

# Sexual Selection and the Adaptive Evolution of Mammalian Ejaculate Proteins

Steven A. Ramm,<sup>\*1</sup> Peter L. Oliver,<sup>†1</sup> Chris P. Ponting,<sup>†</sup> Paula Stockley,<sup>\*</sup> and Richard D. Emes<sup>‡2</sup>

<sup>\*</sup>Mammalian Behaviour and Evolution Group, Faculty of Veterinary Science, University of Liverpool, Liverpool, United Kingdom; <sup>†</sup>MRC Functional Genetics Unit, Department of Physiology, Anatomy and Genetics, University of Oxford, Oxford, United Kingdom; and <sup>‡</sup>Department of Biology, University College London, London, United Kingdom

An elevated rate of substitution characterizes the molecular evolution of reproductive proteins from a wide range of taxa. Although the selective pressures explaining this rapid evolution are yet to be resolved, recent evidence implicates sexual selection as a potentially important explanatory factor. To investigate this hypothesis, we sought evidence of a high rate of adaptive gene evolution linked to postcopulatory sexual selection in muroid rodents, a model vertebrate group displaying a broad range of mating systems. Specifically, we sequenced 7 genes from diverse rodents that are expressed in the testes, prostate, or seminal vesicles, products of which have the potential to act in sperm competition. We inferred positive Darwinian selection in these genes by estimation of the ratio of nonsynonymous ( $d_N$ , amino acid changing) to synonymous ( $d_S$ , amino acid retaining) substitution rates ( $\omega = d_N/d_S$ ). Next, we tested whether variation in this ratio among lineages could be attributed to interspecific variation in mating systems, as inferred from the variation in these rodents' relative testis sizes (RTS). Four of the 7 genes examined (*Prml1*, *Sva*, *Acrv1*, and *Svs2*, but not *Svp2*, *Msmb*, or *Spink3*) exhibit unambiguous evidence of positive selection. One of these, the seminal vesicle-derived protein *Svs2*, also shows some evidence for a concentration of positive selection in those lineages in which sperm competition is common. However, this was not a general trend among all the rodent genes we examined. Using the same methods, we then reanalyzed previously published data on 2 primate genes, *SEMG1* and *SEMG2*. Although *SEMG2* also shows evidence of positive selection concentrated in lineages subject to high levels of sperm competition, no such trend was found for *SEMG1*. Overall, despite a high rate of positive selection being a feature of many ejaculate proteins, these results indicate that the action of sexual selection potentially responsible for elevated evolutionary rates may be difficult to detect on a gene-by-gene basis. Although the extreme diversity of reproductive phenotypes exhibited in nature attests to the power of sexual selection, the extent to which this force predominates in driving the rapid molecular evolution of reproductive genes therefore remains to be determined.

## Introduction

Reproduction-related genes are known to have evolved unusually rapidly across diverse animal groups (Swanson and Vacquier 2002; Emes et al. 2003; Bustamante et al. 2005; Richards et al. 2005; Clark et al. 2006). Of several factors suggested as major influences on the evolutionary rate of reproductive genes (Swanson and Vacquier 2002), those hypothesized to contribute most involve continual selective pressure deriving from coevolutionary cycles of adaptation and counteradaptation. Such coevolutionary cycles could occur either between or within genomes. Thus, rapid evolution could be driven by host–pathogen coevolution because many of the genes expressed in reproductive tissues possess important immune-related functions (Good and Nachman 2005). Alternatively or additionally, the rapid evolution of reproductive genes may result from episodes of sexual selection and coevolution between the sexes (Swanson and Vacquier 2002). In support of the latter idea, many rapidly evolving genes are sex limited or sex biased in their expression (e.g., Pröschel et al. 2006); for example, male accessory gland-specific proteins found in seminal fluid evolve rapidly for both vertebrate (Clark and Swanson 2005) and invertebrate (Swanson, Clark, et al. 2001; Andrés et al. 2006) taxa.

Despite the potentially pervasive influence of sexual selection in driving adaptation at the molecular level, very few studies to date have reported evidence directly linking

episodes of positive selection to changes in relevant physiological or behavioral traits. So far, these studies have concentrated on examining whether species' differences in evolutionary rates correlate with diversity in mating systems or with other proxy measures of sexual selection intensity (Dorus et al. 2004; Wagstaff and Begun 2005; Turner and Hoekstra 2006; Herlyn and Zischler 2007; Hurle et al. 2007; Nadeau et al. 2007; Wagstaff and Begun 2007). One such study concluded that differences in female promiscuity among primate species are responsible for variation in the rate of evolution of *SEMG2*, which encodes the major protein present in primate seminal fluid (Dorus et al. 2004). However, a subsequent reanalysis of *SEMG2* and related genes based on a larger data set failed to establish support for this conclusion (Hurle et al. 2007). Nevertheless, others have now shown that the evolutionary rate of the primate sperm ligand zonadhesin (*ZAN*) correlates negatively with sexual size dimorphism and that an avian pigmentation gene (*MC1R*) evolves fastest in lineages with greatest sexual dichromatism (Herlyn and Zischler 2007; Nadeau et al. 2007). More broadly, Wagstaff and Begun (2005, 2007) have noted the generally higher rates of *Drosophila* accessory gland protein (*Acp*) gene divergence in *repleta* group versus *melanogaster* subgroup flies, potentially driven by the higher rates of female remating in *repleta* group females. Although these studies clearly highlight sexual selection's potential explanatory power, whether or not variation in sexual selection intensity can generally account for interspecific diversity in evolutionary rates for single genes clearly warrants further investigation.

Here, we address this question by investigating the molecular evolution of several genes expressed in the male reproductive tract of muroid rodents. We seek evidence of adaptive evolution specifically for those lineages that are associated with large relative testis size (RTS). This is

<sup>1</sup> These authors contributed equally to this work.

<sup>2</sup> Present address: Institute for Science and Technology in Medicine, School of Medicine, Keele University, Staffordshire, United Kingdom.

Key words: adaptive evolution, positive selection, primates, rodents, sexual selection, sperm competition.

E-mail: r.d.emes@hfac.keele.ac.uk.

*Mol. Biol. Evol.* 25(1):207–219. 2008

doi:10.1093/molbev/msm242

Advance Access publication November 20, 2007

because relative male investment in testes covaries with the intensity of sperm competition, and RTS is thus a suitable quantifiable proxy for the intensity of postcopulatory sexual selection on male reproductive genes (Parker et al. 1997). As the genes we have studied are implicated in postcopulatory sexual selection, those genes that have evolved adaptively only within lineages associated with large, or else small, extant RTS would represent excellent candidates for genes that determine reproductive phenotypes under sexual selection. The majority of our study focused on rodents. This group is ideal for such investigations because rodents exhibit a wide range of mating systems, and adaptations to postcopulatory sexual selection at behavioral and morphological levels are already well established (Ramm et al. 2005). In addition, the availability of biomedical and genomic data for this clade is unprecedented. Consequently, rodents are increasingly being recognized as an appropriate model group for molecular evolutionary studies (Good and Nachman 2005; Podlaha et al. 2005; Vamathevan et al. 2007).

The 7 rodent genes we examined were chosen on the basis of their male reproductive tract-limited expression and on functional information that is suggestive of roles in mediating postcopulatory sexual selection. Specifically, we examined 5 seminal fluid proteins derived predominantly from either the seminal vesicles (*Svs2*, *Svp2*, and *Sva* gene products) or prostate (serine protease inhibitor kazal-type 3 [*Spink3*] and *Msbm* gene products) and 2 testis-expressed genes whose products are associated with the sperm head (*Acrvl* and *Prml*).

*Svs2* encodes the major component of the rodent copulatory plug, the coagulated mass of protein that forms upon mixing with a prostate-derived transglutaminase at ejaculation and which thereafter occludes the female reproductive tract (see Ramm et al. 2005). *Svs2* belongs to the rapidly evolving seminal vesicle-transcribed (REST) gene family (Lundwall and Lazure 1995), as does the second gene investigated here, *Svp2* (also known as *Svs4*). The REST family also contains the primate semenogelins (*SEMG1* and *SEMG2*) that were the subjects of previous studies (Dorus et al. 2004; Hurlle et al. 2007) and will be discussed further below. Although encoded by paralogous genes, the protein encoded by the rodent *Svp2* gene is much shorter than that encoded by the rodent *Svs2* gene, and it functions differently as an anti-inflammatory and immune-modulating agent in the female reproductive tract (Ialenti et al. 2001). The *Sva* gene product acts directly on sperm, where it functions as a suppression factor regulating capacitation (Huang et al. 2005, 2007).

*Spink3* is expressed in the ventral prostate, anterior prostate, and seminal vesicle (Mills et al. 1987; Needham et al. 1988). The function of its protein product is poorly characterized, but protease inhibitors may be important in preventing sperm degradation and/or interfering with female responses to insemination (Lung et al. 2002; Fedorka and Zuk 2005). Beta-microseminoprotein (encoded by the *Msbm* gene, also known as PSP-94) is expressed solely in the prostate (Xuan et al. 1999) and in pigs has been linked to a role in regulating sperm motility through binding to the midpiece and posterior portion of spermatozoa (Chao et al. 1996). It is also known to be extremely resistant to prote-

olysis when deposited in the female reproductive tract (Xuan et al. 1999).

The testis-specific acrosomal protein Sp10 (encoded by the *Acrvl* gene) is located chiefly on the principal segment and posterior bulb of the equatorial segment of the sperm acrosome, where it may be involved in sperm–zona binding or penetration (Foster et al. 1994). In contrast, the second sperm protein we investigated, encoded by *Prml*, is not directly involved in sperm–egg binding but instead functions in chromatin condensation in the sperm head during spermatogenesis, potentially influencing sperm morphology. Several previous studies investigating the molecular evolution of *Prml*, predominantly in primates (e.g., Rooney and Zhang 1999; Wyckoff et al. 2000; Swanson, Zhang, et al. 2001; Torgerson et al. 2002), have found that a large proportion of the gene may have been subject to positive selection (although see Clark and Civetta 2000).

We sequenced the coding regions of the 7 genes of interest for up to 18 diverse rodents and analyzed these in combination with predicted gene sequences obtained from *Mus musculus* and *Rattus norvegicus* genome assemblies. We used programs from the phylogenetic analyses by maximum likelihood (PAML) (Yang 1997) package to infer the past action of positive selection. Depending on the available evidence, positive selection of specific sites can be predicted either across all lineages or within individual lineages. We first sought evidence of adaptive evolution in these genes across all lineages. To test the hypothesis that postcopulatory sexual selection is the predominant force driving the evolution of these genes, we then sought evidence for rapid evolution of genes restricted to those lineages with large (or else small) RTS, our proxy for sexual selection intensity.

Finally, we applied the same methods to previously published data in order to analyze the molecular evolution of 2 primate ejaculate proteins encoded by *SEMG1* and *SEMG2*. These genes have been the subject of previous studies investigating the evolution of reproductive genes thought to be under sexual selection (Jensen-Seaman and Li 2003; Kingan et al. 2003; Dorus et al. 2004; but see Hurlle et al. 2007).

## Materials and Methods

### Tissue Sources

In addition to genomic data obtained for *M. musculus* and *R. norvegicus*, sequencing of the 7 rodent genes of interest was attempted in a total of 18 species: *Acomys cahirinus*, *Acomys cilicicus*, *Apodemus sylvaticus*, *Arvicola terrestris*, *Clethrionomys glareolus*, *Coelomys pahari*, *Meriones unguiculatus*, *Micromys minutus*, *Microtus agrestis*, *Mus musculus castaneus*, *Mus macedonicus*, *Mus spicilegus*, *Peromyscus californicus*, *Peromyscus leucopus*, *Peromyscus maniculatus*, *Peromyscus polionotus*, *Phodopus roborovski*, and *Phodopus sungorus*. Subject species were selected to maximize diversity in mating systems, as well as to maximize the range of divergence times across the rodent phylogeny (supplementary table S1 and fig. S1, Supplementary Material online). We therefore

included several species from within single genera (e.g., *Mus*, *Peromyscus*), and the 2 most divergent clades (murids and cricetids) split approximately 24 MYA (Steppan et al. 2004). In each case, reproductive tissues were dissected either from fresh or (where necessary) previously frozen cadavers and then frozen prior to further analysis.

### Primer Design and Gene Amplification

Primers for reverse transcriptase–polymerase chain reaction (RT-PCR) were designed to amplify the entire coding region of each gene. Initially, these were chosen to be located in the untranslated regions (UTRs) approximately 100 bp from the translation start and stop codons in regions conserved between the genome sequences of *M. musculus* (C57BL/6J) and *R. norvegicus* (<http://www.ensembl.org/index.html>). When RT-PCR products could not be obtained, sequences from closest related species that had been successfully amplified were used to design additional UTR primers. Primer sequences are available from the authors upon request.

Frozen reproductive tract samples were thawed on ice, and tissue was dissected from the testis, epididymis, vas deferens, and seminal vesicles up to a maximum of 150 mg total wet weight. Total RNA was purified from the combined tissue sample using the RNeasy Midi kit (Qiagen Inc., Valencia, CA); samples were homogenized using a Rotor Star homogenizer (Pro Scientific, Oxford, CT) for 30 s in a 4-ml total volume of buffer RLT as recommended by the manufacturer. In addition, DNase treatment was performed during the purification using the RNase-free DNase set (Qiagen). Reverse transcription used Expand Reverse Transcriptase (Roche, Indianapolis, IN) with 20 ng oligo-d(T) primers and 1  $\mu$ g of total RNA followed by RT-PCR carried out using Expand High-Fidelity polymerase (Roche); the optimum annealing temperatures for each primer pair were determined experimentally, in each case, by adjusting the annealing temperature and/or MgCl<sub>2</sub> concentration as necessary. RT-PCR products were purified on agarose gels using the Qiaquick gel purification kit (Qiagen) prior to sequencing in both directions with the relevant PCR primers. For those products over 1 kb in size or that provided a poor yield, gel-purified products were cloned into the pCR4-TOPO TA-cloning vector (Invitrogen, Carlsbad, CA), and mini-prep DNA was purified from positive clones prior to sequencing using standard M13 forward and reverse primers in addition to gene-specific primers. Sequence traces were initially curated and assembled using Chromas 2.0 (<http://www.technelysium.com.au/chromas.html>) and Bioedit (<http://www.mbio.ncsu.edu/BioEdit/bioedit.html>). Owing to their varying degrees of sequence divergence, which affected the efficacy of the degenerate primers, this approach resulted in between 6 and 18 species' genes being sequenced.

### Identification of Sites Subject to Positive Selection

For many proteins, adaptive evolution is likely to have acted only on a small proportion of amino acid residues, and hence, an average of substitution rates across the gene often does not provide the most informative indicator of positive

selection (Yang and Bielawski 2000). Instead, data are fitted to codon-based substitution models that allow the ratio of nonsynonymous ( $d_N$ ) to synonymous ( $d_S$ ) substitutions ( $d_N/d_S = \omega$ ) to vary among codons, with parameters estimated using a maximum likelihood method (Yang and Bielawski 2000; Bielawski and Yang 2003). Analysis was conducted using the codeml application from PAML package version 3.15 (Yang 1997). For each pair of nested models, the log likelihood values are compared by the likelihood ratio test (LRT). If the model allowing positive selection estimates a  $d_N/d_S$  value of greater than 1 and twice the log likelihood difference is greater than the chi-square critical value for the appropriate degrees of freedom (df), then positive selection is inferred (Yang 1998). Bayesian statistics are then used to identify those codons that have been subject to adaptive evolution. In this study, we used 2 pairs of models: M1 (neutral) versus M2 (selection) (Nielsen and Yang 1998) and M7 (beta) versus M8 (beta +  $\omega$ ) (Yang et al. 2000). M1 allows 2  $d_N/d_S$  site classes with  $\omega_0 < 1$  estimated from the data or  $\omega_1 = 1$ , whereas M2 allows an additional  $\omega_2$  value to be estimated from the data which may be  $>1$ . M7 fits  $\omega$  to 10 site classes between 0 and 1 using a beta distribution, whereas M8 adds an additional site class which may be  $>1$  to be estimated from the data. Both M1–M2 and M7–M8 comparisons were performed using 2 df. All analyses were conducted in triplicate using different initial  $d_N/d_S$  values for the likelihood iterations; the greatest log likelihood value for each model was chosen to conduct the LRT. To ascertain if independent runs reached convergence at a global optimal peak, the difference of greatest and second greatest log likelihood values were constrained to be less than or equal to 0.01.

When an alignment exhibits a signature of adaptive evolution according to the LRTs, the empirical Bayes method (Nielsen and Yang 1998; Wong et al. 2004) was used to identify individual codons with  $d_N/d_S > 1$ . To identify such codons, consistency between different evolutionary models was required; hence, only those sites with posterior probabilities of  $>0.95$  in both nested models were considered to indicate positive selection.

### Identification of Sites and Lineages Subject to Positive Selection

Although the sites analysis described above identifies codons with  $d_N/d_S > 1$ , the method assumes that the estimated parameters are invariant across all lineages of the tree. As an alternative approach to identify episodic periods of adaptive evolution, the branch-site test for positive selection (referred to as Test 2 in Zhang et al. [2005]; Yang and Nielsen 2002) was also used. In this analysis, the phylogenetic tree is partitioned into foreground and background branches, where a site class of  $d_N/d_S > 1$  is restricted to occur only along the foreground branch. In the null hypothesis, sites are partitioned into 4 site categories: class 0, purifying selection ( $0 < \omega_0 < 1$ ); class 1, neutral evolution ( $\omega_1 = 1$ ); class 2a, purifying selection for background branches and neutral evolution ( $\omega_2 = 1$ ) in foreground branches; and class 2b, neutral evolution ( $\omega_2 = 1$ ) along both foreground and background lineages. This model is compared with the alternative, where  $d_N/d_S > 1$  can occur

along the foreground branch ( $\omega_2 > 1$ ). An LRT is conducted between the null and alternative models and compared with a chi-squared distribution to determine significance. To avoid potential violations of model assumptions and small sample sizes, we used the conservative  $\chi_1^2$  distribution to conduct the LRT. As above, the posterior probability of each site residing in the  $d_N/d_S > 1$  class is calculated using the Bayes empirical Bayes method. Each branch is labeled as foreground in turn, and the Bonferroni correction was employed to control the type I error rate (Gordi and Khamis 2004). Specific partitions used to explore evolutionary hypotheses are described below.

### Proxy Measures of Sexual Selection

Our analysis differed from previous analyses in partitioning extant species into those predicted to experience more or less intense postcopulatory sexual selection on male reproductive genes and then testing for a difference between the molecular evolutionary rate of genes belonging to these 2 partitions.

For rodents, we classified species with relatively high levels of sperm competition ( $RTS > 1$ ) as experiencing more intense postcopulatory sexual selection and species with relatively low levels of sperm competition ( $RTS < 1$ ) as experiencing less intense postcopulatory sexual selection on male reproductive genes.  $RTS$  was calculated from observed testes mass/expected testes mass. Expected testes mass ( $E = 0.035 \times \text{body mass}^{0.72}$ ) is based on the regression equation derived by Kenagy and Trombulak (1986) in their survey of mammalian testis size in relation to body size. Data were obtained either from published sources (Kenagy and Trombulak 1986; Breed and Taylor 2000) or from dissections carried out as part of this study. For the correlation analyses (see Results), residual testis masses were also employed. These are the standardized residuals from the regression of log testis mass on log body mass. All morphological data utilized are summarized in supplementary table S1 (Supplementary Material online).

For primates, we similarly classified species into those likely to experience more or less intense sexual selection on male reproductive genes, this time on the basis of mating system classifications provided by Hurle et al. (2007). Thus, those with low levels of sperm competition and, hence, less intense postcopulatory sexual selection are the owl monkey (*Aotus nancymae*), and common marmoset (*Callithrix jacchus*; “monogamous” mating systems), and gorilla (*Gorilla gorilla*), and black and white colobus (*Colobus guereza*; “polygynous” mating systems). Species with high levels of sperm competition and, hence, more intense postcopulatory sexual selection are the rhesus macaque (*Macaca mulatta*), olive baboon (*Papio anubis*), chimpanzee (*Pan troglodytes*), ring-tailed lemur (*Lemur catta*), vervet (*Cercopithecus aethiops*), and squirrel monkey (*Saimiri boliviensis*; “multimale–multifemale” mating system) as well as the orangutan (*Pongo pygmaeus*) and galago (*Otolemur garnettii*; “dispersed” mating system). Humans were included as a species with low levels of sperm competition and, hence, less intense postcopulatory sexual selection (Dixson 1998a). All these mating system classifications

of sperm competition level (except the orangutan) correspond to those based on  $RTS$  (Dixson 1998a).

### Generation of Input Alignments

For each family of orthologues, predicted protein sequences were aligned using Muscle (Edgar 2004a, 2004b) and manually adjusted. The *Svp2* family of proteins has 2 serine-rich regions that made unambiguous assignment of homologous positions problematic. These poorly aligned regions of alignments were removed using Gblocks (Castresana 2000) with the following options: maximum number of contiguous nonconserved positions = 10, minimum length of a block = 5, and gap positions allowed in all sequences. Only the Gblock parsed alignment was used as input into codeml. Verified protein alignments were used to generate corresponding cDNA alignments.

### Generation of Input Phylogenies

To allow the comparison of sites and “branch-sites” methods between different genes, unrooted input trees were generated from species phylogenies for the muroid rodents (e.g., Steppan et al. 2004; supplementary fig. S1, Supplementary Material online) by removal of unrepresented species. Branch lengths (substitutions per codon) for input phylogenies were calculated under the M0 codon model and were fixed for subsequent model comparisons.

### Protein Structure

Genes predicted to have been subject to positive selection were used to search for homologous sequences in the PDB database of protein structures (<http://www.rcsb.org/pdb/home/home.do>) using Blast (Altschul et al. 1997). Swiss-PDB Viewer (Guex et al. 1999) was used for all structural manipulations, and POVray (<http://www.povray.org>) was used to generate structural images.

## Results

### Evidence for Widespread Adaptive Evolution in Rodent Ejaculate Proteins

Two of the 7 genes analyzed (*Msemb* and *Spink3*) yielded no evidence of positive selection for these rodents. A third gene, *Svp2*, showed partial evidence of positive selection, with 1 of 2 nested models providing significant results (M7 vs. M8;  $P = 1.8 \times 10^{-3}$ ). The remaining 4 genes (*Acrv1*, *Svs2*, *Prm1*, and *Sva*) showed robust evidence of positive selection with significant LRTs for both sets of nested models (all  $P < 0.05$ ; table 1). Although analysis of all 4 genes predicted a proportion of sites with  $d_N/d_S > 1$ , for only 3 of these 4 genes (*Svs2*, *Prm1*, and *Sva*) was the resolution of these individual sites possible. Two positively selected codons were predicted for *Svs2* (S226 and S374) and *Prm1* (T44 and M47), with Bayesian posterior probabilities ( $p_B$ ) of  $p_B > 95\%$  and  $99\%$ , respectively. For *Sva*, M8/M7 and M2/M1 models predicted 20 and 8 codons, respectively, with  $p_B > 95\%$ . By aligning *Sva* to its homologue, bovine complement component 3 alpha-chain

**Table 1**  
**Summary of Sites Tests for Positive Selection**

Gene	N	LRT Statistic		Parameter Estimates			Sites under Positive Selection	
		M1 versus M2	M7 versus M8	% Sites with $d_N/d_S > 1$			M2	M8
				M2	M8	$d_N/d_S$		
<i>Msemb</i>	6	0.01	0.04	NA	NA	1.09	None	None
<i>Spink3</i>	8	3.85	4.06	NA	NA	1.97	None	None
<i>Svp2</i>	15	3.04	12.67**	NA	22	1.53	None	S43*
<i>Acrv1</i>	10	6.65*	11.48**	4	8.8	2.34	None	None
<i>Svs2</i>	6	15.37***	17.21***	6	7	3.27	S226** S374*	S226** S374**
<i>Prml</i>	20	16.44***	17.14***	2	2	14.71	T44** M47**	T44** M47**
<i>Sva</i>	5	29.68***	29.68***	33	33	7.54	T82** E85** R86** T120* L121* S124* D134* S144*	A26* Y30* H34* N35* M46* S53* N56* T82** H84* E85** R86** Y87* S114* Y118* T120* L121** S124* K131* D134** S144**

NOTE.—NA, not applicable. N = number of species analyzed. M1 versus M2: LRT statistic for model M1 versus M2. M7 versus M8: LRT statistic for model M7 versus M8. Parameter estimates: percentage of sites in  $d_N/d_S > 1$  category and estimated  $d_N/d_S$  parameter under model M8. \*Significance with  $P < 0.05$ ; \*\*significance with  $P < 0.01$ ; \*\*\*significance with  $P < 0.001$ .

(C3 $\alpha$ ), whose protein tertiary structure is known (Fredslund et al. 2006) (fig. 1A and B), we find that 5 of the 8 positively selected sites observed from both nested models form a solvent-exposed spatial cluster that would be accessible to potential binding ligands (fig. 1C). The other 3 residues lie outside of the C3 $\alpha$  homologous region. The prediction of homology between *Sva* and C3 $\alpha$  arises from a PSI-Blast (Altschul and Koonin 1998) database search using the C57BL/6J *Sva* sequence as query, which identified *Cyprinus carpio*  $\alpha$ 2-macroglobulin 4 (ABC47740.1) with significant similarity (2 iterations;  $E = 8 \times 10^{-4}$ ); the latter protein, in turn, is a homologue of bovine C3 $\alpha$  (PDB:2B39;  $E = 0.03$ , in a Blast search of the sequences of known protein structures); see also Kitano et al. (2006).

#### Evidence for Lineage-Specific Positive Selection

The sites methods that assume similar evolutionary constraints across all branches of a tree may not identify episodic adaptive evolution in the specific rodent lineages we were interested in. To address this, we used the branch-sites methods that allow  $d_N/d_S$  to vary between lineages (Yang and Nielsen 2002; Zhang et al. 2005). Using this approach, we identified lineage-specific positive selection in *Svp2* (2 lineages;  $P < 0.01$ ), *Svs2* (1 lineage;  $P < 0.001$ ), and *Sva* (2 lineages;  $P < 0.05$ ) (table 2 and fig. 2). No evidence for such lineage-specific events was obtained for *Msemb*, *Spink3*, *Acrv1*, or *Prml*. Thus, 3 of the 7 genes (*Svp2*, *Svs2*, and *Sva*) exhibit evidence for positive selection both across the entire clade we examined and for particular lineages.

For *Svp2*, positive selection was inferred for 2 lineages: the *A. cilicicus* terminal lineage ( $P < 0.01$ ), with 4 sites (25K, 66S, 72S, and 92S) identified as positively selected ( $p_B > 95\%$ ), and the internal branch separating the *Mus* clade (*Mus castaneus*, *Mus musculus musculus*, *M. macedonicus*, and *C. pahari*) from other species ( $P < 0.001$ ), where 4 different positively selected sites (47H, 48M, 50R, and 51S) were predicted with  $p_B > 95\%$  (fig.

2A). By way of contrast, positive selection was predicted for *Svs2* along a different internal lineage that separating *A. sylvaticus* and *R. norvegicus* from other species ( $P < 0.001$ ). No positively selected sites along this lineage were identified with significance ( $p_B > 95\%$ ). Positive selection was also detected in the terminal lineage leading to *A. sylvaticus* ( $P < 0.001$ ; 211S identified as positively selected with  $p_B > 95\%$ ; fig. 2B). Finally, for *Sva*, both *C. pahari* and *M. musculus musculus* terminal lineages were predicted to have experienced positive selection (8 and 16 codons identified, respectively;  $p_B > 95\%$ ). Interestingly, a stretch of 7 codons (81N-87Y, *M. musculus musculus*, PDB 2B39 E864-F871) harbors 5 sites (82T, 84H, 85E, 86R, and 87Y) predicted to have been under positive selection along the *C. pahari* lineage and 2 sites (81N, 85E) under positive selection along the *M. musculus musculus* lineage (fig. 2C). These sites correspond to C3 $\alpha$  residues E864-F871 encompassing positively selected residues homologous to C3 $\alpha$  residues L866, P869, and A870 (fig. 1C).

#### Evidence for Postcopulatory Sexual Selection Driving Positive Selection

We conducted a branch-sites analysis where the phylogenetic tree was partitioned between those species with relatively high levels of sperm competition and, hence, more intense postcopulatory sexual selection ( $RTS > 1$ ) and other species with relatively low levels of sperm competition and, hence, less intense postcopulatory sexual selection ( $RTS < 1$ ). Using this approach, we attempted to address whether interspecific differences in mating systems, and associated sperm competition intensity, have influenced the relative evolutionary rates for each of the 7 genes under investigation.

Six of the 7 genes failed to exhibit significant evidence of positive selection along  $RTS > 1$  or  $RTS < 1$  lineages when the tree was partitioned in this manner (table 3). Only an analysis of *Svs2* that considered 4  $RTS < 1$  species (*M. castaneus*, *M. musculus musculus*, *M. macedonicus*, and *C.*

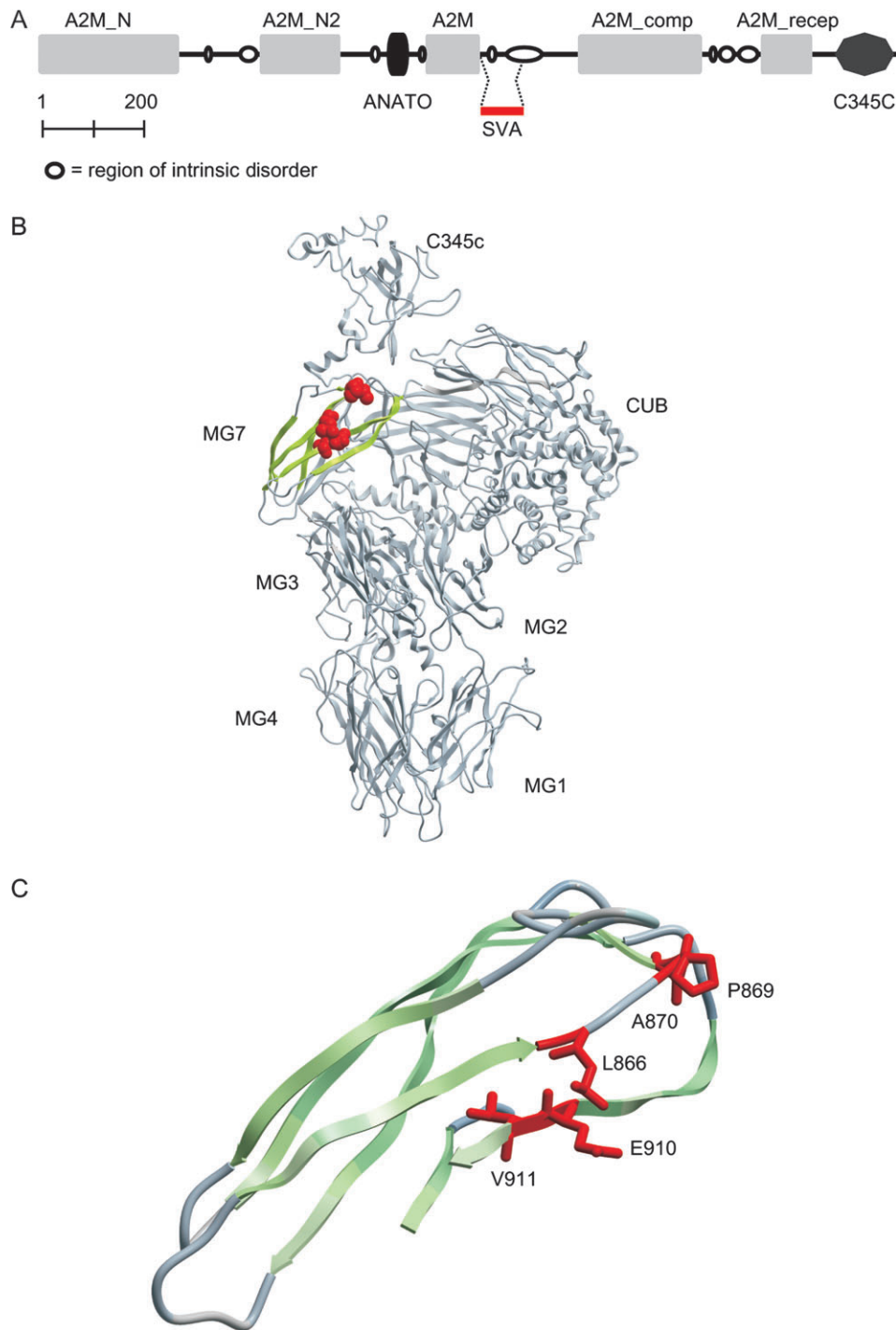


FIG. 1.—Region of similarity between SVA and the homologous MG7 domain of the C3 $\alpha$  (Fredslund et al. 2006). (A) Schematic of C3 protein domains predicted using SMART (Schultz et al. 2000) and the location of region homologous to SVA. (B) Crystal structure of C3 $\alpha$  (Fredslund et al. 2006) PDB 2B39. SVA homologous region shown in blue, positively selected sites highlighted as red space-filled atoms. (C) Magnified view of the MG7 domains of 2B39 corresponding to residues 1787–1867. Positively selected sites are highlighted red and labeled with single-letter amino acid codes.

*pahari*) and 2 RTS > 1 species (*A. sylvaticus* and *R. norvegicus*) revealed significant evidence of positive selection (fig. 3). In this test, although below our cutoff, 2 *Svs2* codons (143V and 374S) were predicted to have evolved under positive selection along those branches leading to RTS > 1 species ( $p_B = 91.7\%$ ). One of these codons,

374S, was also detected as having been positively selected in the analysis of all branches ( $M2 p_B = 98.5\%$ ,  $M8 p_B = 99.4\%$ ). To test the likely robustness of this result, we retested all possible permutations for partitioning the species tree into 2 foreground and 4 background terminal lineages. Of the 15 possible trees with 2 foreground terminal

**Table 2**  
**Summary of Branch-Sites Tests for Positive Selection**

Gene	Number of Sequences	Number of Codons	Significance of Tests for Positive Selection		
			Sites		Branch Sites
			M2	M8	All Branches
<i>Msemb</i>	6	113	NS	NS	NS
<i>Spink3</i>	8	80	NS	NS	NS
<i>Svp2</i>	15	92	NS	0.01	2 branches ( $P < 0.01$ )
<i>Acrv1</i>	10	271	0.05	0.01	NS
<i>Svs2</i>	6	451	0.001	0.001	1 branch ( $P < 0.001$ )
<i>Prml</i>	20	52	0.001	0.001	NS
<i>Sva</i>	5	158	0.001	0.001	2 branches ( $P < 0.05$ )

NOTE.—NS, not significant.

branches,  $\binom{6}{2}$ , one (with *R. norvegicus* and *C. pahari* set as foreground branches) provided greater statistical support for adaptive evolution along these foreground branches than the partition based on RTS. Thus, although our analysis identifies significant evidence of positive selection along RTS > 1 branches in the *Svs2* gene, this may be dependent on the inclusion of the more divergent rat sequence.

We also sought evidence for an influence of sperm competition on lineage-specific evolutionary rates using the less conservative approach adopted previously (e.g., Dorus et al. 2004). For this, we estimated lineage-specific  $d_N/d_S$  values under the free-ratio model and tested whether values for terminal branches correlated with the corresponding RTS values of extant species. No such significant correlations were found for any of the genes examined (supplementary table S2, Supplementary Material online).

#### Analysis of Primate Ejaculate Proteins

To compare our methodology for inferring evolutionary rates with those of others, and in an attempt to resolve the conflicting findings of Dorus et al. (2004) and Hurle et al. (2007), we reinvestigated the evolution of primate semenogelins. Dorus et al. (2004) proposed that the rate of molecular evolution of primate *SEMG2* genes is significantly correlated with the degree of female promiscuity among primate species. Although Hurle et al. (2007) found evidence of positive selection in both primate *SEMG1* and *SEMG2* genes, they only detected differential evolutionary rates between lineages for *SEMG1*. Most importantly, they were not able to confirm a statistically significant association between promiscuity and evolutionary rates (based on differences in mating systems).

Using data from Hurle et al. (2007), we applied the sites method employed above and found, in support of both previous studies, an elevated value of  $d_N/d_S$  for both *SEMG1* and *SEMG2* genes across the primate lineages. However, in contrast to Hurle et al. (2007), when using the branch-sites analysis, we were able to identify specific lineages in which both *SEMG1* and *SEMG2* genes have evolved under positive selection. In the analyses of the *SEMG1* genes, we found that the lineages leading to the

squirrel monkey, common marmoset, and black and white colobus monkey exhibited significant evidence of positive selection ( $P < 0.001$ ), as did the branch separating the New World monkey clade (squirrel monkey and marmoset) from the Old World monkeys (rhesus macaque, olive baboon, vervet monkey, and black and white colobus) and great ape (gorilla and chimpanzee) clades (fig. 4A). For the analyses of the *SEMG2* genes, we found that the lineages leading to the vervet and owl monkey showed evidence of positive selection (both  $P < 0.05$ , fig. 4B).

Separation of species into those with mating systems associated with low levels of sperm competition and, hence, less intense postcopulatory sexual selection versus those associated with high levels of sperm competition and, hence, more intense postcopulatory sexual selection allowed comparison of evolutionary rates and mating strategy (these classifications were based on those employed previously by Hurle et al. 2007). Analysis by branch-sites models analogous to those described above for the rodent RTS > 1/RTS < 1 partitions revealed evidence for positive selection in primate *SEMG2* genes only for high sperm competition species, in line with previous conclusions reached by Dorus et al. (2004) (fig. 5 and table 4).

#### Discussion

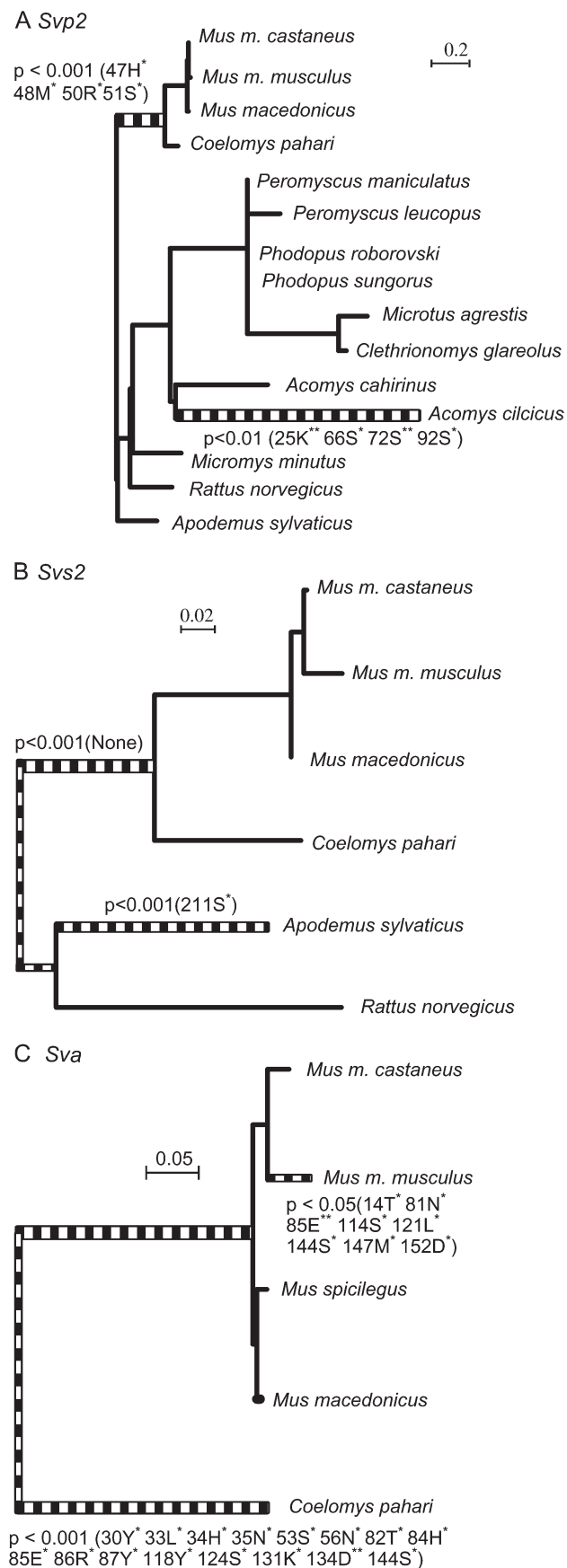
##### Widespread Adaptive Evolution of Rodent Ejaculate Proteins

Our results suggest that adaptive evolution among rodent reproductive genes is widespread, with 4 of the 7 rodent genes examined (*Acrv1*, *Svs2*, *Prml*, and *Sva*) exhibiting unambiguous evidence of positive selection and with partial evidence for a fifth gene (*Svp2*). Simple measures of rates of evolution among these genes (median  $d_N/d_S$  values for pairwise alignments range from 2.25 [*Sva*] to 0.25 [*Svs2*]) exceed those that would be expected if the 7 genes included in the study were a random sample of rodent genes: the median  $d_N/d_S$  value for 11,503 single orthologues in *Mus* and *Rattus* is 0.11 (Gibbs et al. 2004). This fits with a general assumption that reproductive genes are frequent targets of positive selection (Swanson and Vacquier 2002) and is consistent with the generally high levels of phenotypic diversity seen among reproductive traits (e.g., Eberhard 1985; Andersson 1994; Birkhead and Møller 1998).

It should be noted that for *Prml*, our study suggests that, contrary to earlier conclusions based on only 3 rodent species (Rooney and Zhang 1999), the pattern of molecular evolution in rodents is similar to that seen in primates but with positive selection predicted at a substantially lower proportion of sites (2% in rodents; compared with 28–46% of sites for primates) (Swanson, Zhang, et al. 2001; Torgerson et al. 2002).

##### Semen Coagulation Proteins as Targets of Sexual Selection

Evidence for adaptive evolution being associated with sexual selection was found for only 1 of the 7 rodent genes examined in this study, namely *Svs2*. That *Svs2* should evolve under sexual selection is consistent with results



of previous studies linking the molecular evolution of primate seminal coagulation proteins to sexual selection (see below). It is also consistent with evidence that the molecular mass of the copulatory plug protein encoded by *Svs2* covaries with sperm competition intensity across rodent species (Ramm SA, McDonald L, Hurst JL, Beynon RJ, Stockley P, unpublished data) and that variation in the number of tandem repeats responsible for coagulation shows some evidence of a correlation with mating systems in the equivalent primate gene, *SEMG2* (Hurle et al. 2007; see below). However, even for *Svs2*, our branch-site analyses of lineage-specific evolutionary rates were not wholly unambiguous: the foreground partitioning based on RTS, although a significantly better fit than the null model assuming equal evolutionary rates in all lineages, was not the best fit of all possible partitioning permutations.

Nevertheless, in support of the assertion that semen coagulation proteins are evolving under sexual selection, our analysis of primate *SEMG2* employing the same methods reached the same conclusion as for rodent *Svs2* that positive selection is concentrated in lineages subject to high levels of sperm competition. This finding is consistent with the correlation between sperm competition level and lineage-specific evolutionary rate previously reported for primate *SEMG2* by Dorus et al. (2004) and suggests that mating systems do indeed exert a significant influence on the rate of evolution of this gene. This result was obtained despite the fact that Hurle et al. (2007) did not find evidence for significant interlineage heterogeneity in evolutionary rates in *SEMG2*. Our own approach to this issue—using the Hurle et al. (2007) data set—was to partition the phylogeny into “high” and “low” sperm competition lineages and then to fit models that tested for adaptive evolution in these partitions. This procedure may result in increased power to detect positive selection because it compares only 2 evolutionary rates. However, it does assume that the partitioning is biologically meaningful and that genes grouped together are evolving under the same selective pressures (Zhang et al. 2005).

Although the precise means through which a coagulum/plug enhances male reproductive success under sperm competition continues to be debated (Ramm et al. 2005), there is a clear rationale for expecting to find variation in the strength of selection on a plug/coagulum-encoding gene according to the prevailing mating system of the species in whose genome it resides. This is based on the taxonomically widespread evolution of a seminal fluid-derived coagulum/mating plug as an adaptation to sperm competition (review in Ramm et al. 2005) and an evidence

←

FIG. 2.—Genes showing significant evidence of adaptive evolution along individual branches (branch-sites analysis). Individual branches identified as having significant evidence of episodic evolution are hatched. The significance ( $P$  value) for each branch is given for each positive branch. Sites identified to have evolved with  $d_N/d_S > 1$  along a branch are shown if  $>95\%$  posterior probability cutoff (“\*” represents more than 0.95 and “\*\*” represents more than 0.99 posterior probabilities, respectively). Branch lengths in substitutions per codon are estimated under the M0 model using codeml. A = *Svp2*, B = *Svs2*, and C = *Sva*.

**Table 3**  
**Summary of Analyses Based on High (RTS > 1) and Low (RTS < 1) Sexual Selection Partitions**

Gene	Number of Branches		LRT statistic	Parameter Estimates under Free Model			$d_N/d_S$		Sites under Positive Selection
	All	Foreground		p0 <sup>a</sup>	p1 <sup>b</sup>	p2 <sup>c</sup>	Background	Foreground	
<b>RTS &lt; 1</b>									
<i>Msemb</i>	9	2	0.00	0.37	0.63	0.00	0.12	1.00	NA
<i>Spink3</i>	13	6	0.07	0.41	0.45	0.14	0.00	1.58	NA
<i>Svp2</i>	27	8	0.00	0.29	0.22	0.50	0.10	1.00	NA
<i>Acry1</i>	17	6	0.05	0.70	0.26	0.04	0.11	1.44	NA
<i>Svs2</i>	9	4	0.00	0.72	0.29	0.00	0.05	1.00	NA
<i>Prm1</i>	37	11	0.00	0.73	0.27	0.00	0.04	1.00	NA
<i>Sva</i>	7	ND	ND	ND	ND	ND	ND	ND	NA
<b>RTS &gt; 1</b>									
<i>Msemb</i>	9	4	1.10	0.68	0.15	0.16	0.37	3.40	NA
<i>Spink3</i>	13	2	2.93	0.48	0.38	0.15	0.03	3.91	NA
<i>Svp2</i>	27	9	0.00	0.60	0.41	0.00	0.19	1.00	NA
<i>Acry1</i>	17	4	0.00	0.72	0.28	0.00	0.11	1.00	NA
<i>Svs2</i>	9	2	16.68**	0.73	0.24	0.03	0.08	41.25	None
<i>Prm1</i>	37	11	0.00	0.61	0.18	0.21	0.00	1.00	NA
<i>Sva</i>	7	ND	ND	ND	ND	ND	ND	ND	NA

NOTE.—NA, not available. ND, not determined. Number of branches: count of branches partitioned as foreground. LRT statistic: 2 times log likelihood difference of fixed versus free nested models (see Materials and Methods). \*\*Significant at  $P < 0.01$ .

<sup>a</sup> Proportion of sites in category p0 where selection is purifying throughout the tree ( $0 < \omega_0 < 1$ ).

<sup>b</sup> Proportion of sites in category p1 where selection is neutral throughout the tree ( $\omega_1 = 1$ ).

<sup>c</sup> Proportion of sites in category p2 where selection is positive in foreground branches; sum of category 2a: purifying selection in background but positive selection in foreground ( $0 < \omega_0 < 1$ ,  $\omega_2 > 1$ ); and category 2b: neutral selection in background but positive selection in foreground ( $\omega_1 = 1$ ,  $\omega_2 > 1$ ).

that the degree of coagulation of seminal fluid in primates (Dixson and Anderson 2002), the size of the plug in rodents (Ramm et al. 2005), and the size of the plug-producing accessory glands in both these groups (Dixson 1998b; Ramm et al. 2005) all correlate significantly and positively with sperm competition level.

Our results for the other primate gene we examined, *SEMG1*, do not support a heightened rate of evolution in lineages subject to more intense postcopulatory sexual selection. In fact, they tend to support the opposite trend (results not shown), with several branches subject to low sperm competition levels appearing to exhibit evidence of adaptive evolution (fig. 4). Previously, Kingan et al. (2003) and Jensen-Seaman and Li (2003) failed to find evidence of significant evolutionary rate heterogeneity in *SEMG1* among the (4–6) species they sampled, but both studies noted the higher rate of evolution in the *Pan–Homo* clade compared with *Gorilla*, a pattern that may be attributable to positive selection. Although these findings may require additional investigation, our results for primate *SEMG1* support our conclusions for rodents that a correlation of evolutionary rates with mating systems is not an inevitable property of all reproduction-related genes with putative functions in postcopulatory sexual selection (see below).

#### Not All Reproductive Genes Exhibit a Signature of Sexual Selection

Although these genes were initially selected on the basis of their products' potential influence on male reproductive success, and evidence of adaptive evolution exists in many cases, of the 6 other putatively sexually selected rodent genes examined in our study, none showed clear evidence that lineage-specific rates of evolution correlate with

sexual selection intensity. Although our results do not call into question this functional assignment, they do suggest that the selection regime on these genes may not fluctuate strongly with changes in sexual selection brought about by divergence in mating systems. Despite this, several of the genes examined did exhibit evidence of adaptive evolution, suggesting that they are not simply subject to selective constraints that maintain their function across the sampled species. The adaptive benefits of the molecular changes we

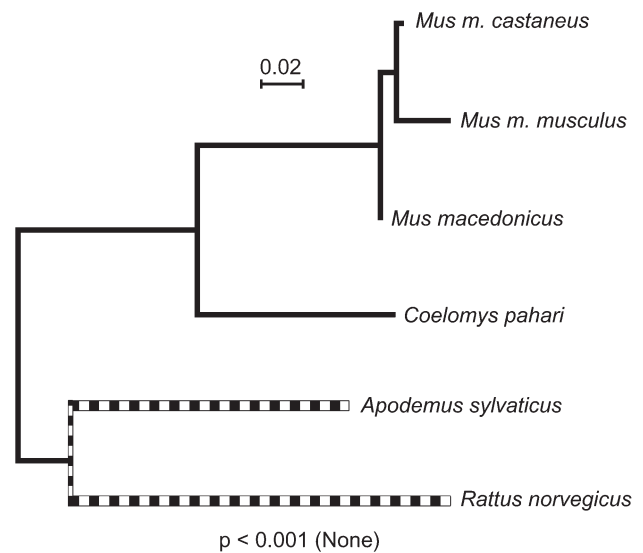


FIG. 3.—Episodic evolution of rodent *Svs2* gene linked to sperm competition. RTS > 1 branches identified as exhibiting significant evidence of episodic evolution are hatched. The significance is shown. No sites were identified to have evolved with  $d_N/d_S > 1$  above the 95% posterior probability cutoff. Branch lengths in substitutions per codon are estimated under the M0 model using codeml.

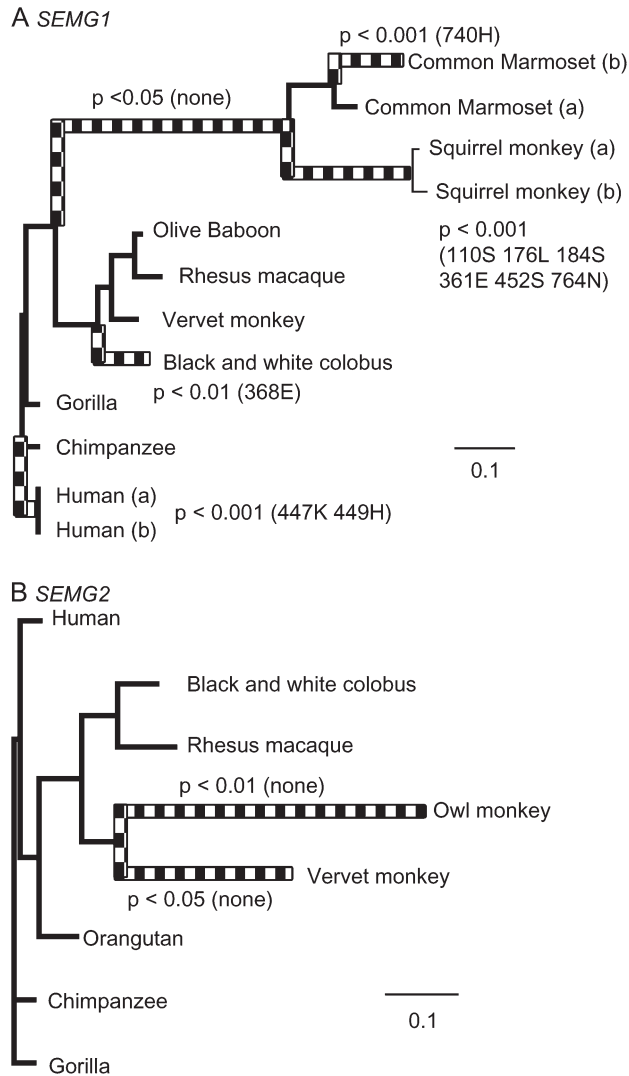


FIG. 4.—Branch-sites analysis of primate SEMG1 and SEMG2 genes. Individual branches identified as having significant evidence of episodic evolution are hatched. The significance ( $P$  value) for each branch is given for each positive branch. Sites identified to have evolved with  $d_N/d_S > 1$  along a branch are shown if  $>95\%$  posterior probability cutoff (“\*” represents more than 0.95 and “\*\*\*” represents more than 0.99 posterior probabilities, respectively). Branch lengths in substitutions per codon are estimated under the M0 model using codeml. A = SEMG1 and B = SEMG2. Sequence accession numbers (DP000036–DP000048) were obtained from Hurlle et al. (2007), human (a) = NM\_003007.2, human (b) = NM\_198139.1.

have documented in *Acrv1*, *Prml1*, and *Sva* (and possibly *Svp2*), however, remain to be determined.

#### Detecting Molecular Adaptation Driven by Sexual Selection

The branch-sites methods employed here to detect adaptive evolution driven by sexual selection have been shown to be robust in analyzing molecular evolution but may be limited in power when sample size or divergence times are low (Zhang et al. 2005). Additionally, where multiple branches are considered as foreground, power will be reduced unless the same sites and selective constraints are occurring along all

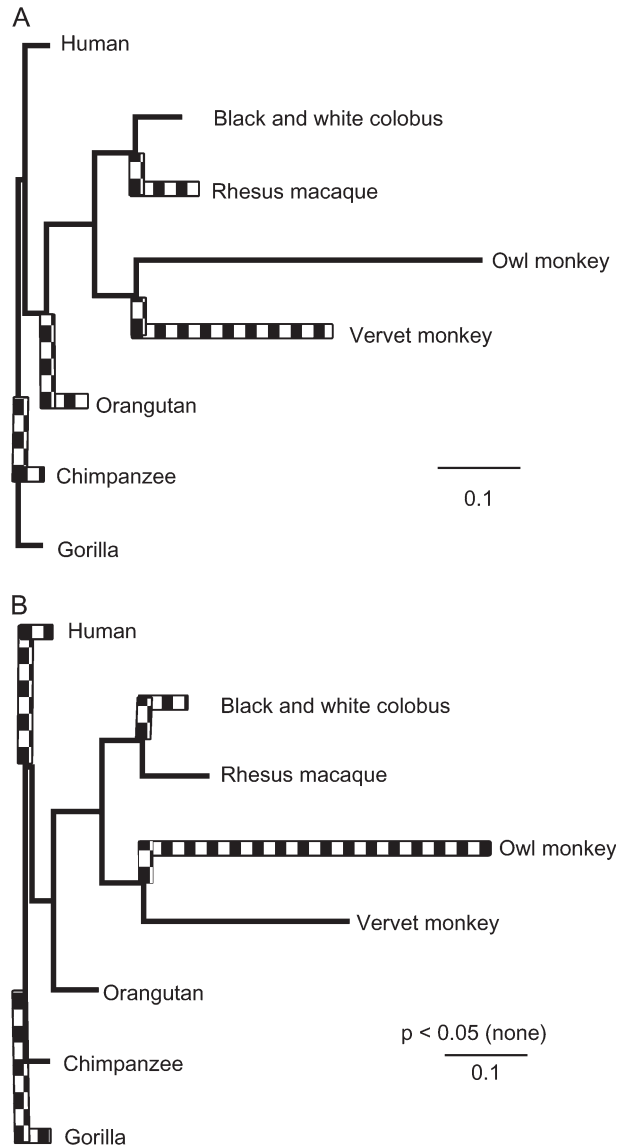


FIG. 5.—Episodic evolution of primate SEMG2 gene linked to sperm competition. For both low (A) and high (B) sperm competition, foreground partition branches are hatched. When high sperm competition branches were grouped as foreground, (B) positive selection was detected ( $P < 0.05$ ). No sites were identified to have evolved with  $d_N/d_S > 1$  above the 95% posterior probability cutoff. Branch lengths in substitutions per codon are estimated under the M0 model using codeml.

foreground branches (Zhang et al. 2005). In the context of linking adaptive evolution to sexual selection, this assumption may be unrealistic because there may often be several different mutational routes to increased reproductive fitness. This interlineage heterogeneity also applies more generally across, as well as within, particular genes because many traits under sexual selection are likely to be quantitative characters controlled by multiple genes.

For this study, we must also consider the different timescales of testis and molecular evolution. Although the former responds readily to selection for increased sperm production to gain paternity under sperm competition (Hosken and Ward 2001; see also Preston et al. 2003; Schulte-Hostedde and Millar 2004), the lineage-specific

**Table 4**  
**Summary of Branch-Sites Analysis of Primate *SEMGI* and *SEMGI2* Genes Linked to Sexual selection**

Gene	Number of Branches		LRT Statistic	Parameter Estimates under Free Model			$d_N/d_S$		Sites under Positive Selection
	All	Foreground		p0 <sup>a</sup>	p1 <sup>b</sup>	p2 <sup>c</sup>	Background	Foreground	
Low sperm competition									
<i>Semg1</i>	21	6	0	0.22	0.78	0	0.03	1.00	NA
<i>Semg2</i>	13	4	4.012	0.18	0.67	0.15	0.00	3.27	NA
High sperm competition									
<i>Semg1</i>	21	6	0	0.22	0.78	0	0.03	1.00	NA
<i>Semg2</i>	13	4	6.852*	0.18	0.77	0.06	0.00	6.26	None

NOTE.—NA, not available. Number of branches: count of branches partitioned as foreground. LRT statistic: 2 times log likelihood difference of fixed versus free nested models (see Materials and Methods). \*Significant at  $P < 0.05$  Bonferroni-corrected critical values ( $P < 0.05$ : 5.024,  $P < 0.01$ : 7.879).

<sup>a</sup> Proportion of sites in category p0 where selection is purifying throughout the tree ( $0 < \omega_0 < 1$ ).

<sup>b</sup> Proportion of sites in category p1 where selection is neutral throughout the tree ( $\omega_1 = 1$ ).

<sup>c</sup> Proportion of sites in category p2 where selection is positive in foreground branches; sum of category 2a: purifying selection in background but positive selection in foreground ( $0 < \omega_0 < 1$ ,  $\omega_2 > 1$ ); and sum of category 2b: neutral selection in background but positive selection in foreground ( $\omega_1 = 1$ ,  $\omega_2 > 1$ ).

rates we estimated are in effect averaged values over longer timescales, specifically, since the last common ancestor of the species in question and its closest sampled neighbor in the phylogeny. In the absence of a fully saturated phylogenetic sampling of extant and extinct species (together with information on their testis size or another proxy for mating system), each evolutionary rate estimate will represent both a proportion of the evolutionary history about which we have informative data on sexual selection intensity and a (potentially overriding) proportion of the evolutionary history about which we are largely ignorant. The actual proportions will depend upon the range and depth of phylogenetic sampling, the divergence times of the species under study, and the rate of evolution of the proxy measure for sexual selection intensity employed, all of which may mean that detecting a molecular signature of sexual selection will be difficult in all but the most clear-cut of cases.

Given the above considerations, we suggest that our failure to detect a widespread signature of sexual selection among rodent reproductive genes cannot easily be attributed to the absence of sexual selection acting on these genes. Alternative approaches may in some cases avoid some of the methodological issues we have raised. For example, because sexual selection may often account for instances of episodic adaptive evolution, but these need not always affect the same genes or the same sites within genes, in different lineages, the most likely level at which any influence of sexual selection on the rate of molecular evolution will be detected may be as an elevated average rate of—or predisposal to—diversification over the reproductive genome as a whole when comparing relatively promiscuous lineages with less promiscuous ones. Thus the generally higher rate of female remating among *Drosophila* species of the *repleta* group compared with the *melanogaster* subgroup may explain why on average Acps in the former group exhibit faster rates of sequence divergence and a greater propensity to accumulate duplications (Wagstaff and Begun 2005, 2007). Alternatively, because models of sexual selection predict coevolution of interacting male and female traits, it might also be possible to test whether the lineage-specific evolutionary rates of male and female genes encoding the proteins involved in such interactions are correlated because adaptation in one sex

is predicted to lead to counteradaptation in the other sex (see also Civetta 2003).

## Conclusions

Sexual selection drives the rapid diversification of many behavioral, morphological, and physiological traits, operating at both the pre- and postcopulatory stages of reproduction. Comparative molecular evolutionary studies now offer the possibility of detecting bouts of adaptive evolution driven by sexual selection at the molecular level, and it has been suggested that it is this force that is responsible for the rapid evolution of many genes expressed in reproductive tissues. However, despite evidence for widespread adaptive evolution in reproductive genes, linking interspecific variation in the intensity of sexual selection to rates of molecular evolution is not straightforward. Notwithstanding the evidence we have presented linking semen coagulation proteins to sexual selection in both rodents and primates, the more general conclusion to emerge from our study is that there is not necessarily a simple link to be made between variation in mating systems among species and lineage-specific evolutionary rates, at least not on gene-by-gene basis. The extent to which sexual selection can account for the rapid evolution of reproductive genes therefore remains to be determined, and other potential causative factors (e.g., host–pathogen coevolution, because many reproductive genes have immune-related functions) should not be prematurely discounted.

## Supplementary Material

Supplementary fig. S1 and tables S1 and S2 are available at *Molecular Biology and Evolution* online (<http://www.mbe.oxfordjournals.org/>).

## Acknowledgments

We are extremely grateful to the following for supplying biological samples: M. Bennett, K. Blasdell, and G. A. Parker (University of Liverpool, UK); S. Siah and S. Unwin (North of England Zoological Society, Cheshire, UK); A. Loudon (University of Manchester, UK);

J. Crossland (Peromyscus Genetic Stock Center, University of South Carolina); Robert Deacon (University of Oxford, UK); and A. Orth (Université Montpellier 2, France). RDE was funded by a MRC UK training fellowship (G90/112). PLO and CPP were funded by the UK Medical Research Council. SAR was funded by a studentship from the University of Liverpool (to PS and G. A. Parker).

## Literature Cited

- Altschul SF, Koonin EV. 1998. Iterated profile searches with PSI-BLAST—a tool for discovery in protein databases. *Trends Biochem Sci.* 23:444–447.
- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25:3389–3402.
- Andersson M. 1994. *Sexual selection*. Princeton (NJ): Princeton University Press.
- Andrés JA, Maroja LS, Bogdanowicz SM, Swanson WJ, Harrison RG. 2006. Molecular evolution of seminal proteins in field crickets. *Mol Biol Evol.* 23:1574–1584.
- Bielawski JP, Yang Z. 2003. Maximum likelihood methods for detecting adaptive evolution after gene duplication. *J Struct Funct Genomics.* 3:201–212.
- Birkhead TR, Møller AP. 1998. *Sperm competition and sexual selection*. New York: Academic Press.
- Breed WG, Taylor J. 2000. Body mass, testes mass, and sperm size in murine rodents. *J Mammal.* 81:758–768.
- Bustamante CD, Fledel-Alon A, Williamson S, et al. (14 co-authors). 2005. Natural selection on protein-coding genes in the human genome. *Nature.* 437:1153–1157.
- Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol.* 17:540–552.
- Chao CF, Chiou ST, Jeng H, Chang WC. 1996. The porcine sperm motility inhibitor is identical to beta-microseminoprotein and is a competitive inhibitor of Na<sup>+</sup>,K<sup>+</sup>-ATPase. *Biochem Biophys Res Commun.* 218:623–628.
- Civetta A. 2003. Shall we dance or shall we fight? Using DNA sequence data to untangle controversies surrounding sexual selection. *Genome.* 46:925–929.
- Clark NL, Aagaard JE, Swanson WJ. 2006. Evolution of reproductive proteins from animals and plants. *Reproduction.* 131:11–22.
- Clark AG, Civetta A. 2000. Evolutionary biology—Protamine wars. *Nature.* 403:261–263.
- Clark NL, Swanson WJ. 2005. Pervasive adaptive evolution in primate seminal proteins. *PLoS Genet.* 1:335–342.
- Dixon AF. 1998a. *Primate sexuality*. Oxford: Oxford University Press.
- Dixon AF. 1998b. Sexual selection and evolution of the seminal vesicles in primates. *Folia Primatol.* 69:300–306.
- Dixon AF, Anderson MJ. 2002. Sexual selection, seminal coagulation and copulatory plug formation in primates. *Folia Primatol.* 73:63–69.
- Dorus S, Evans PD, Wyckoff GJ, Choi SS, Lahn BT. 2004. Rate of molecular evolution of the seminal protein gene SEMG2 correlates with levels of female promiscuity. *Nat Genet.* 36:1326–1329.
- Eberhard WG. 1985. *Sexual selection and animal genitalia*. Cambridge (MA): Harvard University Press.
- Edgar RC. 2004a. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics.* 5:1–19.
- Edgar RC. 2004b. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32:1792–1797.
- Emes RD, Goodstadt L, Winter EE, Ponting CP. 2003. Comparison of the genomes of human and mouse lays the foundation of genome zoology. *Hum Mol Genet.* 12:701–709.
- Fedoroka KM, Zuk M. 2005. Sexual conflict and female immune suppression in the cricket, *Allonemobius socius*. *J Evol Biol.* 18:1515–1522.
- Foster JA, Klotz KL, Flickinger CJ, Thomas TS, Wright RM, Castillo JR, Herr JC. 1994. Human Sp-10—acrosomal distribution, processing, and fate after the acrosome reaction. *Biol Reprod.* 51:1222–1231.
- Fredslund F, Jenner L, Husted LB, Nyborg J, Andersen GR, Sottrup-Jensen L. 2006. The structure of bovine complement component 3 reveals the basis for thioester function. *J Mol Biol.* 361:115–127.
- Gibbs RAGM, Weinstock ML, Metzker DM, et al. (230 co-authors). 2004. Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature.* 428:493–521.
- Good JM, Nachman MW. 2005. Rates of protein evolution are positively correlated with developmental timing of expression during mouse spermatogenesis. *Mol Biol Evol.* 22:1044–1052.
- Gordi T, Khamis H. 2004. Simple solution to a common statistical problem: interpreting multiple tests. *Clin Ther.* 26:780–786.
- Guex N, Diemand A, Peitsch MC. 1999. Protein modelling for all. *Trends Biochem Sci.* 24:364–367.
- Herlyn H, Zischler H. 2007. Sequence evolution of the sperm ligand zonadhesin correlates negatively with body weight dimorphism in primates. *Evolution.* 61:289–298.
- Hosken DJ, Ward PI. 2001. Experimental evidence for testis size evolution via sperm competition. *Ecol Lett.* 4:10–13.
- Huang YH, Chen YH, Lin CM, Ciou YY, Kuo SP, Chen CT, Shih CM, Chang EE. 2007. Suppression effect of seminal vesicle autoantigen on platelet-activating factor-induced mouse sperm capacitation. *J Cell Biochem.* 100:941–951.
- Huang YH, Kuo SP, Lin MH, Shih CM, Chu ST, Wei CC, Wu TJ, Chen YH. 2005. Signals of seminal vesicle autoantigen suppresses bovine serum albumin-induced capacitation in mouse sperm. *Biochem Biophys Res Commun.* 338:1564–1571.
- Hurle B, Swanson W, NISC Comparative Sequencing Program, Green ED. 2007. Comparative sequence analyses reveal rapid and divergent evolutionary changes of the WFDC locus in the primate lineage. *Genome Res.* 17:276–286.
- Ialenti A, Santagada V, Caliendo G, et al. (13 co-authors). 2001. Synthesis of novel anti-inflammatory peptides derived from the amino-acid sequence of the bioactive protein SV-IV. *Eur J Biochem.* 268:3399–3406.
- Jensen-Seaman MI, Li WH. 2003. Evolution of the hominoid semenogelin genes, the major proteins of ejaculated semen. *J Mol Evol.* 57:261–270.
- Kenagy GJ, Trombulak SC. 1986. Size and function of mammalian testes in relation to body size. *J Mammal.* 67:1–22.
- Kingan SB, Tatar M, Rand DM. 2003. Reduced polymorphism in the chimpanzee semen coagulating protein, semenogelin I. *J Mol Evol.* 57:159–169.
- Kitano T, Tian W, Umetsu K, Yuasa I, Yamazaki K, Saitou N, Osawa M. 2006. Origin and evolution of gene for prolactin-induced protein. *Gene.* 383:64–70.

- Lundwall A, Lazure C. 1995. A novel gene family encoding proteins with highly differing structure because of a rapidly evolving exon. *FEBS Lett.* 374:53–56.
- Lung O, Tram U, Finnerty CM, Eipper-Mains MA, Kalb JM, Wolfner MF. 2002. The *Drosophila melanogaster* seminal fluid protein Acp62F is a protease inhibitor that is toxic upon ectopic expression. *Genetics.* 160:211–224.
- Mills JS, Needham M, Parker MG. 1987. A secretory protease inhibitor requires androgens for its expression in male sex accessory tissues but is expressed constitutively in pancreas. *EMBO J.* 6:3711–3717.
- Nadeau NJ, Burke T, Mundy NI. 2007. Evolution of an avian pigmentation gene correlates with a measure of sexual selection. *Proc R Soc Lond B Biol Sci.* 274:1807–1813.
- Needham M, Mills JS, Parker MG. 1988. Organization and upstream DNA-sequence of the mouse protease inhibitor gene. *Nucleic Acids Res.* 16:6229.
- Nielsen R, Yang ZH. 1998. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics.* 148:929–936.
- Parker GA, Ball MA, Stockley P, Gage MJG. 1997. Sperm competition games: a prospective analysis of risk assessment. *Proc R Soc Lond B Biol Sci.* 264:1793–1802.
- Podlaha O, Webb DM, Tucker PK, Zhang J. 2005. Positive selection for indel substitutions in the rodent sperm protein catsper1. *Mol Biol Evol.* 22:1845–1852.
- Preston BT, Stevenson IR, Pemberton JM, Coltman DW, Wilson K. 2003. Overt and covert competition in a promiscuous mammal: the importance of weaponry and testes size to male reproductive success. *Proc R Soc Lond B Biol Sci.* 270:633–640.
- Pröschel M, Zhang Z, Parsch J. 2006. Widespread adaptive evolution of *drosophila* genes with sex-biased expression. *Genetics.* 174:893–900.
- Ramm SA, Parker GA, Stockley P. 2005. Sperm competition and the evolution of male reproductive anatomy in rodents. *Proc R Soc Lond B Biol Sci.* 272:949–955.
- Richards S, Liu Y, Bettencourt BR, et al. (52 co-authors). 2005. Comparative genome sequencing of *Drosophila pseudoobscura*: chromosomal, gene, and cis-element evolution. *Genome Res.* 15:1–18.
- Rooney AP, Zhang JZ. 1999. Rapid evolution of a primate sperm protein: relaxation of functional constraint or positive Darwinian selection? *Mol Biol Evol.* 16:706–710.
- Schulte-Hostedde AI, Millar JS. 2004. Intraspecific variation of testis size and sperm length in the yellow-pine chipmunk (*Tamias amoenus*): implications for sperm competition and reproductive success. *Behav Ecol Sociobiol.* 55:272–277.
- Schultz J, Copley RR, Doerks T, Ponting CP, Bork P. 2000. SMART: a web-based tool for the study of genetically mobile domains. *Nucleic Acids Res.* 28:231–234.
- Steppan SJ, Adkins RM, Anderson J. 2004. Phylogeny and divergence-date estimates of rapid radiations in muroid rodents based on multiple nuclear genes. *Syst Biol.* 53:533–553.
- Swanson WJ, Clark AG, Waldrip-Dail HM, Wolfner MF, Aquadro CF. 2001. Evolutionary EST analysis identifies rapidly evolving male reproductive proteins in *Drosophila*. *Proc Natl Acad Sci USA.* 98:7375–7379.
- Swanson WJ, Vacquier VD. 2002. The rapid evolution of reproductive proteins. *Nat Rev Genet.* 3:137–144.
- Swanson WJ, Zhang ZH, Wolfner MF, Aquadro CF. 2001. Positive Darwinian selection drives the evolution of several female reproductive proteins in mammals. *Proc Natl Acad Sci USA.* 98:2509–2514.
- Torgerson DG, Kulathinal RJ, Singh RS. 2002. Mammalian sperm proteins are rapidly evolving: evidence of positive selection in functionally diverse genes. *Mol Biol Evol.* 19:1973–1980.
- Turner LM, Hoekstra HE. 2006. Adaptive evolution of fertilization proteins within a genus: variation in ZP2 and ZP3 in deer mice (*Peromyscus*). *Mol Biol Evol.* 23:1656–1669.
- Vamathevan J, Holbrook JD, Emes RD. Forthcoming. The mouse genome as a rodent model in evolutionary studies. *Encycl Life Sci.*
- Wagstaff BJ, Begun DJ. 2005. Molecular population genetics of accessory gland protein genes and testis-expressed genes in *Drosophila mojavensis* and *D. arizonae*. *Genetics.* 171:1083–1101.
- Wagstaff BJ, Begun DJ. 2007. Adaptive evolution of recently duplicated accessory gland protein genes in desert *Drosophila*. *Genetics.* 177:1023–1030.
- Wong WSW, Yang ZH, Goldman N, Nielsen R. 2004. Accuracy and power of statistical methods for detecting adaptive evolution in protein coding sequences and for identifying positively selected sites. *Genetics.* 168:1041–1051.
- Wyckoff GJ, Wang W, Wu CI. 2000. Rapid evolution of male reproductive genes in the descent of man. *Nature.* 403:304–309.
- Xuan JW, Kwong J, Chan FL, Ricci M, Imasato Y, Sakai H, Fong GH, Panchal C, Chin JL. 1999. cDNA, genomic cloning, and gene expression analysis of mouse PSP94 (Prostate secretory protein of 94 amino acids). *DNA Cell Biol.* 18:11–26.
- Yang Z, Nielsen R, Goldman N, Pedersen AM. 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics.* 155:431–449.
- Yang ZH. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci.* 13:555–556.
- Yang ZH. 1998. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Mol Biol Evol.* 15:568–573.
- Yang ZH, Bielawski JP. 2000. Statistical methods for detecting molecular adaptation. *Trends Ecol Evol.* 15:496–503.
- Yang ZH, Nielsen R. 2002. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Mol Biol Evol.* 19:908–917.
- Zhang JZ, Nielsen R, Yang ZH. 2005. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol Biol Evol.* 22:2472–2479.

Jianzhi Zhang, Associate Editor

Accepted October 30, 2007