RESEARCH ARTICLES

Adaptive Evolution of Fertilization Proteins within a Genus: Variation in ZP2 and ZP3 in Deer Mice (*Peromyscus*)

Leslie M. Turner and Hopi E. Hoekstra

Division of Biological Sciences, University of California, San Diego

Rapid evolution of reproductive proteins has been documented in a wide variety of taxa. In internally fertilized species, knowledge about the evolutionary dynamics of these proteins between closely related taxa is primarily limited to accessory gland proteins in the semen of *Drosophila*. Investigation of additional taxa and functional classes of proteins is necessary in order to determine if there is a general pattern of adaptive evolution of reproductive proteins between recently diverged species. We performed an evolutionary analysis of 2 egg coat proteins, ZP2 and ZP3, in 15 species of deer mice (genus *Peromyscus*). Both of these proteins are involved in egg–sperm binding, a critical step in maintaining species-specific fertilization. Here, we show that *Zp2* and *Zp3* gene trees are not consistent with trees based on nonreproductive genes, *Mc1r* and *Lcat*, where species formed monophyletic clades. In fact, for both of the reproductive genes, intraspecific amino acid variation was extensive and alleles were sometimes shared across species. We document positive selection acting on ZP2 and ZP3 and identify specific amino acid sites that are likely targets of selection using both maximum likelihood approaches and patterns of parallel amino acid change. In ZP3, positively selected sites are clustered in and around the region implicated in sperm binding in *Mus*, suggesting changes may impact egg–sperm binding and fertilization potential. Finally, we identify lineages with significantly elevated rates of amino acid substitution using a Bayesian mapping approach. These findings demonstrate that the pattern of adaptive reproductive protein evolution found at higher taxonomic levels can be documented between closely related mammalian species, where reproductive isolation has evolved recently.

Introduction

Over the past decade, a pattern of rapid evolution of proteins involved in reproduction has emerged from research in taxa ranging from diatoms to primates (Singh and Kulathinal 2000; Swanson and Vacquier 2002a, 2002b). Investigations of reproductive protein evolution have examined sperm–egg recognition proteins in marine invertebrates (Swanson and Vacquier 2002a, 2002b), accessory gland proteins in semen of *Drosophila* (Civetta and Singh 1995; Cirera and Aguade 1997; Tsaur and Wu 1997; Aguade 1999; Begun et al. 2000; Swanson, Clark et al. 2001), proteins expressed in the female reproductive tract of *Drosophila* (Swanson et al. 2004), and male and female reproductive proteins in mammals (Wyckoff et al. 2000; Swanson, Yang et al. 2001; Torgerson et al. 2002; Jansa et al. 2003; Kingan et al. 2003; Swanson et al. 2003; Dorus et al. 2004; Clark and Swanson 2005). Rapid divergence and positive selection have been documented in many of these reproductive proteins.

Adaptive evolution of egg and sperm interaction proteins, specifically, has been documented at several taxonomic levels in marine invertebrates (Swanson and Vacquier 2002b; Galindo et al. 2003; Geyer and Palumbi 2003; Mah et al. 2005). In these broadcast spawners, maintenance of species-specific binding of gametes has been proposed as a possible explanation for their rapid divergence (Swanson and Vacquier 2002a). However, the selective pressures driving reproductive protein evolution may vary in different taxa (Swanson and Vacquier 2002a). For example, in internally fertilized species, organisms have more control over which individuals exchange gametes, and the environment in which gametes interact is different (Eisenbach and Giojalas 2006). Thus, the selective pressures on reproductive proteins in internally fertilized species may differ substantially from those acting on species with external fertilization.

The seminal proteins produced by accessory glands of *Drosophila* (Acps) are the most well-studied class of reproductive proteins in internally fertilized species. Rapid evolution and positive selection have been documented for many Acps, both between closely and distantly related species (Begun et al. 2000; Swanson, Clark et al. 2001; Begun and Lindfors 2005; Mueller et al. 2005). Recent work has demonstrated that proteins expressed in the testes, ovaries, and female reproductive tracts of *Drosophila* also evolve rapidly, although not as dramatically as Acps (Swanson et al. 2004; Jagadeeshan and Singh 2005). However, detailed examination of the evolution of reproductive proteins in different taxa and functional classes is necessary to determine if evolution of *Drosophila* Acps reflects a general pattern in internally fertilized species.

Research on the evolution of reproductive proteins in mammals has thus far focused primarily on identifying genes that have experienced positive selection by analyzing sequences from distantly related species. To our knowledge, there has only been one study addressing patterns of evolution of a reproductive protein within a mammalian genus (Jansa et al. 2003); the authors provided evidence that the egg protein ZP3 had experienced positive selection in the *Mus* genus; however, when their analysis was repeated without including outgroup sequences, there was no longer evidence for positive selection (see Supplementary Material online for details). Lack of significance may be due to limited sampling; therefore, we decided to test extensively for positive selection on egg proteins in the evolution of a single genus.

Here, we extend previous work in mammals by documenting patterns of evolution of egg coat proteins in closely related species of deer mice (genus *Peromyscus*). As in *Drosophila*, pairs of *Peromyscus* taxa with varying degrees

Key words: adaptive evolution, fertilization, *Peromyscus*, positive selection, reproductive isolation, reproductive proteins.

E-mail: lturner@biomail.ucsd.edu.

Mol. Biol. Evol. 23(9):1656–1669. 2006
doi:10.1093/molbev/msl035
Advance Access publication June 14, 2006

of reproductive isolation may be sampled, including populations, subspecies, sister species, species, and species groups (Hooper 1968). In addition, sperm competition and sexual conflict have been proposed as important factors driving reproductive protein evolution (Wyckoff et al. 2000; Price et al. 2001; Swanson, Yang et al. 2001; Torgerson et al. 2002). *Peromyscus* has well documented variation in mating system (Kleiman 1977; Wolff 1989); thus, we are also able to compare evolution of fertilization proteins between closely related species where the selective environment may differ.

The fertilization process, and specifically egg–sperm interactions, is better understood in mammals than in other internally fertilized species, providing a large number of candidate genes. We focused on 2 proteins that are directly involved in egg–sperm binding because this step of fertilization is critical to species-specific fertilization (Wassarman et al. 2001). The egg proteins ZP2 (zona pellucida glycoprotein 2, *Zp2*) and ZP3 (zona pellucida glycoprotein 3, *Zp3*) are 2 of the proteins that make up the zona pellucida, or egg coat, and they are both necessary for binding of the egg and sperm (Wassarman and Litscher 2001). We chose to focus initially on the egg component of this interaction because the identity and function of the sperm proteins involved are less well defined (Jansen et al. 2001).

The goal of this study was to determine patterns of evolution of ZP2 and ZP3 in *Peromyscus*. We identify differences in tree topologies and patterns of intraspecific variation between these egg coat proteins and nonreproductive proteins. We document positive selection acting on ZP2 and ZP3 and determine the spatial pattern and identity of amino acid sites under selection. Finally, we identify lineages with significantly elevated rates of amino acid substitution in ZP2 and ZP3. Together, these results suggest that positive selection is driving divergence of egg coat proteins in closely related species and allow us to nominate candidate amino acid sites that may contribute to reduced fertilization potential between sister taxa.

Materials and Methods

Extraction, Amplification, and Sequencing

To maximize genetic variation, 1–3 geographic locales for each of 15 *Peromyscus* species were sampled (table 1, see Supplementary Material online for details). For each locale, 1–2 individuals were included, for a total of 44 individuals (*Zp2*) and 48 individuals (*Zp3*). An additional 2 individuals of an outgroup species, *Onychomys torridus*, were sequenced for each gene. Genomic DNA was extracted from frozen or ethanol-preserved tissue samples (tail, liver, or kidney) using DNeasy tissue kits (Qiagen, Valencia, CA).

The entire genomic sequence of each reproductive gene and some 5′ and 3′ flanking sequence was determined in *Peromyscus polionotus*, totaling 12,755 bp for *Zp2* and 11,518 bp for *Zp3* (see Supplementary Material online for complete genomic sequences). Initially, 2- to 4-kb regions of each gene were amplified using primers designed in conserved regions, based on aligned exon sequences from mammalian species available in GenBank. Resulting polymerase chain reaction (PCR) products were cloned (pGEM-T system, Promega, San Luis Obispo, CA) and sequenced using T7 and SP6 primers and internal sequencing primers. Sequences were edited and contigs assembled using SEQUENCHER (Gene Codes, Ann Harbor, MI). Once sequences were verified as the correct targets based on identity with *Mus* sequences, a genome walking approach (Universal GenomeWalker Kit, Clontech, Mountain View, CA) was used to amplify and sequence 5′ and 3′ of cloned regions in the same *P. polionotus* individual until the entire genomic sequence was determined. The predicted amino acid sequence was aligned to *Mus* and *Rattus* sequences using ClustalW (Chenna et al. 2003).

Table 1
Samples of 15 *Peromyscus* Species and the Outgroup *Onychomys torridus*

Species	Sampling Locations
P. aztecus	Michoacan, Mexico; Guerrero, Mexico; Chiapas, Mexico
P. boylii	Culberson Co., TX; Nuevo Leon, Mexico; San Luis Obispo Co., CA; Monterey Co., CA[a] (2)
P. *californicus*	Los Angeles Co., CA; San Diego Co., CA[b,c] (2); Monterey Co., CA[b,c] (2)
P. crinitus	Yuma Co., AZ[a] (2); Tooele Co., UT
P. difficilis	Lincoln Co., NM; Cibola Co., NM; Zacatecas, Mexico
P. eremicus	Yuma Co., AZ; Dona Ana Co., NM; Soccorro Co., NM
P. eva	Baja California Sur, Mexico[b,c] (2)
P. fraterculus	San Diego Co., CA
P. gossypinus	Decatur Co., TN; Bradley Co., AR; Bowie Co., TX
P. leucopus	Avery Co., NC; Antelope Co., NE; Lake Co., OH
P. maniculatus	Washtenaw Co., MI; Boulder Co., CO; Coconino Co., AZ; Mono Co., CA[a] (2); San Diego Co, CA
P. melanophrys	Zacatecas, Mexico; Durango, Mexico; Jalisco, Mexico
P. mexicanus	Guanacaste, Costa Rica; Francisco Morazan, Honduras[a] (2); Selva Negra, Nicaragua
P. polionotus	Santa Rosa Co., FL (2); Marion Co., FL; Lee Co., AL[c]; Lake Co., FL
P. truei	Durango, Mexico; Yavapai Co., AZ; Armstrong Co., TX
O. torridus	Kern Co., CA; Nye Co., NV

NOTE.—One individual was sampled at each site, except for those sites indicated with (2), where 2 individuals were sampled. Both *Zp2* and *Zp3* were sequenced except for those samples indicated below. See Supplementary Material online for sample sources and accession numbers.

[a] one individual from site sequenced for *Zp3* only.

[b] *Zp2* only.

[c] *Zp3* only.

After the entire *P. polionotus* nucleotide sequence for both fertilization genes was complete, *Peromyscus* specific primers were designed to amplify exons 8–10 of *Zp2* (2,102 bp) and exons 6–7 of *Zp3* (790 bp) (fig. 1). These regions were chosen because they contain several sites identified as targets of selection in an analysis of divergent mammalian species (Swanson, Yang et al. 2001). In addition, the region chosen for *Zp3* contains the sperm-combining region, which is necessary for ZP3's role in fertilization (Wassarman and Litscher 2001). PCR was performed under standard conditions (Supplementary Material online).

In order to determine whether phylogenies for the fertilization proteins are representative of species relationships, we sequenced 2 nonreproductive nuclear genes, the melanocortin-1 receptor (*Mc1r*) and lecithin cholesterol acyl transferase (*Lcat*). Mc1r is a G-protein–coupled receptor involved in pigmentation (Barsh 1996). Lcat is an enzyme in the glycerophospholipid metabolism pathway (Kuivenhoven et al. 1997). An 869-bp fragment containing most of the single-exon–coding region of *Mc1r* and

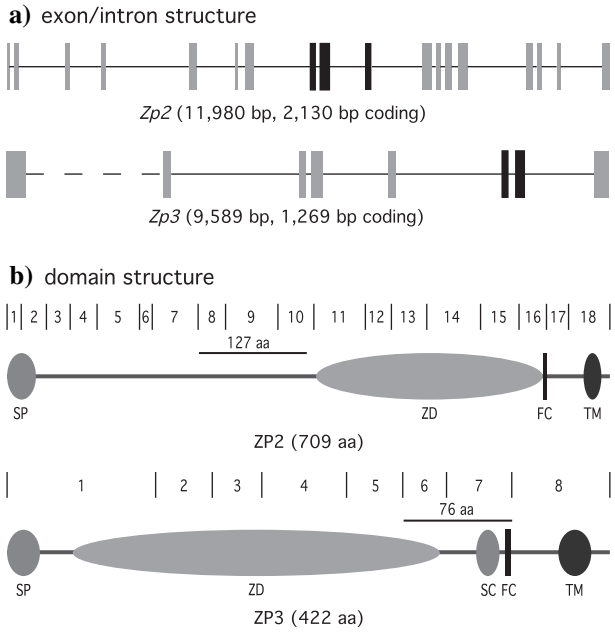

FIG. 1.—(*a*) Exon/intron structure of fertilization protein genes in *Peromyscus polionotus* drawn to scale within each gene. Boxes indicate exons; areas amplified for interspecific analysis are in black. Intron 1 of *Zp3* is dashed because this region was not fully sequenced; length was estimated from *Mus*. (*b*) Protein structure with predicted functional domains based on alignment with *Mus* and *Rattus* (Akatsuka et al. 1998; Jovine et al. 2004). Domains are drawn to scale within each protein. Region amplified for interspecific analysis is indicated with a black bar. Exon boundaries are indicated above with vertical bars and exon numbers between them. Abbreviations for domains are: SP, signal peptide; ZD, zona domain; FC, furin cleavage site; TM, transmembrane domain, and SC, sperm-combining region.

a 487-bp fragment containing most of exon 6 of *Lcat* were amplified under standard conditions (Supplementary Material online) using published primers (Robinson et al. 1997; Nachman et al. 2003).

PCR products were purified using a MinElute PCR purification kit (Qiagen) or a PerfectPrep PCR cleanup 96 kit (Eppendorf, Westbury, NY) if a single band was present. If multiple bands were present, PCR products were purified using the MinElute Gel Extraction kit (Qiagen). Purified PCR products were directly sequenced on an ABI 3100 automated sequencer (Applied Biosystems, Foster City, CA) using both PCR amplification primers and internal sequencing primers. For *Zp2* and *Zp3*, if an individual was heterozygous at more than one site, PCR products were cloned (TOPO-TA, Invitrogen, Carlsbad, CA) and sequenced using T7 and T3 primers to determine phase. Base calls were confirmed by eye, and sequences were aligned in SEQUENCHER. Coding region sequences analyzed for each gene included: *Zp2* (381 bp, 127 aa), *Zp3* (228 bp, 76 aa), *Mc1r* (756 bp, 252 aa), and *Lcat* (445 bp, 148 aa). Sequences were deposited in GenBank (Accession numbers DQ482843–DQ482899; DQ668051–DQ668343).

Phylogenetic Reconstruction

Bayesian gene trees were constructed using MrBayes (Huelsenbeck and Ronquist 2001, General Time Reversible [GTR] + Γ, partitioned by position in codon, 10 million generation Markov Chain Monte Carlo [MCMC]) both with and without outgroup sequences included. The first 500,000 generations were excluded as burn-in. *Mc1r* and *Lcat* were concatenated, and the data set was partitioned by both gene and position in codon; separate trees were generated using sequences from the same individuals included in the *Zp2* and *Zp3* data sets. We determined the appropriate model for each gene using hierarchical likelihood ratio tests (LRTs) comparing nested models (Huelsenbeck and Crandall 1997). Likelihoods of the resulting highest posterior probability tree were determined under alternative models available in MrBayes (nst = 1, 2, 6) using PAUP*(v.4b10, Swofford 2002). Trees were rooted using outgroup (*O. torridus*) sequences, if included. Neighbor-joining (NJ) and maximum likelihood (ML) trees were generated in PAUP* (GTR + Γ). In order to determine support values, ML analysis was repeated for 100 bootstrap data sets generated using the program SEQBOOT from the PHYLIP package (Felsenstein 2004). Because the gene data were partitioned by position in codon, we generated bootstrap data sets by resampling at the codon level rather than the nucleotide level.

Detection of Positive Selection

We tested for evidence of positive selection by comparing the nonsynonymous substitution rate (d_N) to the synonymous substitution rate (d_S). If a gene is evolving neutrally, $\omega = d_N/d_S$ is expected to equal one, whereas ω greater than one is considered strong evidence that a gene experiences positive selection. We used several ML approaches to test for evidence of positive selection on these fertilization proteins. The first approach, developed by Nielsen and colleagues (hereafter referred to as NY models), involves comparisons of a neutral codon substitution model with ω constrained to be <1 to a selection model where a class of sites has $\omega > 1$ (Nielsen and Yang 1998; Yang et al. 2000). As neutral models are nested within the corresponding selection models, a LRT can be used to compare them. The test statistic $-2\Delta\ln L$ ($\Delta\ln L$ = the difference in log likelihoods of the 2 models) follows a χ^2 distribution with degrees of freedom (df) equal to the difference in number of parameters between models. In the specific models implemented, ω varies between codons as a discrete (neutral: M0, M1; selection: M3, M2) or beta distribution (neutral: M7, M8A; selection: M8). We implemented models M0, M1, M2, M3, M7, and M8 (Wong et al. 2004) with the codeml program in PAML (v.3.14, Yang 2000). In order to account for uncertainty in the phylogeny, we performed the analysis using the 10 most probable trees from MrBayes as well as the NJ tree. Results of 3 model comparisons (M3 vs. M0, M2 vs. M1, M8 vs. M7) were consistent; here, we present data for the M8 versus M7 comparison, as this comparison is considered a more stringent test of positive selection (Yang and Nielsen 2002). We performed an additional test comparing results from M8 to a modified version of the model where the selection class has ω set to 1 (model M8A, Swanson et al. 2003). This test rules out the possibility that the neutral model is rejected because of a poor fit of the beta distribution for neutral and negatively selected sites. The test statistic follows a 50:50 mix of a χ^2 distribution with one df and a point mass at zero. Amino acid sites experiencing

positive selection were identified using the Bayes empirical Bayes (BEB) procedure (Yang et al. 2005). The BEB procedure is a modified version of an empirical Bayes' approach (Nielsen and Yang 1998) that identifies the most likely ω class for each codon site. Those sites that are most likely to be in the positive selection class ($\omega > 1$) are identified as likely targets of selection. The BEB procedure is an improvement over the previous approach as it takes into account sampling error in the ML estimates of parameters.

As our data include multiple alleles from each species, there is a possibility that recombination has occurred between alleles within species. In addition, if only a short time elapsed between speciation events, recombinant alleles from a polymorphic ancestor may have fixed in closely related species. Recombination can reduce the accuracy of the NY models (Anisimova et al. 2003) because different sites can have different phylogenetic histories. Specifically, differences in topology can result in patterns that look like recurrent substitution, and differences in branch lengths can result in variation in synonymous divergence among sites. We accounted for differences in topology in part by applying the NY models to multiple trees for each egg protein; however, it is possible that phylogenetic histories for all sites were not sampled. In order to address the issue of differences in branch lengths between sites, we applied additional methods to test for positive selection.

Although the NY models allow for variation in the nonsynonymous substitution rate, the synonymous rate is fixed across the sequence. Recently, several methods for detecting positive selection that allow for variation in synonymous rate have been proposed. These methods are new implementations of the 3 general classes of previous models, counting methods, fixed effects methods, and random effects methods. Counting methods map changes onto the phylogeny to estimate ω on a site-by-site basis. Kosakovsky Pond and Frost (2005b) propose a version called the single-likelihood ancestor counting (SLAC) method, which calculates the number of nonsynonymous and synonymous substitutions that have occurred at each site using ML reconstructions of ancestral sequences. Kosakovky Pond and Frost additionally introduce a version of a fixed effect approach, which estimates ω on a site-by-site basis. Their fixed effect likelihood (FEL) method uses ML estimation and treats shared parameters (branch lengths, tree topology, and nucleotide substitution rates) as fixed. The random effects likelihood (REL) method is similar to the NY model M3; however, both nonsynonymous and synonymous rates vary as gamma distributions with 3 rate classes (Kosakovsky Pond and Frost 2005b; Kosakovsky Pond and Muse 2005). The SLAC and FEL methods were implemented using the web interface DATAMONKEY (Kosakovsky Pond and Frost 2005a), and the REL method was implemented in HYPHY (Kosakovsky Pond et al. 2005).

Mapping of Amino Acid Substitutions

Nucleotide substitutions in both the reproductive genes *Zp2* and *Zp3* and nonreproductive genes *Mc1r* and *Lcat* were mapped onto the *Mc1r*/*Lcat* ML trees using maximum parsimony. Combined *Mc1r*/*Lcat* trees were used because of potential inaccuracies in the topology of gene trees for *Zp2* and *Zp3* due to parallel amino acid substitutions (see Results). In addition, substitutions in *Zp2* and *Zp3* were mapped using a Bayesian method (Nielsen 2002) with the program SIMMAP (Bollback 2006). Because, by definition, the parsimony method assumes that evolution has occurred in the fewest possible number of mutational steps, this approach provides a biased estimate. The degree to which parsimony underestimates the number of mutations depends on branch length and mutational parameters. The Bayesian method provides an advantage over parsimony because it accounts for uncertainty in the topology and model parameters by simulating mappings based on their probability of occurrence (Nielsen 2002). The Bayesian mappings were performed for *Zp2* and *Zp3* data sets that were modified such that, for sites that were variable within a species, only the derived state was included; this modification ensured that substitutions that were not fixed were not counted more than once and resulted in conservative estimates of the number of substitutions at these sites. Substitutions were mapped onto 1,000 samples from posterior distributions of trees generated in MrBayes based on both the data set for the gene and on the concatenated *Mc1r*/*Lcat* data sets (11 million generation MCMC, 1 million generation burn-in, GTR + Γ). We used the GTR + Γ model for mapping; mutational parameters were sampled from the posterior distribution for the *Zp2* and *Zp3* data sets. Ten realizations (mappings generated that are consistent with the data) were generated for each amino acid site for each of the 1,000 trees for each data set.

We performed an additional mapping analysis to determine if there was significant variation in rate of substitution across lineages. For this analysis, we focused on the branches on the tree where substitutions occur. As with parsimony mapping, we mapped changes onto the ML trees based on the *Mc1r*/*Lcat* data from the same individuals. Using SIMMAP, we determined the mean total number of nonsynonymous and synonymous substitutions that occurred on each branch over 1,000 realizations per codon. In order to determine if patterns of change were different from expectations (i.e., if there was no increase in rate of substitution for any particular branch), results for the observed data were compared with a null distribution based on 100 simulated data sets each generated from 1,000 realizations for each codon with the same mutational parameters. The rate class and starting state for each codon realization were determined by passing from the tips to the root of the tree and determining conditional likelihoods of rate/state at each node. States at the tips of the tree were then simulated using that rate category (Bollback 2006). Observed values were considered significantly different from expected if they fell outside 95% of the probability density of the simulated distribution.

Results

Structure and Sequence Variation of Egg Proteins

Intron/exon structure for both *Zp2* and *Zp3* is conserved between *Peromyscus* and *Mus* (fig. 1). Sequence identity between *P. polionotus* and *Mus* is 85% for *Zp2* and 84% for *Zp3*. Protein length is largely conserved with *Mus*; ZP2 is identical in length, and the few amino acid insertions/deletions in ZP3 (3 indels, 1–2 aa each) are very small. Conservation of length and ability to align the entire

amino acid sequence suggest that these proteins probably retain the domain structures predicted in *Mus*. There are, however, some potentially important differences between *Peromyscus* and *Mus* ZP3. Namely, numerous gains and losses of glycosylation sites have occurred; these changes may have functional consequences, as some evidence indicates glycosylation is critical to ZP3 function (Chen et al. 1998; but see Dean 2004). Three of 6 N-glycosylation sites found in *Mus* and *Rattus* were lost in *Peromyscus*, and 2 N-glycosylation sites in different positions were gained. In addition, 2 O-glycosylation sites (*Mus* Ser-332 and Ser-334), which have been identified as essential for sperm binding by ZP3 (Chen et al. 1998; but see Dean 2004), are conserved in mouse, rat, hamster, and human. One of these sites (331, homologous to *Mus* 332) has been lost in *Peromyscus*.

Phylogenetic Reconstruction

Mc1r/*Lcat* trees produced by ML, Bayesian, and NJ methods were consistent with each other with the exception of lineages within *Peromyscus maniculatus*; thus, only the ML tree is presented (fig. 2). Topologies were also consistent with published species trees based on morphological and molecular data (Avise et al. 1974; Rogers and Engstrom 1992; Tiemann-Boege et al. 2000). Species formed monophyletic groups with 2 exceptions: 1) a single *Peromyscus aztecus* individual fell outside the clade containing the other 2 *P. aztecus* and *Peromyscus boylii* individuals (this individual was placed outside the clade for *Zp2* and *Zp3* as well; thus, the taxonomic identity of that sample is uncertain) and 2) *Peromyscus leucopus* was paraphyletic. Similar to the *Mc1r*/*Lcat* phylogeny, the egg protein gene trees generated by different methods did not differ in topology, although some clades in the NJ tree were unresolved in the ML and Bayesian trees; thus, only the ML trees are presented (fig. 2). Strikingly, gene trees of *Zp2* and *Zp3* were not consistent with each other, with *Mc1r*/*Lcat* trees, or with published phylogenies. The topology of these gene trees may reflect cases where the same amino acid substitution occurred independently in more than one lineage (see Mapping of Amino Acid Substitutions). For example, the ML tree for *Zp2* groups 2 alleles from *Peromyscus truei* with the *maniculatus* and *leucopus* species groups, a relationship not consistent with any other phylogeny. Exclusion of sites that changed in parallel in multiple lineages resulted in topologies for *Zp2* and *Zp*3 that were more similar to *Mc1r*/*Lcat* trees and published phylogenies (data not shown).

Both *Zp2* and *Zp3* experienced numerous amino acid substitutions during the evolution of the *Peromyscus* genus. Twenty-five of 127 amino acid sites (19.7%) in exons 8–10 of *Zp2* (fig. 3*a*) and 22 of 76 sites (28.9%) in exons 6–7 of *Zp3* (fig. 3*b*) were variable, with several sites having multiple substitutions. For comparison, only 13.9% of sites in *Mc1r* and 5.4% of sites in exon 6 of *Lcat* were variable. However, overall estimates of ω for *Zp2* (0.38) and *Zp3* (0.31) were less than one, indicating that if these genes experienced positive selection, selection acted on a subset of amino acid sites.

Intraspecific Variation

Despite the limited number of alleles sampled for each species, we found extensive intraspecific amino acid variation in both *Zp2* and *Zp3*. For example, we identified 4 alleles of *Zp2* in 3 *P. aztecus* individuals with 5 variable amino acid sites. For *Zp3*, we found 3 alleles in 3 *P. truei* individuals, again with 5 variable amino acid sites. Alleles from a single species did not always form monophyletic groups, indicating alleles from different species were sometimes more similar than alleles within species.

Amino Acid Sites under Selection

Results from all 4 ML approaches for detecting selection indicated that a proportion of amino acid sites of both egg proteins have evolved adaptively. For *Zp2*, the LRTs comparing NY selection model M8 with neutral models (M7 and M8A) were significant ($P < 0.05$), with 1–2% of sites in the positively selected class with a mean $\omega = 7.93$ (range 7.77–9.27). The BEB procedure identified sites 239 and 321 as likely targets of positive selection. Results of the NY models for *Zp2* were consistent among analyses using the 10 most probable Bayesian trees and the NJ tree; significance of the LRTs and sites identified as targets of positive selection did not differ. Estimates of ω and posterior probabilities were similar for the 10 Bayesian trees but differed somewhat for the NJ tree; average values for the Bayesian trees are presented in table 2. For *Zp3*, LRTs comparing M8 with M8A were significant for all trees, and LRTs comparing M8 with M7 were significant for 4 of the 11 trees, and all tests had $P < 0.10$. Results averaged across the 10 most probable Bayesian trees are presented in table 2. The nonsignificance of the LRT comparing M8 with M7 in some cases may be due to the fact that the comparison of the test statistic to a χ^2 with 2 df is an approximation causing the test to be conservative, particularly for short, closely related sequences (Anisimova et al. 2001). Parameter estimates indicate that 2–3% of sites are in the positively selected class with a mean $\omega = 5.26$ (range 4.80–6.82). The BEB approach identified sites 343 and 345 as targets of positive selection for all analyses and site 316 in 4 of the 11 analyses.

The SLAC method did not identify any sites in ZP2 or ZP3 with evidence of positive selection significant at the $P < 0.05$ level; however, both sites identified with the BEB method in ZP2 (239 and 321) and one of the sites in ZP3 (345) had $P < 0.20$ of positive selection (table 2). Lack of significance at the 0.05 level is not surprising, as counting methods have low power with sequences of low divergence, and analyses of simulated data sets of similar size indicate that *P* values for the SLAC and FEL methods <0.20 have a true Type I error rate of $<5\%$ (Kosakovsky Pond and Frost 2005b). Using the FEL method, sites 239 and 321 of ZP2 and 345 of ZP3 were significant at the $P < 0.05$ level. An additional 3 sites in ZP3 were significant at the $P < 0.20$ level. The REL method also identified sites 239 and 321 of ZP2 as positively selected, although the posterior probability of selection for site 321 (0.76) was lower than with the BEB method (0.90). REL estimates of ω were higher than NY estimates for both sites. For ZP3, REL identified 15 sites with posterior probabilities >0.5 of positive selection (290, 295, 296, 308, 316, 320, 324, 326, 329, 330, 335, 337, 340, 343, and 345). However, in most cases, high ω values were due to low d_S rather than high d_N; of the 15, only sites 343 and 345 (sites identified by all BEB analyses)

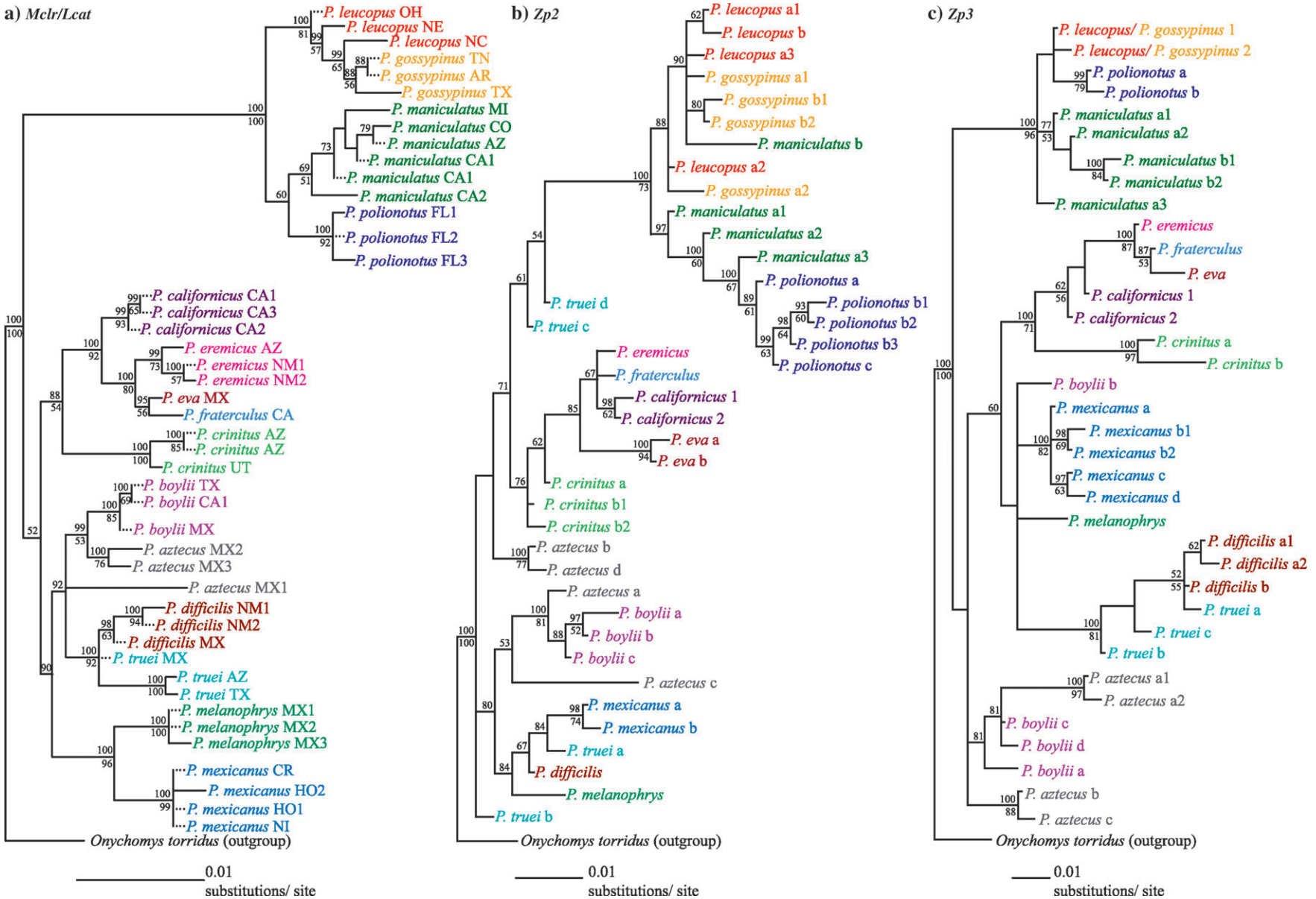

FIG. 2.—ML gene trees of (*a*) *Mc1r*/*Lcat*, (*b*) *Zp2*, and (*c*) *Zp3*. In (*a*), individuals are identified by species name and location, and dashed lines indicate branches with a length of zero. In (*b*) and (*c*), only unique alleles are included and are indicated by species name; letters indicate different coding alleles and numbers indicate alleles that differ by synonymous substitutions only. Species are coded by color. Scale for each gene is indicated below each tree. Posterior probabilities (%) >50 are given above and bootstrap percent values >50 are given below each branch.

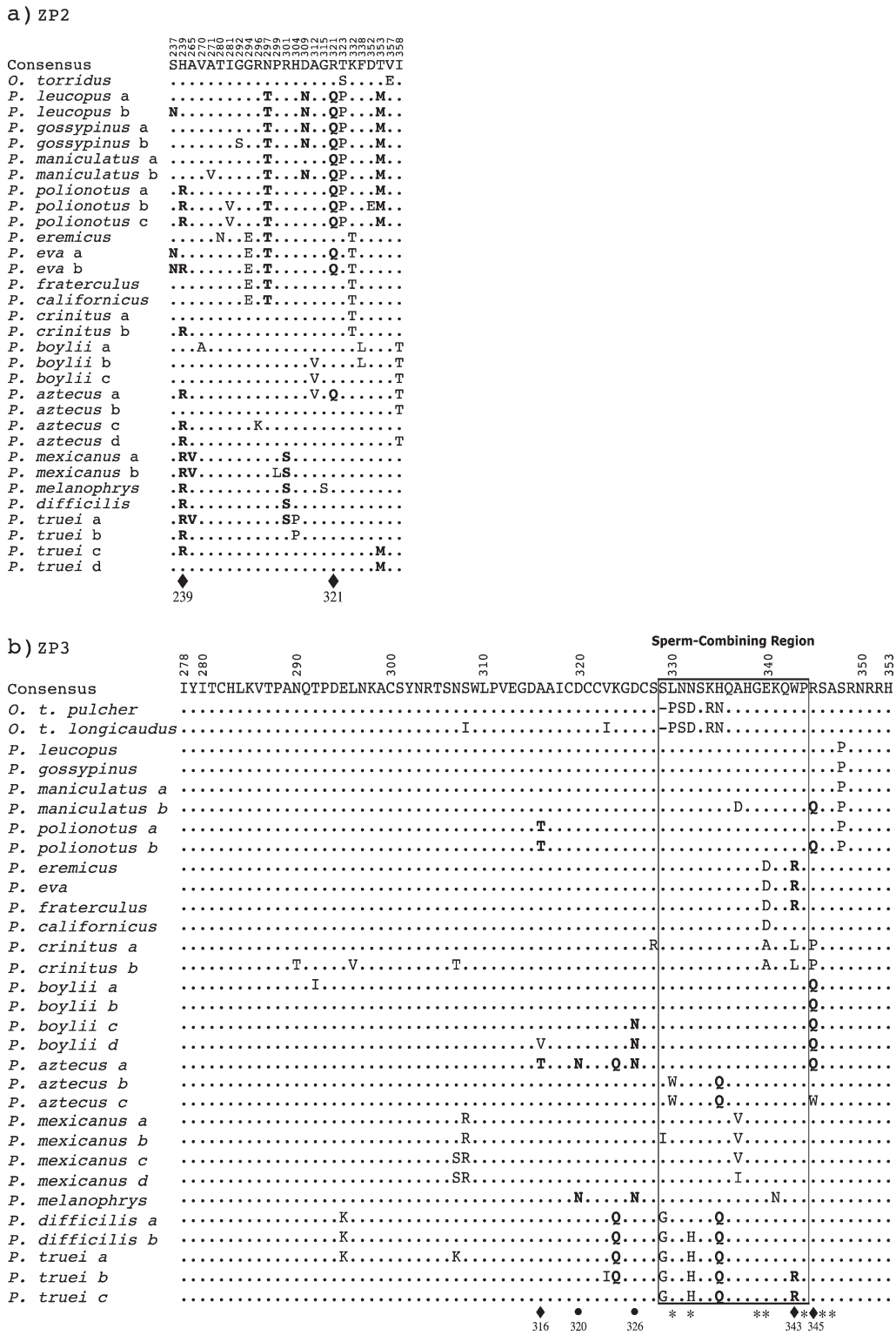

FIG. 3.—Amino acid sequence alignment of (*a*) variable amino acid sites in exons 8–10 (aa 236–362) of ZP2 and (*b*) all amino acid sites in exons 6–7 (aa 278–353) of ZP3; all sites are given for ZP3 in order to show spatial pattern of variable sites. In (*a*), numbering is consistent with *Mus* ZP2, and region in (*b*) aligns to aa 279–354 of *Mus* ZP3 (Boja et al. 2003). Alleles with identical amino acid sequences have been collapsed into a single haplotype, unless they are from different species. *Onychomys torridus* sequences are included as the outgroup. Dots indicate identity with the consensus sequence. Diamonds indicate amino acid sites that have a posterior probability >50% of being under positive selection in *Peromyscus* in one or more BEB analysis; amino acid site numbers are below the diamonds. Amino acids that have been substituted independently in more than one *Peromyscus* lineage are in bold. In (*b*), the box delimits the sperm-combining region (Kinloch et al. 1995), asterisks indicate sites that had a posterior probability >50% of being under positive selection in a phylogenetically diverse set of mammals (Swanson, Yang et al. 2001), and closed circles indicate 2 sites that had correlated change in both *Peromyscus aztecus* and *Peromyscus melanophrys*.

Table 2
Positive Selection on Egg Coat Proteins in *Peromyscus*

Gene	N	L_C	S	d_N/d_S	Selection Parameters	P (M8 vs. M7)	P (M8 vs M8A)	Positively Selected Sites (M8)	Positively Selected Sites (SLAC)	Positively Selected Sites (FEL)	Positively Selected Sites (REL)
Zp2	88	127	0.64	0.38	$p_s = 0.02$ $\omega_s = 7.77$	**0.016**	**0.002**	239 (0.99) $\omega = 5.52$	239 (0.07)	239 (**0.03**)	239 (0.99) $\omega = 13.55$
								321 (0.90) $\omega = 5.13$	321 (0.16)	321 (**0.05**)	321 (0.76) $\omega = 10.57$
Zp3	96	76	1.15	0.31	$p_s = 0.02$ $\omega_s = 5.35$	0.059	**0.011**	316 (0.52) $\omega = 2.09$		316 (0.07)	316 (0.81) $\omega = 3.41$
										329 (0.20) 337 (0.15)	
								343 (0.64) $\omega = 2.86$			343 (0.70) $\omega = 3.69$
								345 (0.99) $\omega = 4.50$	345 (0.06)	345 (**0.03**)	345 (0.98) $\omega = 13.65$

NOTE.—N, number of alleles sequenced; L_C, length of sequence in codons; S, tree length in substitutions per codon; d_N/d_S, ratio averaged across all sites and lineages (estimated with PAML, M0 [Yang 2000]); Selection parameters: average values from M8 estimated in PAML using 10 most probable Bayesian trees, p_s, proportion of sites in the "$\omega > 1$" class; ω_s, ω estimate for that class; P, P value of LRT comparing models listed, significant values (0.05 level) are *bold*; positively selected sites (M8): sites with posterior probability >0.5 of being in positively selected class in any analysis given as amino acid position, posterior probability that the site experiences positive selection, and the average ML estimate of ω at the site. Positively selected sites (SLAC)/(FEL): sites (and P values) identified using the SLAC and fixed effects likelihood (FEL) methods (Kosakovsky Pond and Frost 2005b) with P values consistent with a Type I error rate less than 5% ($P < 0.20$). Positively selected sites (REL): results of the REL method for sites identified by BEB procedure are given as amino acid position, posterior probability that the site experiences positive selection, and the estimate of ω at the site.

were assigned to the class with the highest d_N. Those 2 sites and the third site identified in some of the BEB analyses (316) also had the highest estimates of ω. Thus, the results were relatively consistent with BEB, although as for ZP2, estimates of ω were higher for each site.

In summary, all 4 ML approaches identified sites 239 and 321 of ZP2 and site 345 of ZP3 as likely targets of selection. For ZP3, an additional 2 sites (316 and 343) were identified by some, but not all methods, as targets of selection.

Mapping of Amino Acid Substitutions

The pattern of amino acid change based on parsimony mapping provides further evidence that ZP2 and ZP3 evolved under positive selection (fig. 4). Eight amino acid sites in ZP2 and 7 sites in ZP3 changed independently to the same amino acid in 2 or more *Peromyscus* lineages (table 3). For example, site 239 in ZP2 changed from arginine (R) to histidine (H) in 3 different *Peromyscus* lineages, and the reverse change occurred in 4 lineages. In ZP3, site 345 changed from arginine (R) to glutamine (Q) in 3 lineages. This change, although classified as conservative based on Grantham's distance, which takes into account amino acid size, hydrophobicity, charge, and polarity, was a change from a positively charged to a noncharged residue. Parallel evolution at the amino acid sequence level can be interpreted as evidence of adaptive evolution (Zhang 2003); consequently, sites that have changed in parallel are likely targets of selection in addition to those identified with the ML approach.

There are 2 additional sites in ZP3 (337 and 340) with patterns of substitution consistent with positive selection. Both of these sites had 2 substitutions in a single lineage over a relatively short period of time. For example, site 340 changed from glutamic acid (E) to aspartic acid (D) in the clade containing *californicus/eremicus/eva/fraterculus/crinitus* and subsequently from aspartic acid (D) to alanine (A) in *Peromyscus crinitus*.

As previously mentioned, sequence comparisons between closely related species introduces the possibility of recombination, which could result in patterns that look like recurrent substitution. To address this potential problem, we compared patterns of parallel substitution between the egg protein genes and the nonreproductive genes (*Mc1r* and *Lcat*). Recombination within extant species or a polymorphic ancestor is expected to generate similar patterns for both sets of genes, assuming they experience similar recombination rates, and would affect patterns of both nonsynonymous and synonymous substitutions in a similar manner. First, our results show that parallel amino acid substitutions were rare or absent in the nonreproductive genes. Second, the ratio of sites with parallel nonsynonymous substitutions to sites with parallel synonymous substitutions was significantly higher in reproductive genes than in nonreproductive genes (Fisher's exact test, $P = 0.002$). In addition, proportions of parallel sites did not differ significantly between genes within each class. These results support the supposition that positive selection, rather than recombination, is the cause of parallel patterns of amino acid substitutions in ZP2 and ZP3.

In order to determine potential functional consequences of amino acid substitutions, we examined the spatial pattern of nucleotide substitution in *Zp2* and *Zp3*. In addition to considering the location of adaptively evolving sites, overall patterns of synonymous and nonsynonymous change across the sequenced regions were determined through Bayesian mapping (fig. 5). Amino acid substitutions in ZP2 were not localized in any one region, and there was no clustering of the sites identified as positively selected by the ML methods (239 and 321) or sites that experienced parallel changes (fig. 3*a*). These results are consistent with the dispersed pattern of sites identified as positively selected in an analysis of ZP2 in a diverse set of mammals (Swanson, Yang et al. 2001). In addition, the specific functional roles of different domains of ZP2 are not well characterized; therefore, it is difficult to predict whether changes at these

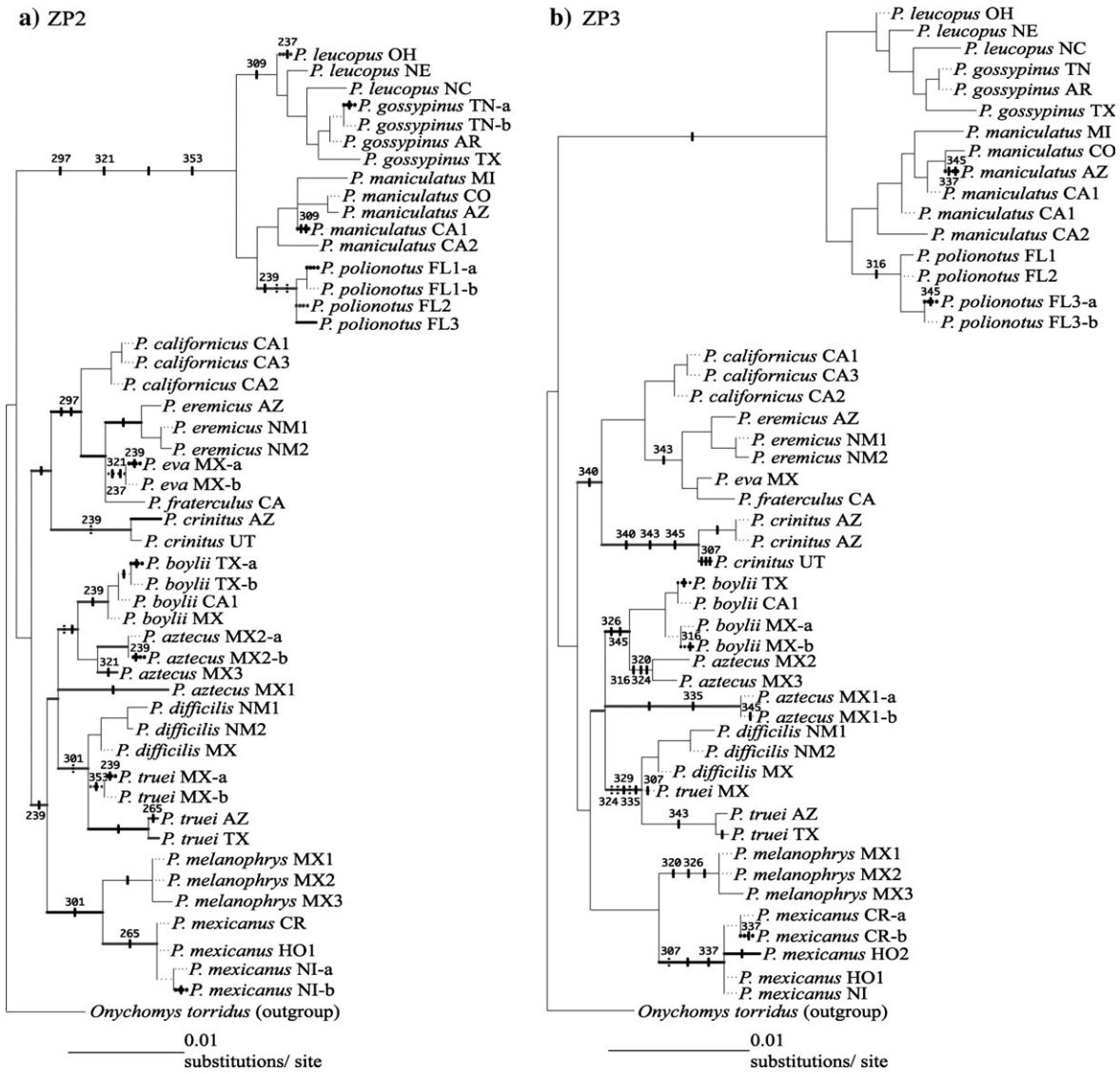

FIG. 4.—Amino acid substitutions in egg proteins mapped onto *Mc1r*/*Lcat* gene trees, using parsimony and a Bayesian method for the same individuals included in the (*a*) ZP2 and (*b*) ZP3 data sets. Scale for each gene is indicated below each tree. Dashed branches have a length of zero. Each vertical bar represents a single amino acid substitution determined by parsimony. Substitutions at sites that have changed independently in more than one lineage have the amino acid position indicated. As determined by Bayesian analysis, branches with significantly higher nonsynonymous substitution rates are in bold black, and those where synonymous rates are also elevated are in bold gray.

sites might impact egg–sperm binding. In contrast, amino acid substitutions in ZP3 were concentrated in and around the region homologous to the *Mus* sperm-combining site (fig. 5). In addition, sites identified as positively selected by more than one ML method (316, 343, and 345) and sites that have changed in parallel clustered in this region (fig. 3*b*). Interestingly, these sites neighbor but are not the same as those sites identified as positively selected in ZP3 in more divergent mammalian taxa (Swanson, Yang et al. 2001).

For both ZP2 and ZP3, the amount of amino acid change varied across lineages, both in absolute terms and in relation to the amount of synonymous change (fig. 4). Differences are apparent when comparing the number of amino acid substitutions on each branch determined by parsimony to branch length determined by overall substitution. For example, in ZP3, parsimony mapping suggests that there have been 3 amino acid substitutions in the *P. crinitus* lineage, and 4 sites were variable within the species, yet only one substitution has occurred in the rest of the clade. In ZP2, a different pattern is observed for this clade: *P. crinitus* has not fixed any substitutions, but there have been 5 substitutions in the rest of the clade. In addition, we did not observe a consistent pattern of decrease in substitution rate in ZP2 and ZP3 in the 2 monogamous taxa, *P. polionotus* and *Peromyscus californicus*.

Bayesian mapping allowed us to test whether elevated rates of amino acid substitution in some lineages were significantly different from neutral expectations. In general, results from the Bayesian approach were in agreement with patterns inferred from parsimony results, but there were cases where 2 branches of similar length had the same number of nonsynonymous changes but one was significantly

Table 3
Parallel Amino Acid Substitutions in ZP2 and ZP3

Change	Number of Independent Changes	Starting amino acid Class	Ending amino acid Class	Charge Changing	Grantham's distance	Type of Change	Possible Alternative amino acid Substitutions
ZP2							
S237N	2	P	P	N	46	C	5
R239H	**3**	+	+	**N**	**29**	**C**	**4**
H239R	**4**	+	+	**N**	**29**	**C**	**6**
A265V	2	NP	NP	N	64	MC	5
N297T	2	P	P	N	65	MC	6
R301S	2	+	P	Y	110	MR	4
D309N	2	−	P	Y	23	C	6
R321Q	**3**	**P**	+	**Y**	**43**	**C**	**3**
T353M	2	NP	P	N	81	MC	3
ZP3							
A316T	2	NP	P	N	58	MC	5
D320N	2	−	P	Y	23	C	6
K324Q	2	+	P	Y	53	MC	5
D326N	2	−	P	Y	23	C	6
H335Q	2	+	P	Y	24	C	6
W343R	**2**	**NP**	+	**Y**	**101**	**MR**	**4**
R345Q	**3**	+	**P**	**Y**	**43**	**C**	**5**

NOTE.—Number of independent changes determined by parsimony mapping of ZP2 and ZP3 amino acid substitutions onto *Mc1r*/*Lcat* ML trees. Amino acid types: ''+,'' positively charged; ''−,'' negatively charged; P, polar; NP, nonpolar. Grantham's distance between starting and ending amino acid (Grantham 1974). Types of change: C, conservative (Grantham's distance <50); MC, moderately conservative (51–100); MR, moderately radical (101–150); R, radical (>150) (Li et al. 1984). Possible alternative amino acid substitutions: number of possible amino acid substitutions given the starting codon and a single nucleotide mutation. Rows in bold are sites identified as likely targets of selection by the NY codon models/BEB procedure.

elevated and the other was not (fig. 4). This discrepancy is the result of substitution parameters that are not taken into account by parsimony mapping, such as the rate category for the codon and the type of nucleotide substitution that occurred. For example, rates of change from C to T in *Zp3* were approximately 5 times higher than rates of change from A to C; branches with several substitutions that tend to be rare were more likely to be identified as having elevated rates by the Bayesian method. For *Zp2*, there were several branches that had high nonsynonymous rates but no amino acid substitutions as determined by parsimony. These were cases where mean rates for all realizations were very low but nonzero due to a small proportion of realizations in SIMMAP that were inconsistent with parsimony. If the means for all simulated data sets were zero, then the very small values for the observed data were significantly elevated. Several branches had elevated nonsynonymous rates in both *Zp2* and *Zp3*, including the branches leading to the *boylii*/*aztecus* and the *truei*/*difficilis* lineages. This pattern suggests that, although these 2 egg proteins are involved in different stages of the fertilization process (Wassarman and Litscher 2001), selection may have acted on both proteins in the same lineages.

Discussion

In the past decade, rapid evolution of reproductive proteins has been documented in a wide variety of taxa (Swanson and Vacquier 2002a). In internally fertilized species, research on patterns of evolution of reproductive proteins in closely related taxa has been primarily limited to *Drosophila* species (but see Jansa et al. 2003). Here, we provide strong evidence that the egg proteins ZP2 and ZP3 have experienced positive selection in the *Peromyscus* genus. In addition, we identify specific amino acid sites in ZP2 and ZP3 that are likely targets of selection. In ZP3, these sites are clustered in and around the functionally important sperm-combining region. We show that some amino acids changed in parallel in multiple lineages, providing further support that these changes are adaptive and suggesting that the number of available pathways of adaptive evolution may be constrained. Finally, using a Bayesian method to map amino acid substitutions, we identify lineages with elevated rates of nonsynonymous change in both ZP2 and ZP3. These data confirm that patterns of evolution of reproductive proteins across mammals are reflective of processes at lower taxonomic levels and suggest future avenues of investigation to characterize the potential functional consequences of amino acid change on fertilization potential.

For both ZP2 and ZP3, we have identified several species that have variation in amino acid sequence. In some cases, alleles were not monophyletic with respect to species. This pattern could be a result of incomplete lineage sorting; however, the same individuals were monophyletic or unresolved with respect to species for the autosomal, nonreproductive genes *Mc1r* and *Lcat*. This pattern suggests that selection may be maintaining divergent *Zp2* and *Zp3* alleles within species. However, more detailed intraspecific analysis is needed to confirm and explain this result. Similar patterns of extensive polymorphism and divergence have been found in other adaptively evolving reproductive proteins, including the sea urchin sperm protein bindin and *Drosophila* Acps, whereas other reproductive proteins, including abalone lysin, appear to have experienced

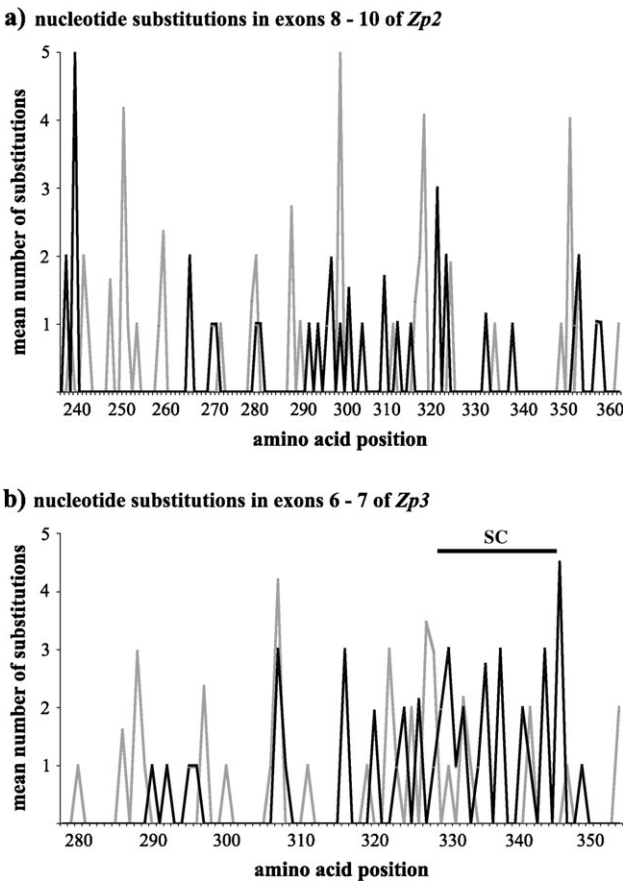

FIG. 5.—Patterns of nucleotide substitution in (a) *Zp2* and (b) *Zp3*. Substitutions were mapped onto *Mc1r*/*Lcat* trees using the same individuals included in each gene data set. Bayesian mapping was performed using SIMMAP. Mean numbers of substitutions are shown from 10 realizations per codon per tree for each of 1,000 samples from a posterior distribution of trees generated in MrBayes (10 million generation MCMC, 5 million generation burn-in, GTR + Γ). Gray lines indicate synonymous substitutions, and black lines indicate nonsynonymous substitutions. The black bar labeled "SC" indicates the sperm-combining region of *Zp3* (Kinloch et al. 1995).

selective sweeps resulting in very little intraspecific variation (Swanson and Vacquier 2002b).

In addition to the ML approaches, the parallel pattern of change at several sites in ZP2 and ZP3 provided evidence that these proteins have evolved adaptively. Parallel or convergent evolution at the amino acid sequence level can be interpreted as evidence of adaptive evolution; examples include lysozymes of cows and langurs (Stewart et al. 1987), butterfly and vertebrate opsins (Briscoe 2001), and HIV envelope protein genes between different lineages within a patient (Holmes et al. 1992). However, some sites that had parallel changes were not identified as targets of positive selection through the ML approaches, which assume that ω is consistent through time at a particular codon. The expectation that selective pressure remains constant is unrealistic; however, statistical methods that account for variation in ω both among codons in a sequence and through time require a large amount of variation and have thus far been applied successfully only to evolution of viral sequences, where rates of evolution are exceptionally high (Guindon et al. 2004). The parallel pattern of amino acid change allowed us to identify sites that are likely targets of positive selection, but where, response to selection was limited to specific lineages and/or to specific times during the *Peromyscus* radiation.

This repeated pattern of amino acid change suggests that there may be a finite number of ways to change adaptively. If all substitutions that occurred multiple times were conservative in terms of amino acid properties, we might infer that this pattern is a result of the negative consequences of radical change, even in the context of positive selection. However, for both ZP2 and ZP3, several repeated changes were not conservative, as defined by changes in charge or by Grantham's distance (Grantham 1974). Such nonconservative changes have been found to occur much less frequently than expected under neutrality (Li et al. 1984). Thus, the nonconservative changes we observed seem more likely to have consequences for protein structure and/or function.

In addition to parallel changes at single sites across taxa, we also observed correlated amino acid change in 2 sites that occurred in independent lineages. Two substitutions at sites 320 and 326 of ZP3, both from aspartic acid (D) to asparagine (N), occurred in the *P. aztecus* and *Peromyscus melanophrys* lineages (fig. 3*b*). This pattern is intriguing as, in addition to the fact that these substitutions are charge changing, the change at site 326 created a potential N-glycosylation site. In fact, this site is also an N-glycosylation site in *Mus* and *Rattus* and is known to be occupied in *Mus* (Boja et al. 2003). Evidence indicates that glycosylation of ZP3 in *Mus* is required for sperm binding (Chen et al. 1998; but see Dean 2004); consequently, changes at glycosylation sites may have a direct impact on egg–sperm binding.

Variation in the amount of nonsynonymous change that has occurred in different *Peromyscus* lineages suggests that the selective forces acting on these genes have not remained the same throughout the evolution of the genus. Patterns of variation within species and between members of a sister species pair varied across taxa (fig. 3). For example, there were differences in ZP3 between the sister species *P. maniculatus* and *P. polionotus* as well as variation within each species. In contrast, the amino acid sequence was identical for all samples of the sister species pair *P. leucopus* and *Peromyscus gossypinus*, which share a similar divergence time with *P. maniculatus*/*P. polionotus* (Blair 1950).

Application of a Bayesian method for mapping nucleotide substitutions allowed us to identify variation in substitution rate both along the length of each gene and between lineages. Although these patterns can be inferred by examining sequence alignments and by parsimony mapping of substitutions, the Bayesian method provides a quantitative estimate of the amount of change that has occurred and allows statistical tests of elevated lineage-specific substitution rates. Using this approach, we identified several branches with significantly elevated rates of amino acid substitution in both ZP2 and ZP3.

In internally fertilized species, sperm competition and sexual conflict have been proposed as important factors driving reproductive protein evolution (Wyckoff et al. 2000; Price et al. 2001; Swanson, Yang et al. 2001; Torgerson et al. 2002). Thus, variation in rates of evolution between species with different mating systems is predicted. Specifically, monogamous species may have lower rates

of reproductive protein evolution because of the lack of sperm competition and reduced sexual conflict. For example, rates of evolution of 2 genes encoding semen proteins that are components of the mating plug are correlated with female promiscuity in primates (Kingan et al. 2003; Dorus et al. 2004), and rates of *Acp* evolution are higher in *Drosophila* species with higher remating rates (Wagstaff and Begun 2005). Variation in mating system is found in *Peromyscus*; although most *Peromyscus* species are promiscuous, monogamy has evolved independently in 2 of the species sampled in this study, *P. californicus* and *P. polionotus* (Kleiman 1977; Foltz 1981; Ribble 1991, 2003). We did not find a consistent reduction of rate of evolution of ZP2 or ZP3 in the monogamous taxa. Although *P. californicus* has had very little change in either of the proteins, the *P. polionotus* lineage has had multiple amino acid substitutions in both ZP2 and ZP3 (fig. 4). Proteins involved in sperm morphology and performance may be more appropriate candidates to detect evidence of mating system effects on evolutionary rates.

Conclusions

Recent empirical and theoretical studies suggest that rapid evolution of reproductive proteins may play an important role in the evolution of reproductive isolation (Price et al. 2001; Servedio 2001; Coyne and Orr 2004). Allopatric populations that have limited overall phenotypic divergence may have significant divergence of reproductive proteins, leading to postmating, prezygotic reproductive incompatibilities upon secondary contact (gametic isolation), and potentially resulting in reinforcement. In marine invertebrates, patterns of evolution of egg–sperm binding proteins suggest that changes have potentially contributed to reinforcement and reproductive isolation (Galindo et al. 2003; Geyer and Palumbi 2003). However, in internally fertilized species, evidence for adaptive evolution of reproductive proteins within a genus is limited to *Drosophila* (Begun et al. 2000; Swanson, Clark et al. 2001; Swanson et al. 2004; Wagstaff and Begun 2005) and *Mus* (Jansa et al. 2003, but see Supplementary Material online).

Tests for positive selection that do not require population samples, and those which can identify specific sites subject to positive selection, generally have been applied to higher level taxa, where amino acid variation is more likely to be sufficient to significantly reject neutral models (Anisimova et al. 2002; Yang and Nielsen 2002). Here, we used a combination of several ML approaches and parallel patterns of substitution to detect selection and identify the specific amino acid sites that are evolving adaptively. Our results documenting positive selection acting on ZP2 and ZP3 within a genus confirm that we can successfully extend work documenting adaptive evolution of reproductive proteins across mammals (Wyckoff et al. 2000; Swanson, Yang et al. 2001; Torgerson et al. 2002; Jansa et al. 2003; Swanson et al. 2003) to look at how these proteins have changed between closely related species, where isolating barriers act and have evolved recently. It is certainly possible that adaptive change of ZP2 and ZP3 did not contribute to reproductive isolation between *Peromyscus* species, either because changes were not sufficient to prevent fertilization or because other isolating barriers had evolved before the egg proteins had diverged sufficiently to cause incompatibilities. More detailed intraspecific analysis is necessary to determine if differences in *Zp2* or *Zp3* genotype correlate with incipient reproductive isolation between populations.

In order to confirm a role of reproductive protein evolution in gametic isolation, functional consequences of amino acid change on fertilization potential must be determined. Here, we identified specific amino acid sites likely to be targets of selection in ZP2 and ZP3. Particularly appropriate for functional studies are the adaptively evolving sites in ZP3, which are clustered in and around the region known to be critical to successful egg–sperm binding in *Mus* (Wassarman and Litscher 2001). Evidence that sites in or around this region have evolved adaptively has been found previously in analysis of a taxonomically diverse set of mammals (Swanson, Yang et al. 2001). Interestingly, the specific amino acid sites identified here are adjacent but not identical to the sites evolving adaptively across mammals, underscoring the value of examining evolutionary processes at multiple taxonomic levels. These data identify this region of ZP3, and the positively selected sites specifically, as promising targets for future functional assays of allelic differences in sperm-binding ability.

Supplementary Material

Supplementary materials are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

The authors wish to thank J. R. Kohn,V. D. Vacquier, and 3 anonymous reviewers for helpful discussion and comments on the manuscript. J. P. Bollback generously provided a modified version of SIMMAP. J. P. Huelsenbeck and P. Andolfatto provided guidance with data analysis. E. Chuong and F. Kondrashov wrote perl scripts to summarize mapping and codon bootstrap results. R. J. Fisher, L. M. Mullen, and B. J. Haeck assisted with fieldwork. G. Arnaud, M. Kalcounis-Rueppell, M. Nachman, K. Nutt, J. Storz, A. Suazo, M. Wooten, the M. L. Bean Life Science Museum (Brigham Young University), the Museum of Southwestern Biology (University of New Mexico), the Museum of Texas Tech University, the *Peromyscus* Genetic Stock Center, and the Museum of Vertebrate Zoology (University of California, Berkeley) kindly provided tissue samples. This research was supported by a Howard Hughes Medical Institute Predoctoral Fellowship, the American Society of Mammalogists Grants-in-Aid of Research, the University of California Reserve System Mildred Mathias Award (to L.M.T.), National Science Foundation (NSF) doctoral dissertation improvement grant DEB 0608030 (to L.M.T. and H.E.H.), and NSF DEB 0344710 (to H.E.H.).

Literature Cited

Aguade M. 1999. Positive selection drives the evolution of the Acp29AB accessory gland protein in *Drosophila*. Genetics 152:543–51.

Akatsuka K, Yoshida-Komiya H, Tulsiani DRP, Orgebin-Crist M-C, Hiroi M, Araki Y. 1998. Rat zona pellucida glycoproteins: molecular cloning and characterization of the three major components. Mol Reprod Dev 51:454–67.

Anisimova M, Bielawski JP, Yang Z. 2002. Accuracy and power of Bayes prediction of amino acid sites under positive selection. Mol Biol Evol 19:950–8.

Anisimova M, Bielawski JP, Yang ZH. 2001. Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. Mol Biol Evol 18:1585–92.

Anisimova M, Nielsen R, Yang ZH. 2003. Effect of recombination on the accuracy of the likelihood method for detecting positive selection at amino acid sites. Genetics 164:1229–36.

Avise JC, Smith MH, Selander RK. 1974. Biochemical polymorphism and systematics in the genus *Peromyscus* part 6: the *Peromyscus boylii* group. J Mammal 55:751–63.

Barsh GS. 1996. The genetics of pigmentation: from fancy genes to complex traits. Trends Genet 12:299.

Begun DJ, Lindfors HA. 2005. Rapid evolution of genomic *Acp* complement in the *melanogaster* subgroup of *Drosophila*. Mol Biol Evol 22:2010–21.

Begun DJ, Whitley P, Todd BL, Waldrip-Dail HM, Clark AG. 2000. Molecular population genetics of male accessory gland proteins in *Drosophila*. Genetics 156:1879–88.

Blair WF. 1950. Ecological factors in speciation of *Peromyscus*. Evolution 4:253–75.

Boja ES, Hoodbhoy T, Fales HM, Dean J. 2003. Structural characterization of native mouse zona pellucida proteins using mass spectrometry. J Biol Chem 278:34189–202.

Bollback JP. 2006. SIMMAP: stochastic character mapping of discrete traits on phylogenies. BMC Bioinformatics 7:88.

Briscoe AD. 2001. Functional diversification of lepidopteran opsins following gene duplication. Mol Biol Evol 18: 2270–9.

Chen J, Litscher ES, Wassarman PM. 1998. Inactivation of the mouse sperm receptor, mZP3, by site-directed mutagenesis of individual serine residues located at the combining site for sperm. Proc Natl Acad Sci USA 95:6193–7.

Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, Higgins DG, Thompson JD. 2003. Multiple sequence alignment with the Clustal series of programs. Nucleic Acids Res 31: 3497–500.

Cirera S, Aguade M. 1997. Evolutionary history of the sex-peptide (*Acp70A*) gene region in *Drosophila melanogaster*. Genetics 147:189–97.

Civetta A, Singh RS. 1995. High divergence of reproductive tract proteins and their association with postzygotic reproductive isolation in *Drosophila melanogaster* and *Drosophila virilis* group species. J Mol Evol 41:1085–95.

Clark NL, Swanson WJ. 2005. Pervasive adaptive evolution in primate seminal proteins. PLoS Genetics 1:335–42.

Coyne JA, Orr HA. 2004. Speciation. Sunderland, MA: Sinauer Associates.

Dean J. 2004. Reassessing the molecular biology of sperm-egg recognition with mouse genetics. Bioessays 26:29–38.

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: 1326–9.

Eisenbach M, Giojalas LC. 2006. Sperm guidance in mammals—an unpaved road to the egg. Nat Rev Mol Cell Biol 7:276.

Felsenstein J. 2004. PHYLIP (phylogeny inference package). Version 3.6. Seattle, WA: Department of Genome Sciences, University of Washington.

Foltz DW. 1981. Genetic evidence for long-term monogamy in a small rodent, *Peromyscus polionotus*. Am Nat 117: 665–75.

Galindo BE, Vacquier VD, Swanson WJ. 2003. Positive selection in the egg receptor for abalone sperm lysin. Proc Natl Acad Sci USA 100:4639–43.

Geyer LB, Palumbi SR. 2003. Reproductive character displacement and the genetics of gamete recognition in tropical sea urchins. Evolution 57:1049–60.

Grantham R. 1974. Amino-acid difference formula to help explain protein evolution. Science 185:862–4.

Guindon S, Rodrigo AG, Dyer KA, Huelsenbeck JP. 2004. Modeling the site-specific variation of selection patterns along lineages. Proc Natl Acad Sci USA 101:12957–62.

Holmes EC, Zhang LQ, Simmonds P, Ludlam CA, Brown AJL. 1992. Convergent and divergent sequence evolution in the surface envelope glycoprotein of human immunodeficiency virus type 1 within a single infected patient. Proc Natl Acad Sci USA 89:4835–9.

Hooper ET. 1968. Classification. In: King JA, editor. Biology of *Peromyscus* (Rodentia). American Society of Mammalogists. p 27–74.

Huelsenbeck JP, Crandall KA. 1997. Phylogeny estimation and hypothesis testing using maximum likelihood. Annu Rev Ecol Syst 28:437–66.

Huelsenbeck JP, Ronquist F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17:754–5.

Jagadeeshan S, Singh RS. 2005. Rapidly evolving genes of *Drosophila*: differing levels of selective pressure in testis, ovary, and head tissues between sibling species. Mol Biol Evol 22:1793–801.

Jansa SA, Lundrigan BL, Tucker PK. 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, Ekhlasi-Hundrieser M, Toepfer-Petersen E. 2001. Sperm adhesion molecules: structure and function. Cells Tissues Organs 168:82–92.

Jovine L, Qi HY, Williams Z, Litscher ES, Wassarman PM. 2004. A duplicated motif controls assembly of zona pellucida domain proteins. Proc Natl Acad Sci USA 101:5922–7.

Kingan SB, Tatar M, Rand DM. 2003. Reduced polymorphism in the chimpanzee semen coagulating protein, semenogelin I. J Mol Evol 57:159–69.

Kinloch RA, Sakai Y, Wassarman PM. 1995. Mapping the mouse ZP3 combining site for sperm by exon swapping and site-directed mutagenesis. Proc Natl Acad Sci USA 92:263–7.

Kleiman DG. 1977. Monogamy in mammals. Q Rev Biol 52: 39–69.

Kosakovsky Pond SL, Frost SDW. 2005a. Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. Bioinformatics 21:2531–3.

Kosakovsky Pond SL, Frost SDW. 2005b. Not so different after all: a comparison of methods for detecting amino acid sites under selection. Mol Biol Evol 22:1208–22.

Kosakovsky Pond SL, Frost SDW, Muse SV. 2005. HyPhy: hypothesis testing using phylogenies. Bioinformatics 21: 676–79.

Kosakovsky Pond SL, Muse SV. 2005. Site-to-site variation of synonymous substitution rates. Mol Biol Evol 22:2375–85.

Kuivenhoven JA, Pritchard H, Hill J, Frohlich J, Assmann G, Kastelein J. 1997. The molecular pathology of lecithin cholesterol acyltransferase (LCAT) deficiency syndromes. J Lipid Res 38:191–205.

Li WH, Wu CI, Luo CC. 1984. Nonrandomness of point mutation as reflected in nucleotide substitutions in pseudogenes and its evolutionary implications. J Mol Evol 21:58–71.

Mah SA, Swanson WJ, Vacquier VD. 2005. Positive selection in the carbohydrate recognition domains of sea urchin sperm receptor for egg jelly (suREJ) proteins. Mol Biol Evol 22: 533–41.

Mueller JL, Ram KR, McGraw LA, Bloch Qazi MC, Siggia ED, Clark AG, Aquadro CF, Wolfner MF. 2005. Cross-species comparison of *Drosophila* male accessory gland protein genes. Genetics 171:131–43.

Nachman MW, Hoekstra HE, D'Agostino SL. 2003. The genetic basis of adaptive melanism in pocket mice. Proc Natl Acad Sci USA 100:5268–73.

Nielsen R. 2002. Mapping mutations on phylogenies. Syst Biol 51:729–39.

Nielsen R, Yang Z. 1998. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. Genetics 148:929–36.

Price CSC, Kim CH, Gronlund CJ, Coyne JA. 2001. Cryptic reproductive isolation in the *Drosophila simulans* species complex. Evolution 55:81–92.

Ribble DO. 1991. The monogamous mating system of *Peromyscus californicus* as revealed by DNA fingerprinting. Behav Ecol Sociobiol 29:161–6.

Ribble DO. 2003. The evolution of social and reproductive monogamy in *Peromyscus*: evidence from *Peromyscus californicus* (the california mouse). In: Reichard UH, Boesch C, editors. Monogamy: mating strategies and partnerships in birds, humans, and other mammals. Cambridge, UK: Cambridge University Press. p 161–6.

Robinson M, Catzeflis F, Briolay J, Mouchiroud D. 1997. Molecular phylogeny of rodents, with special emphasis on murids: evidence from nuclear gene *LCAT*. Mol Phylogenet Evol 8:423–34.

Rogers DS, Engstrom MD. 1992. Evolutionary implications of allozymic variation in tropical *Peromyscus* of the *mexicanus* species group. J Mammal 73:55–69.

Servedio MR. 2001. Beyond reinforcement: the evolution of premating isolation by direct selection on preferences and postmating, prezygotic incompatibilities. Evolution 55:1909–20.

Singh RS, Kulathinal RJ. 2000. Sex gene pool evolution and speciation: a new paradigm. Genes Genet Syst 75:119–30.

Stewart CB, Schilling JW, Wilson AC. 1987. Adaptive evolution in the stomach lysozymes of foregut fermenters. Nature 330:401–4.

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 98:7375–9.

Swanson WJ, Nielsen R, Yang Q. 2003. Pervasive adaptive evolution in mammalian fertilization proteins. Mol Biol Evol 20:18–20.

Swanson WJ, Vacquier VD. 2002a. The rapid evolution of reproductive proteins. Nat Rev Genet 3:137–44.

Swanson WJ, Vacquier VD. 2002b. Reproductive protein evolution. Annu Rev Ecol Syst 33:161–79.

Swanson WJ, Wong A, Wolfner MF, Aquadro CF. 2004. Evolutionary expressed sequence tag analysis of *Drosophila* female reproductive tracts identifies genes subjected to positive selection. Genetics 168:1457–65.

Swanson WJ, Yang Z, Wolfner MF, Aquadro CF. 2001. Positive Darwinian selection drives the evolution of several female reproductive proteins in mammals. Proc Natl Acad Sci USA 98:2509–14.

Swofford DL. 2002. PAUP*: phylogenetic analysis using parsimony (*and other methods). Version 4. Sunderland, MA: Sinauer Associates.

Tiemann-Boege I, Kilpatrick CW, Schmidly DJ, Bradley RD. 2000. Molecular phylogenetics of the *Peromyscus boylii* species group (Rodentia: Muridae) based on mitochondrial *cytochrome b* sequences. Mol Phylogenet Evol 16:366–78.

Torgerson DG, Kulathinal RJ, Singh RS. 2002. Mammalian sperm proteins are rapidly evolving: evidence of positive selection in functionally diverse genes. Mol Biol Evol 19:1973–80.

Tsaur S-C, Wu CI. 1997. Positive selection and the molecular evolution of a gene of male reproduction, *Acp26Aa* of *Drosophila*. Mol Biol Evol 14:544–9.

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: 1083–101.

Wassarman PM, Jovine L, Litscher ES. 2001. A profile of fertilization in mammals. Nat Cell Biol 3:E59–64.

Wassarman PM, Litscher ES. 2001. Towards the molecular basis of sperm and egg interaction during mammalian fertilization. Cells Tissues Organs 168:36–45.

Wolff JO. 1989. Social behavior. In: Kirkland GL, Layne JN, editors. Advances in the study of *Peromyscus* (Rodentia). Lubbock, Tx: Texas Tech University Press.

Wong WSW, Yang Z, 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 168:1041–51. p 271–91.

Wyckoff GJ, Wang W, Wu CI. 2000. Rapid evolution of male reproductive genes in the descent of man. Nature 403:304–9.

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

Yang Z, Nielsen R. 2002. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. Mol Biol Evol 19:908–17.

Yang Z, Nielsen R, Goldman N, Pedersen AK. 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics 155:431–49.

Yang Z, Wong WSW, Nielsen R. 2005. Bayes empirical Bayes inference of amino acid sites under positive selection. Mol Biol Evol 22:1107–18.

Zhang J. 2003. Parallel functional changes in the digestive RNases of ruminants and colobines by divergent amino acid substitutions. Mol Biol Evol 20:1310–7.

Spencer Muse, Associate Editor

Accepted May 19, 2006