially infertile phenotypes in Mus; homozygous males never sire offspring in competitive mating trials with wild-type males (Jansen *et al.* 2001).

In addition to these well-characterized genes, many of the most rapidly evolving genes identified here have gene ontology annotations. For these, we have an indication of the general function of the gene but not its involvement in a specific fertilization step. This group includes genes encoding proteins involved in signal transduction (*Gm1276* and *4930596D02Rik*) and protein binding (*Lrrc50* and *Phf8*). The functional diversity among rapidly evolving genes identified here is consistent with a study that compared rates of evolution of genes expressed at different stages of spermatogenesis in the mouse, which showed that rates of evolution are higher for genes expressed in late stages of spermatogenesis that serve a wide variety of functions (Good and Nachman 2005). These patterns in rodents are also

consistent with previous studies demonstrating adaptive evolution of sperm proteins with diverse functions across more distantly related mammals (Torgerson *et al.* 2002; Swanson *et al.* 2003).

We also identified several rapidly evolving genes with unknown function, some of which have signatures of positive selection at multiple levels of taxonomic divergence. This result underscores the importance of combining analysis of proteins that have well-described function with genomic approaches that facilitate identification of novel targets. Moreover, evolutionary analyses provide valuable data about the molecular processes of reproduction; for example, some of the positively selected genes identified here may play important roles in fertilization.

Targets of selection at various levels of divergence: A large number of proteins participate in the processes of mammalian fertilization, and there is considerable functional redundancy between proteins, particularly in males (Tanphaichitr *et al.* 2007). Even in the unlikely case that there is one predominant form of selection acting on one particular step of fertilization, it is very possible that the individual proteins targeted by selection differ between species.

Since the majority of studies that have demonstrated positive selection on mammalian reproductive proteins have sampled divergent species, it is clear that there are some common targets. Further, some of the genes identified in these analyses have subsequently been found to be subject to positive selection in more closely related taxa. For example, egg coat proteins under selection across divergent mammals (Swanson *et al.* 2001b) are evolving adaptively within Murines (Jansa *et al.* 2003; Swann *et al.* 2007) and within Peromyscus (Turner and Hoekstra 2006, 2008a). In addition, almost all rapidly evolving seminal proteins identified through a comparison of human and chimpanzee sequences have evidence for positive selection when sequenced in a more diverse sample of primate species (Clark and Swanson 2005). In this study, however, evidence for rapid evolution and positive selection for some testis proteins is limited when we investigate evolutionary patterns at three taxonomic levels: within a genus (Peromyscus), within a superfamily (Muroidea), and within an order (Mammalia). For example, some genes with high rates of evolution between Peromyscus and Mus have been identified previously as targets of selection in more divergent mammals (*e.g.*, *Spa17*, Swanson *et al.* 2003; *Hils1*, Good and Nachman 2005), but have no evidence for positive selection within Peromyscus. The inverse pattern, rapid evolution in closely related taxa but not divergent taxa, was also evident; three (*Ddc8*, *Gsg1*, and *Lrrc50*) of five genes that are rapidly evolving in Muroidea and positively selected within Peromyscus have no evidence for positive selection among diverse mammals. Variation in evolutionary pattern across taxonomic levels might result from variation in the selective agent between taxa, differences in levels of redundancy of genes serving different functions during the fertilization processes, or different degrees of pleiotropy of genes with shared function. Clearly, however, studies that wish to link rapid reproductive protein evolution to gametic isolation will have to examine patterns of nucleotide variation between closely related species as well as test the functional effects of amino acid or expression differences on a case-by-case basis.

Success of EST screen at identifying positively selected genes in Peromyscus: The evolutionary EST analysis we employed here determines evolutionary rates between a single species of Peromyscus and two other rodents. High values of ω between Peromyscus and Mus or Rattus, however, may result from rapid evolution solely within Murids or preceding the diversification of the Peromyscus genus. Approximately half of the sequenced genes chosen on the basis of high pairwise ω in comparisons of Peromyscus *vs.* Mus show evidence for positive selection within Peromyscus. Moreover, four out of five positively selected genes have not been identified previously as targets of selection in mammals. Therefore, this approach is a promising one for identifying new genes likely to be rapidly evolving in taxa without sequenced genomes.

Since close to half of the genes identified in the EST screen that we sequenced in multiple Peromyscus species have evidence for positive selection, it is likely that there are additional targets of selection among the remaining genes with high rates of evolution. Further analysis of rapidly evolving genes among closely related species of Mus and Rattus will likely yield similar success in identifying targets of selection within those genera. However, as selection acts on a small proportion of amino acid sites in many genes, choosing genes on the basis of ω-values averaged across large regions certainly will miss important targets (Hughes 2007). EST analysis and other genomic approaches are complementary to choosing genes on the basis of knowledge of their biological functions.

Adaptive change in amino acid sequence and length of testis proteins: Detailed characterization of 11 testis proteins within Peromyscus allowed us to extend our analysis of the functional targets and lineage specificity of positive selection to the amino acid level. For two proteins that are subject to positive selection in Peromyscus, proacrosin and Gm1276, functional information is available, even about specific domains. Further, both of these proteins have evidence for positive selection among divergent mammals, allowing us to compare the specific targets of selection among taxa.

In Peromyscus, we identified a cluster of selection target sites in the C terminus of proacrosin (Figure 3). Evidence for the function of this region differs between species. In boar and human, this region is implicated in binding to the ZP (Mori *et al.* 1995; Furlong *et al.* 2005), but in the mouse there is no evidence of the C

terminus binding to ZP2 (Howes *et al.* 2001). In boar, the C terminus is cleaved during processing of proacrosin to the mature proteolytic form following ZP binding (Mori *et al.* 1995), thus this region is unlikely to have a role in dissolution of the ZP. Although we identified positively selected sites both in a diverse sample of 11 mammals and within Peromyscus, these sites differed. However, two sites in the positive selection class (45 and 48) in mammals (supplemental Table S4) neighbor one of the positively selected sites in Peromyscus (46). These three sites are ~20 aa upstream of one of the two regions implicated in zona pellucida binding in Mus (Jansen *et al.* 2001). The other three positively selected sites in Peromyscus are clustered in the C-terminal region, which unfortunately cannot be aligned reliably between divergent species.

The human homolog of Gm1276 is MS4A13 (also known as NYD-SP21). This protein is a member of the membrane-spanning four-domain (MS4A) family, which is part of the CD20/β subunit of high affinity IgE receptor superfamily (Ishibashi *et al.* 2001). These plasma membrane-bound proteins interact with other cell-surface proteins in oligomeric complexes that have signal transduction functions in a variety of tissues. Gm1276/MS4A13 has highly specific expression in testis in both mouse and human. All 8 positively selected sites in Peromyscus are located in the extracellular C terminus (Figure 3). In the analysis of Gm1276 sequences from eight divergent mammals, we identified 13 positively selected sites including one site (101) in the intracellular loop between TM domains 2 and 3, another site (113) in TM domain 3, and 11 sites in the C-terminal extracellular region. One of these C-terminal sites (142) is positively selected both in this analysis and within Peromyscus. As MS4A13 is a putative signaling protein with receptor activity, we can speculate that substitutions (and/or length variation) in the large extracellular domain might affect ligand binding.

A comparison of mouse–human orthologous pairs showed that sperm-specific proteins have exceptionally high rates of evolution, both amino acid substitution rate and variation in protein length (Torgerson *et al.* 2002), suggesting both may be responses to selection. We found that Gm1276 has evolved rapidly in protein length as well as amino acid substitution rate, both in a phylogenetically diverse sample of mammals and within Peromyscus. Length variation results from the expansion and contraction of a large repeat region. This repeat region is in the C-terminal extracellular region, downstream of the region homologous to other MS4A family members; querying the InterPro database with the repeat sequence failed to identify any homologous protein domain. This region expanded some time between the divergence of Muroids from other mammalian lineages and the time of the divergence of Cricetids (including Peromyscus) from Murids (including Mus and Rattus). In addition, this repeat varies in length by >100 amino acids within Peromyscus. These results suggest that this gene may consistently respond to selection through two different mechanisms of sequence evolution; however, although there is evidence that positive selection promotes amino acid substitution in this protein, additional functional data are required to evaluate whether length variation is the result of selection or reduced constraint.

For both proacrosin and Gm1276, patterns of amino acid change within Peromyscus and across divergent mammals show that selection has repeatedly targeted the same protein regions and sometimes the same amino acid sites. Taken together, the results from our genome-wide and gene-level analyses reveal that selection on testis proteins ranges widely in scope—from effects seen in a single lineage to effects common across lineages, even those that have been diverging for >100 million years.

Implications for fertilization and reproductive isolation: Ultimately, we are interested in finding genes that cause or maintain reproductive isolation between species. In marine invertebrates, through a combination of detailed analysis of evolutionary patterns within and between recently diverged species and functional characterization of positively selected genes, great progress has been made in identifying the selective forces promoting divergence of sperm proteins (Geyer and Palumbi 2003; Levitan and Ferrell 2006; Riginos *et al.* 2006) and determining the consequences of protein divergence on fertilization potential between species (Lyon and Vacquier 1999; Palumbi 1999; Levitan and Ferrell 2006). In mammals, however, a detailed understanding of the causes and consequences of the rapid divergence of reproductive proteins remains elusive (Turner and Hoekstra 2008b). Progress toward this goal requires the identification and comparison of evolutionary dynamics of these proteins across a range of taxonomic levels as well as experimental assessment of the influence of allelic variation on fertilization success in natural populations with incomplete or recently evolved isolating barriers (Coyne and Orr 2004). The timescale of change in reproductive proteins relative to other factors (*e.g.*, ecological specialization, postzygotic isolation) promoting divergence determines whether reproductive genes may be "speciation genes." Here, we have successfully identified five testis protein genes that are evolving rapidly in Peromyscus and potentially play a role in reducing fertilization success between diverging species. Among these, *Acr* and *Gm1276* are strong candidates for intraspecific and functional analysis to identify specific selective forces driving rapid divergence of male reproductive proteins and to assess their contributions to reproductive isolation. Moreover, Peromyscus subspecies and species pairs with evidence for reduced fertility can be crossed in the laboratory, providing the opportunity to investigate the effects of allelic variation in these proteins *in vivo*.

We thank members of the Hoekstra Laboratory, J. Kohn, and V. Vacquier for helpful discussion, H. Fisher and two anonymous reviewers for useful comments on the manuscript, and J. Weston Glenn and M. Dewey for early access to the *P. maniculatus* EST sequences. The Peromyscus Genetic Stock Center (PGSC) kindly provided tissue samples. This research was supported by a Howard Hughes Medical Institute Predoctoral fellowship and National Science Foundation doctoral dissertation improvement grant DEB 0608030 (to L.M.T.), National Science Foundation grant DEB 0344710 (to H.E.H.).

LITERATURE CITED

Andres, J. A., L. S. Maroja, S. M. Bogdanowicz, W. J. Swanson and R. G. Harrison, 2006 Molecular evolution of seminal proteins in field crickets. Mol. Biol. Evol. **23:** 1574–1584.

Bendtsen, J. D., H. Nielsen, G. von Heijne and S. Brunak, 2004 Improved prediction of signal peptides: SignalP 3.0. J. Mol. Biol. **340:** 783–795.

Bradley, R. D., N. D. Durish, D. S. Rogers, J. R. Miller, M. D. Engstrom *et al.*, 2007 Toward a molecular phylogeny for Peromyscus: evidence from mitochondrial cytochrome-b sequences. J. Mammal. **88:** 1146–1159.

Braswell, W. E., J. A. Andres, L. S. Maroja, R. G. Harrison, D. J. Howard *et al.*, 2006 Identification and comparative analysis of accessory gland proteins in Orthoptera. Genome **49:** 1069–1080.

Castillo-Davis, C. I., D. L. Hartl and G. Achaz, 2004 Cis-regulatory and protein evolution in orthologous and duplicate genes. Genome Res. **14:** 1530–1536.

Cheng, G. Y., J. L. Shi, M. Wang, Y. Q. Hu, C. M. Liu *et al.*, 2007 Inhibition of mouse acrosome reaction and sperm-zona pellucida binding by anti-human sperm membrane protein 1 antibody. Asian J. Androl. **9:** 23–29.

Cho, R. J., and M. J. Campbell, 2000 Transcription, genomes, function. Trends Genet. **16:** 409–415.

Clark, A. G., and A. Civetta, 2000 Evolutionary biology: protamine wars. Nature **403:** 261–263.

Clark, N. L., J. E. Aagaard and W. J. Swanson, 2006 Evolution of reproductive proteins from animals and plants. Reproduction **131:** 11–22.

Clark, N. L., and W. J. Swanson, 2005 Pervasive adaptive evolution in primate seminal proteins. PLoS Genet. **1:** 335–342.

Coyne, J. A., and H. A. Orr, 2004 *Speciation*. Sinauer Associates, Sunderland, MA.

Dewey, M. J., and W. D. Dawson, 2001 Deer mice: "The Drosophila of North American mammalogy." Genesis **29:** 105–109.

Dice, L. R., 1933 Fertility relationships between some of the species and subspecies of mice in the genus Peromyscus. J. Mammal. **14:** 298–305.

Duret, L., and D. Mouchiroud, 2000 Determinants of substitution rates in mammalian genes: expression pattern affects selection intensity but not mutation rate. Mol. Biol. Evol. **17:** 68–74.

Edgar, R. C., 2004 MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. **32:** 1792–1797.

Ewing, B., and P. Green, 1998 Base-calling of automated sequencer traces using Phred. II. Error probabilities. Genome Res. **8:** 186–194.

Ewing, B., L. Hillier, M. C. Wendl and P. Green, 1998 Base-calling of automated sequencer traces using Phred. I. Accuracy assessment. Genome Res. **8:** 175–185.

Furlong, L. I., J. D. Harris and M. H. Vazquez-Levin, 2005 Binding of recombinant human proacrosin/acrosin to zona pellucida (ZP) glycoproteins. I. Studies with recombinant human ZPA, ZPB, and ZPC. Fertil. Steril. **83:** 1780–1790.

Gasper, J., and W. J. Swanson, 2006 Molecular population genetics of the gene encoding the human fertilization protein zonadhesin reveals rapid adaptive evolution. Am. J. Hum. Genet. **79:** 820–830.

Geyer, L. B., and S. R. Palumbi, 2003 Reproductive character displacement and the genetics of gamete recognition in tropical sea urchins. Evolution **57:** 1049–1060.

Gibbs, R. A., G. M. Weinstock, M. L. Metzker, D. M. Muzny, E. J. Sodergren *et al.*, 2004 Genome sequence of the Brown Norway rat yields insights into mammalian evolution. Nature **428:** 493–521.

Glassey, B., and A. Civetta, 2004 Positive selection at reproductive ADAM genes with potential intercellular binding activity. Mol. Biol. Evol. **21:** 851–859.

Glenn, J., C.-F. Chen, A. Lewandowski, C.-H. Cheng, C. Ramsdell *et al.*, 2008 Expressed sequence tags from Peromyscus testis and placenta tissue: analysis, annotation, and utility for mapping. BMC Genomics **9:** 300.

Good, J. M., and M. W. Nachman, 2005 Rates of protein evolution are positively correlated with developmental timing of expression during mouse spermatogenesis. Mol. Biol. Evol. **22:** 1044–1052.

Guttman, J. A., Y. Takai and A. W. Vogl, 2004 Evidence that tubulobulbar complexes in the seminiferous epithelium are involved with internalization of adhesion junctions. Biol. Reprod. **71:** 548–559.

Hamm, D., B. S. Mautz, M. F. Wolfner, C. F. Aquadro and W. J. Swanson, 2007 Evidence of amino acid diversity–enhancing selection within humans and among primates at the candidate sperm-receptor gene *PKDREJ*. Am. J. Hum. Genet. **81:** 44–52.

Hawthorne, S. K., G. Goodarzi, J. Bagarova, K. E. Gallant, R. R. Busanelli *et al.*, 2006 Comparative genomics of the sperm mitochondria-associated cysteine-rich protein gene. Genomics **87:** 382–391.

Hoekstra, H. E., K. E. Drumm and M. W. Nachman, 2004 Ecological genetics of adaptive color polymorphism in pocket mice: geographic variation in selected and neutral genes. Evolution **58:** 1329–1341.

Howes, E., J. C. Pascall, W. Engel and R. Jones, 2001 Interactions between mouse ZP2 glycoprotein and proacrosin; a mechanism for secondary binding of sperm to the zona pellucida during fertilization. J. Cell Sci. **114:** 4127–4136.

Huang, X. Q., and A. Madan, 1999 CAP3: A DNA sequence assembly program. Genome Res. **9:** 868–877.

Huelsenbeck, J. P., and F. Ronquist, 2001 MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics **17:** 754–755.

Hughes, A. L., 2007 Looking for Darwin in all the wrong places: the misguided quest for positive selection at the nucleotide sequence level. Heredity **99:** 364–373.

Ishibashi, K., M. Suzuki, S. Sasaki and M. Imai, 2001 Identification of a new multigene four-transmembrane family (MS4A) related to CD20, HTm4 and beta subunit of the high-affinity IgE receptor. Gene **264:** 87–93.

Jansa, S. A., B. L. Lundrigan and P. K. Tucker, 2003 Tests for positive selection on immune and reproductive genes in closely related species of the murine genus Mus. J. Mol. Evol. **56:** 294–307.

Jansen, S., M. Ekhlasi-Hundrieser and E. Toepfer-Petersen, 2001 Sperm adhesion molecules: structure and function. Cells Tissues Organs **168:** 82–92.

Jeong, B. C., C. Y. Hong, S. Chattopadhyay, J. H. Park, E. Y. Gong *et al.*, 2004 Androgen receptor corepressor-19 kDa (ARR19), a leucine-rich protein that represses the transcriptional activity of androgen receptor through recruitment of histone deacetylase. Mol. Endocrinol. **18:** 13–25.

Kanno, K., M. K. Wu, E. F. Scapa, S. L. Roderick and D. E. Cohen, 2007 Structure and function of phosphatidylcholine transfer protein (PC-TP)/StarD2. Biochim. Biophys. Acta-Mol. Cell Biol. Lipids **1771:** 654–662.

Kleiman, D. G., 1977 Monogamy in mammals. Q. Rev. Biol. **52:** 39–69.

Krogh, A., B. Larsson, G. von Heijne and E. L. L. Sonnhammer, 2001 Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J. Mol. Biol. **305:** 567–580.

Kumar, S., K. Tamura, I. B. Jakobsen and M. Nei, 2001 MEGA2: molecular evolutionary genetics analysis software. Bioinformatics **17:** 1244–1245.

Levitan, D. R., and D. L. Ferrell, 2006 Selection on gamete recognition proteins depends on sex, density, and genotype frequency. Science **312:** 267–269.

Li, Y. C., X. Q. Hu, K. Y. Zhang, H. Guo, Z. Y. Hu *et al.*, 2006 Afaf, a novel vesicle membrane protein, is related to acrosome formation in murine testis. FEBS Lett. **580:** 4266–4273.

Liu, J., J. Xia, K.-H. Cho, D. E. Clapham and D. Ren, 2007 CatSperbeta, a novel transmembrane protein in the CatSper channel complex. J. Biol. Chem. **282:** 18945–18952.

Liu, T. T., 1953 The measurement of fertility and its use as an index of reproductive isolation among certain laboratory stocks of Peromyscus. Contrib. Lab. Vertebr. Biol. **59:** 1–12.

Lyon, J. D., and V. D. Vacquier, 1999 Interspecies chimeric sperm lysins identify regions mediating species-specific recognition of the abalone egg vitelline envelope. Dev. Biol. **214:** 151–159.

Maddock, M. B., and W. D. Dawson, 1974 Artificial insemination of deermice (*Peromyscus maniculatus*) with sperm from other rodent species. J. Embryol. Exp. Morphol. **31:** 621–634.

Martianov, I., S. Brancorsini, R. Catena, A. Gansmuller, N. Kotaja *et al.*, 2005 Polar nuclear localization of H1T2, a histone H1 variant, required for spermatid elongation and DNA condensation during spermiogenesis. Proc. Natl. Acad. Sci. USA **102:** 2808–2813.

Mori, E., S. Kashiwabara, T. Baba, Y. Inagaki and T. Mori, 1995 Amino acid sequences of porcine Sp38 and proacrosin required for binding to the zona pellucida. Dev. Biol. **168:** 575–583.

Mulder, N. J., R. Apweiler, T. K. Attwood, A. Bairoch, A. Bateman *et al.*, 2007 New developments in the InterPro database. Nucleic Acids Res. **35:** D224–D228.

Nayernia, K., I. M. Adham, E. Burkhardt-Gottges, J. Neesen, M. Rieche *et al.*, 2002 Asthenozoospermia in mice with targeted deletion of the sperm mitochondrion-associated cysteine-rich protein (*Smcp*) gene. Mol.Cell. Biol. **22:** 3046–3052.

Nielsen, H., and A. Krogh, 1998 Prediction of signal peptides and signal anchors by a hidden Markov model. Proc. Int. Conf. Intell. Syst. Mol. Biol. **6:** 122–130.

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.

Nixon, B., R. J. Aitken and E. A. McLaughlin, 2007 New insights into the molecular mechanisms of sperm-egg interaction. Cell. Mol. Life Sci. **64:** 1805–1823.

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.

Pearson, W. R., 1990 Rapid and sensitive sequence comparison with FASTP and FASTA. Methods Enzymol. **183:** 63–98.

Podlaha, O., and J. Zhang, 2003 Positive selection on protein-length in the evolution of a primate sperm ion channel. Proc. Natl. Acad. Sci. USA **100:** 12241–12246.

Podlaha, O., D. M. Webb, P. K. Tucker and J. Zhang, 2005 Positive selection for indel substitutions in the rodent sperm protein CATSPER1. Mol. Biol. Evol. **22:** 1845–1852.

Podlaha, O., D. M. Webb and J. Zhang, 2006 Accelerated evolution and loss of a domain of the sperm-egg-binding protein SED1 in ancestral primates. Mol. Biol. Evol. **23:** 1828–1831.

Posada, D., and K. A. Crandall, 1998 MODELTEST: testing the model of DNA substitution. Bioinformatics **14:** 817–818.

Queralt, R., R. Adroer, R. Oliva, R. J. Winkfein, J. D. Retief *et al.*, 1995 Evolution of *protamine P1* genes in mammals. J. Mol. Evol. **40:** 601–607.

Retief, J. D., R. J. Winkfein, G. H. Dixon, R. Adroer, R. Queralt *et al.*, 1993 Evolution of *protamine P1* genes in primates. J. Mol. Evol. **37:** 426–434.

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.

Riginos, C., D. Wang and A. J. Abrams, 2006 Geographic variation and positive selection on M7 lysin, an acrosomal sperm protein in mussels (*Mytilus* spp.). Mol. Biol. Evol. **23:** 1952–1965.

Rooney, A. P., J. Zhang and M. Nei, 2000 An unusual form of purifying selection in a sperm protein. Mol. Biol. Evol. **17:** 278–283.

Singh, R. S., and R. J. Kulathinal, 2000 Sex gene pool evolution and speciation: a new paradigm. Genes Genet. Syst. **75:** 119–130.

Sonnhammer, E. L. L., G. von Heijne and A. Krogh, 1998 A hidden Markov model for predicting transmembrane helices in protein sequences. Proc. Int. Conf. Intell. Syst. Mol. Biol. **6:** 175–182.

Stajich, J. E., D. Block, K. Boulez, S. E. Brenner, S. A. Chervitz *et al.*, 2002 The bioperl toolkit: Perl modules for the life sciences. Genome Res. **12:** 1611–1618.

Steppan, S., R. Adkins and J. Anderson, 2004 Phylogeny and divergence-date estimates of rapid radiations in muroid rodents based on multiple nuclear genes. Syst. Biol. **53:** 533–553.

Swann, C. A., S. J. B. Cooper and W. G. Breed, 2007 Molecular evolution of the carboxy terminal region of the zona pellucida 3 glycoprotein in murine rodents. Reproduction **133:** 697–708.

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

Swanson, W. J., A. G. Clark, H. M. Waldrip-Dail, M. F. Wolfner and C. F. Aquadro, 2001a Evolutionary EST analysis identifies rapidly evolving male reproductive proteins in Drosophila. Proc. Natl. Acad. Sci. USA **98:** 7375–7379.

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

Swanson, W. J., R. Nielsen and Q. F. Yang, 2003 Pervasive adaptive evolution in mammalian fertilization proteins. Mol. Biol. Evol. **20:** 18–20.

Swanson, W. J., A. Wong, M. F. Wolfner and C. F. Aquadro, 2004 Evolutionary expressed sequence tag analysis of Drosophila female reproductive tracts identifies genes subjected to positive selection. Genetics **168:** 1457–1465.

Swofford, D. L., 2002 *PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods)*, Version 4. Sinauer Associates, Sunderland, MA.

Tanaka, H., N. Iguchi, A. Isotani, K. Kitamura, Y. Toyama *et al.*, 2005 HANP1/H1T2, a novel histone H1-like protein involved in nuclear formation and sperm fertility. Mol. Cell. Biol. **25:** 7107–7119.

Tanphaichitr, N., E. Carmona, M. B. Khalil, H. B. Xu, T. Berger *et al.*, 2007 New insights into sperm-zona pellucida interaction: involvement of sperm lipid rafts. Front. Biosci. **12:** 1748–1766.

Thomas, P. D., A. Kejariwal, M. J. Campbell, H. Y. Mi, K. Diemer *et al.*, 2003 PANTHER: a browsable database of gene products organized by biological function, using curated protein family and subfamily classification. Nucleic Acids Res. **31:** 334–341.

Torgerson, D. G., R. J. Kulathinal and R. S. Singh, 2002 Mammalian sperm proteins are rapidly evolving: evidence of positive selection in functionally diverse genes. Mol. Biol. Evol. **19:** 1973–1980.

Turner, L. M., and H. E. Hoekstra, 2006 Adaptive evolution of fertilization proteins within a genus: variation in ZP2 and ZP3 in deer mice (Peromyscus). Mol. Biol. Evol. **23:** 1656–1669.

Turner, L. M., and H. E. Hoekstra, 2008a Reproductive protein evolution within and between species: maintenance of divergent ZP3 alleles in Peromyscus. Mol. Ecol. **17:** 2616–2628.

Turner, L. M., and H. E. Hoekstra, 2008b The evolution of reproductive proteins: causes and consequences. Int. J. Dev. Biol. **52:** 769–780.

Walden, P. D., and N. J. Cowan, 1993 A novel 205-kilodalton testis-specific serine threonine protein-kinase associated with microtubules of the spermatid manchette. Mol. Cell. Biol. **13:** 7625–7635.

Wassarman, P. M., L. Jovine and E. S. Litscher, 2001 A profile of fertilization in mammals. Nat. Cell Biol. **3:** E59–E64.

Waterston, R. H., K. Lindblad-Toh, E. Birney, J. Rogers, J. F. Abril *et al.*, 2002 Initial sequencing and comparative analysis of the mouse genome. Nature **420:** 520–562.

Winter, E. E., L. Goodstadt and C. P. Ponting, 2004 Elevated rates of protein secretion, evolution, and disease among tissue-specific genes. Genome Res. **14:** 54–61.

Wolff, J. O., 1989 Social behavior, pp. 271–291 in *Advances in the Study of Peromyscus (Rodentia)*, edited by G. L. Kirkland and J. N. Layne. Texas Tech University Press, Lubbock, TX.

Wu, Z. J., and R. A. Irizarry, 2005 Stochastic models inspired by hybridization theory for short oligonucleotide arrays. J. Comput. Biol. **12:** 882–893.

Wyckoff, G. J., W. Wang and C. I. Wu, 2000 Rapid evolution of male reproductive genes in the descent of man. Nature **403:** 304–309.

Yan, W., L. Ma, K. H. Burns and M. M. Matzuk, 2003 HILS1 is a spermatid-specific linker histone H1-like protein implicated in chromatin remodeling during mammalian spermiogenesis. Proc. Natl. Acad. Sci. USA **100:** 10546–10551.

Yang, Z., 2000 Phylogenetic analysis by maximum likelihood (PAML). University College, London.

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

Yang, Z., W. S. W. Wong and R. Nielsen, 2005 Bayes empirical Bayes inference of amino acid sites under positive selection. Mol. Biol. Evol. **22:** 1107–1118.

Zdobnov, E. M., and R. Apweiler, 2001 InterProScan: an integration platform for the signature-recognition methods in InterPro. Bioinformatics **17:** 847–848.

Communicating editor: D. M. Rand