J Mol Evol (2003) 56:294–307
DOI: 10.1007/s00239-002-2401-6

JOURNAL OF MOLECULAR EVOLUTION

Tests for Positive Selection on Immune and Reproductive Genes in Closely Related Species of the Murine Genus *Mus*

Sharon A. Jansa,[1] Barbara L. Lundrigan,[2] Priscilla K. Tucker[3]

[1] James Ford Bell Museum of Natural History and Department of Ecology, Evolution, and Behavior, University of Minnesota, St. Paul, MN 55108-6098, USA
[2] Michigan State University Museum and Department of Zoology, Michigan State University, East Lansing, MI 48824-1045, USA
[3] Museum of Zoology and Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI 48109-1079, USA

Received: 7 February 2002 / Accepted: 8 October 2002

Abstract. We examine variation among species of *Mus* in four genes involved in reproduction and the immune response for evidence of positive selection: the sperm recognition gene *Zp-3*, the testis-determining locus *Sry,* the testicular cell surface matrix protein *Tcp-1*, and the immune system protein $\beta_2 m$. We use likelihood ratio tests in the context of a well-supported phylogeny to determine whether models that allow for positively selected sites fit the sequences better than models that assume purifying selection. We then apply a Bayesian approach to identify particular sites in each gene that have a high posterior probability of being under positive selection. We find no evidence of positive selection on the *Tcp-1* gene, but for *Zp-3*, *Sry*, and $\beta_2 m$, models that allow for positively selected sites fit the sequences better than alternatives. For each of these genes, we identify sites that have a high (>95%) posterior probability of being positively selected. For *Zp-3*, two of these sites occur near the sperm-binding region, while one occurs in a region whose functional role remains unstudied but where the pattern of change predicts functional importance. A single site in *Sry* shows an elevated rate of replacement substitution but occurs in a region of apparently little functional importance; therefore, relaxation of functional constraints may better explain the rapid evolution of this site. Three sites in $\beta_2 m$ have a posterior probability >50% of being under positive selection. While the functional role for two of these sites is unknown, the third is known to influence the ability of MHC class I molecules to present antigens to the immune system; therefore, the elevated rate of replacement substitutions at this site is consistent with selection acting to promote variability in immune system proteins.

Key words: *Sry* — *Tcp-1* — *Zp-3* — Zona pellucida — β-2-Microglobulin — *Mus* — Positive selection

Introduction

The neutral theory of molecular evolution maintains that the majority of changes at the molecular level are fixed by random drift of selectively equivalent mutations (Kimura 1983). Accordingly, while advantageous mutations may contribute directly to improving an organism's fitness, the rate of positive selection is probably too low for adaptive change to be the driving force of evolution at most genetic loci. For this reason, genes that appear to be evolving under positive selection are of great interest to evolutionary biologists. Such genes seem to fall into discrete functional categories (Endo et al. 1996) and the study of their function and evolution could reveal

Correspondence to: Sharon A. Jansa, *email:* jansa003@tc.umn. edu

much about the nature of adaptive evolution at the molecular level. For example, genes associated with the immune response (Hughes and Nei 1988, 1989), gamete recognition (Lee et al. 1995; Swanson and Vacquier 1995; Metz and Palumbi 1996; Metz et al. 1998), and male reproduction (Tsaur and Wu 1997; Ting et al. 1998; Wycoff et al. 2000) are hypothesized to have experienced long-term positive selection. Detailed study of selection on particular sites in these genes can identify regions of potential functional importance that may play a pivotal role in speciation or the immune response.

Selection at the molecular level is typically detected by comparing the number of nonsynonymous (d_N) and synonymous (d_S) substitutions per site for a given gene. Generally, purifying selection is inferred when d_S is greater than d_N, while positive selection is inferred when the ratio of nonsynonymous-to-synonymous substitutions (d_N/d_S) exceeds 1. Several methods have been developed to estimate d_N, d_S, and their ratio. These methods differ principally in whether they rely on pairwise comparisons of aligned coding sequences among taxa (Nei and Gojobori 1986; Ina 1995; Li et al. 1985; Li 1993) or use a phylogenetic tree to reconstruct the pattern of molecular change among taxa (Goldman and Yang 1994; Nielsen and Yang 1998; Yang 1998; Suzuki and Gojobori 1999).

Pairwise estimates of the d_N/d_S ratio provide an extremely stringent test of positive selection, because, for most genes, only a few amino acids are under positive selection. In such cases, pairwise estimates of the d_N/d_S ratio across the entire gene may be less than 1 despite positive selection operating on particular sites. Moreover, pairwise comparisons do not allow changes to be localized to particular lineages in evolutionary history. For this reason, the importance of having a phylogeny as a basis for comparative studies of gene evolution is increasingly being recognized. Using a phylogenetic framework, one can localize d_N/d_S ratios to particular branches and evaluate whether and when positive selection has played a role in gene evolution (e.g., Messier and Stewart 1997; Willet 2000; Yang et al. 2000b). Recently, several phylogeny-based likelihood tests have been developed to test for positive selection in different lineages (Yang 1998) or among different codon positions in a gene (Nielsen and Yang 1998; Suzuki and Gojobori 1999; Yang et al. 2000a). These methods provide a more sensitive and explicitly statistical test of positive selection than pairwise distance-based estimates (Nielsen and Yang 1998; Bielawski et al. 2000).

In this study, we use phylogeny-based likelihood tests to examine four genes for evidence of positive selection among closely related species in the genus *Mus*. These genes might be expected to have experienced positive selection during their evolution by virtue of being involved in either gamete recognition, male reproduction, or the immune response. While the selective forces driving positive evolution differ among reproductive genes and between reproductive and immune genes, identifying the patterns of adaptive change in each of these genes is the first step toward inferring the selective processes driving their evolution. Zona pellucida-3 (*Zp-3*) is the mammalian sperm recognition locus and is responsible for binding sperm to the egg surface and initiating the acrosome reaction. The Y chromosome-linked gene, *Sry*, codes for testis determination and therefore determines maleness in mammals. The t-complex polypeptide-1 gene (*Tcp-1*) codes for a protein whose specific function is unknown but which is expressed in mouse testes at high levels during spermatogenesis. The immune system gene β-2-microglobulin (β_2m) associates with major histocompatability complex (MHC) class I heavy chains and ensures the proper display of class I antigens. While the pattern of nucleotide substitution has been studied for some of these genes in some of these taxa (*Sry*, β_2m, and *Tcp-1* for selected species of *Mus*; see citations below), one gene (*Zp-3*) has never been examined for any of these taxa, and none of these genes has been examined using methods designed to identify particular sites subject to positive selection.

Methods

Data for this study are taken from Lundrigan et al. (2002). The data consist of partial sequences from the coding region of the Y chromosome-linked sex-determining locus *Sry*; partial sequences from exon 2 of β-2-microglobulin (β_2m); exons 8, 9, and 10 of t-complex polypeptide-1 (*Tcp-1*); and exons 1, 3, 4, 6, and 7 of the sperm recognition gene zona pellucida-3 (*Zp-3*). For the purposes of this study, only coding regions of these genes were used. The genes were sequenced for 13 species and subspecies of *Mus* and for representatives from three additional murine genera (*Mastomys*, *Hylomyscus*, and *Rattus*), which were used as outgroups in phylogenetic analysis. Sequences from *Sry* for *Mus saxicola* were unobtainable, as were sequences from β_2m for *Mastomys* and *Hylomyscus* and *Tcp-1* for *Hylomyscus*. Additional information about all genes, including GenBank accession numbers, is given by Lundrigan et al. (2002). A single best tree results from a combined-data parsimony analysis of these four genes plus sequences from the mitochondrial genes12S rRNA and cytochrome *b*. An identical topology is recovered when these data are analyzed under the best-fit maximum likelihood model (GTR + I + Γ with clock enforced). We use this tree topology (Fig. 1) as the basis for all likelihood-based tests of selection performed in this study. Details of all methods used to obtain this tree are given by Lundrigan et al. (2002).

We used three likelihood ratio tests (LRTs) to examine our data for evidence of positive selection. First, because positive selection may act at discrete points during the evolution of a lineage rather than constantly across an entire phylogeny, we examined whether the d_N/d_S ratio (hereafter referred to as ω) varies across lineages for each gene. For this test, a model that assumes a constant ω across all lineages (one-ratio model, or M0 [Yang et al. 2000a]) is compared with one that allows ω to vary across lineages (free-ratio model [Yang, 1998]). The one-ratio model is a simplified version of

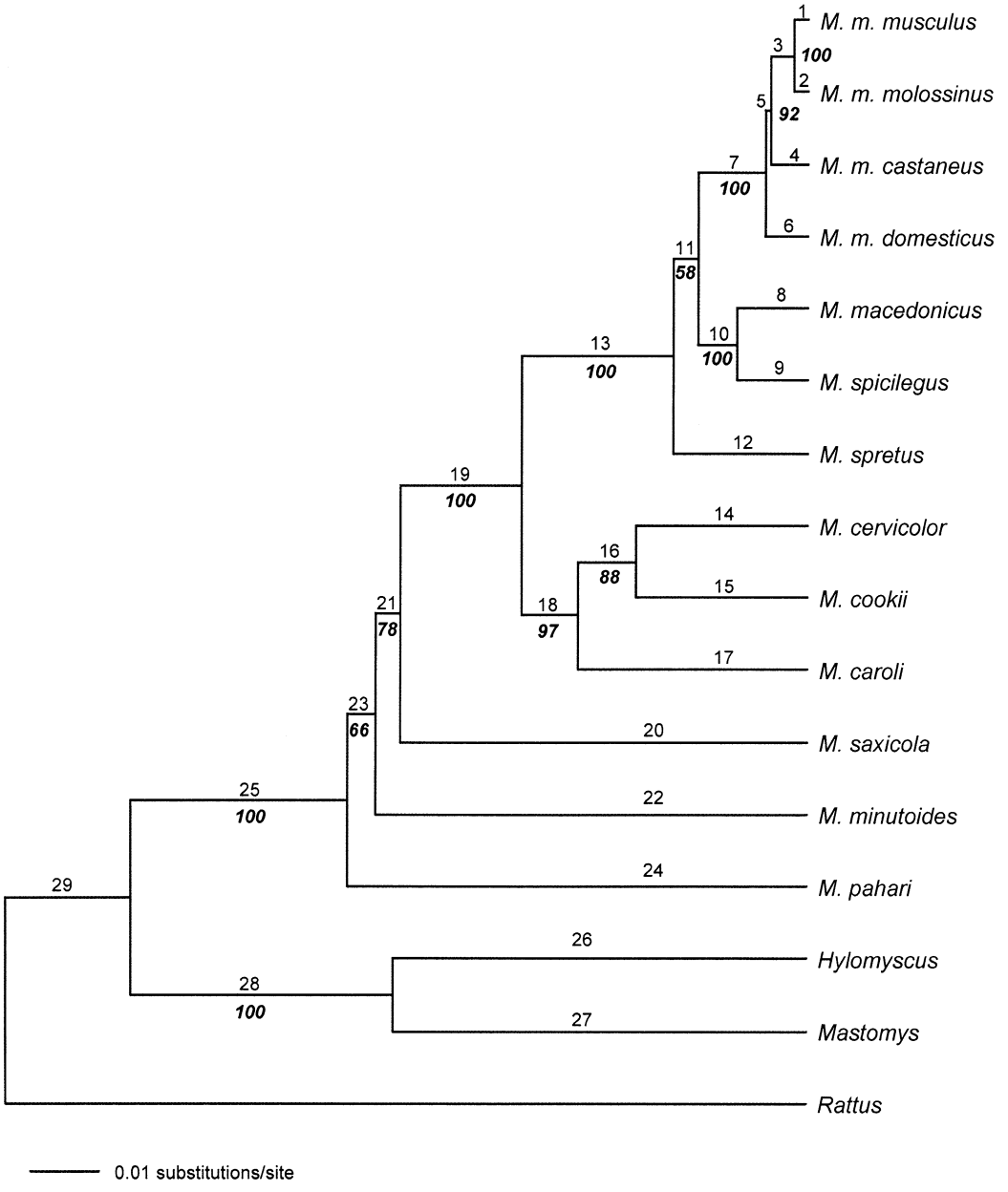

Fig. 1. The tree describing relationships among species of *Mus* that was used as the basis for all phylogeny-based tests of positive selection described in this study (see Lundrigan et al. 2002). This tree results from a maximum-likelihood analysis of DNA sequence from six genes (12S rRNA, cytochrome *b, Sry, Tcp-1, Zp-3,* and $\beta_2 m$) under the best-fit GTR + I + Γ + clock model of sequence evolution. Numbers in *boldface italics* are bootstrap support values; numbers *above the branch* identify lineages referred to in the text.

the Goldman and Yang (1994) model and accounts for transition/transversion rate bias, different base frequencies at codon positions, and the structure of the genetic code (Yang 1998). The free-ratio model assumes the same parameters but allows ω to vary freely across lineages. The log-likelihood value under the one-ratio model can be compared with that under the free-ratio model; the appropriate test statistic compares twice the difference in log-likelihood values ($-2 \ln \Lambda = 2[\ln \lambda_1 - \ln \lambda_2]$, where λ_1 is the likelihood under the more parameter-rich model) with a χ^2 distribution with degrees of freedom equal to the number of branches minus 1. If the free-ratio model provides a significantly better fit, then a constant ω rate (one-ratio model) can be rejected. In addition, we use the free-ratio model to estimate the ω value for each branch in the phylogeny. While this model is parameter-rich and unlikely to produce accurate ω estimates for all branches (Yang et al. 2000b), it is nonetheless useful for identifying lineages where episodes of positive selection might have occurred.

Second, we used the approach outlined by Nielsen and Yang (1998) to test whether positive selection has acted on particular sites in each gene. Briefly, this method tests for the presence of positively selected sites by comparing the log-likelihood under a model of purifying selection (the "neutral" model) with that under a model that allows for positively selected sites. The neutral model (M1) assumes two categories of codon sites in a gene: those for which nonsynonymous changes are neutral, with $\omega = 1$, and those for which nonsynonymous changes are deleterious so that $\omega = 0$. The positive selection model (M2) adds a third category of codon site with ω as a free parameter, thus allowing for positively selected sites (those with $\omega > 1$). The log-likelihood value under the neutral model can be compared with that from the

positive selection model to determine whether a model that allows for positively selected sites fits the data better than one that allows for only neutral and negatively selected sites. The appropriate test statistic compares twice the difference in log-likelihood values ($-2\ln\Lambda$) with a χ^2 distribution with two degrees of freedom. It is appropriate to conclude that positive selection has affected the gene sequences only if the estimate of ω under the positive selection model is greater than 1.

Finally, an LRT comparing results from the positive selection model (M2) and the neutral model (M1) may not be sensitive enough to detect selection at particular sites if the majority of sites in a gene has been neutral or negatively selected (Yang et al. 2000a). We therefore implemented a third LRT that uses models that permit heterogeneous ω ratios among sites according to a β distribution. The null model in this test (M7) does not allow for positively selected sites but assumes that sites are modeled under a β distribution $\beta(p,q)$ with ω limited to the interval [0,1], where 0 corresponds to complete selective constraint (deleterious mutations) and 1 to no selective constraint (neutral mutations). The log-likelihood under this model is compared to that under a model (M8) that allows for a proportion of sites to be drawn from the β distribution and the remaining sites with $\omega > 1$ to be estimated from the data. The M8 model has two more parameters than M7; therefore, the LRT compares twice the difference in log-likelihood values ($-2\ln\Lambda$) with a χ^2 distribution with two degrees of freedom. Comparison of the log-likelihood scores under the M8 model with those from the M7 model constitutes a test for positive selection similar to the comparison of the M2 and M1 models, but where a more complex model describes the distribution of ω. We note that models used to test for selection across sites assume a constant ω ratio across all lineages (Nielsen and Yang 1998; Yang et al. 2000a). While we test for the validity of this assumption with the free-ratio/one-ratio LRT described above, it is not known how sensitive these tests are to violations of this assumption. We therefore report the results of all tests for positive selection, but note that parameter estimates may not be accurate when assumptions of rate homogeneity are violated.

Once the fit of various nested models was assessed using LRTs, we identified particular sites in each gene that were likely to have evolved under positive selection. This was accomplished using an empirical Bayesian approach as outlined by Nielsen and Yang (1998). For this approach, unknown parameters in Bayes' equation (e.g., branch length and the ω distribution across sites) are first estimated from the data using the likelihood function as applied in M2 and M8. Once these parameters have been estimated, Bayes' theorem is used to estimate the posterior probability that a given site came from the class of positively selected sites (Nielsen and Yang 1998; Yang and Bielawski 2000).

All analyses were accomplished using the codeml program in PAML 2.0k (Yang 1997, 2000) with models specified according to descriptions given by Yang et al. (2000a). For each model, equilibrium codon frequencies were estimated from the average nucleotide frequencies at each codon position (CodonFreq = 2), amino acid distances were assumed to be equal (aaDist = 0), and the transition/transversion ratio (κ) was estimated from the data. In addition, for comparative purposes, we calculated pairwise comparisons of d_N/d_s using the model of Yang and Nielson (2000) as implemented in YN00 of PAML 2.0 (Yang 2000).

Results

Zp-3

Only two of the pairwise d_N/d_S values for *Zp-3* exceed 1 (Fig. 2), and both of these comparisons involve *Mus macedonicus* (*M. macedonicus* vs *M. m. domesticus*, $d_N/d_S = 1.54$; vs *M. m. musculus*, $d_N/d_S = 1.16$). Not surprisingly, then, the free-ratio model assigns estimates of $\omega > 1$ for the branch leading to *M. macedonicus* (branch 8; Fig. 1, Table 1). The free-ratio model also assigns $\omega > 1$ to the branch leading to *M. cervicolor* (branch 14; Fig. 1, Table 1), a pattern that was not detected among the pairwise comparisons. While these two lineages show ω values greater than 1, this inference is based on only a few changes per branch and cannot be considered statistically meaningful, because the one-ratio model cannot be rejected for the tree as a whole ($-2\ln\Lambda = 37.799$; $p = 0.127$, df = 29).

Models that allow for heterogeneous ω ratios across sites (M1, M2, M7, and M8) require that ω remains constant across evolutionary time. The observation that a constant ω ratio cannot be rejected for *Zp-3* across this tree suggests that these models can be reliably applied to the *Zp-3* data to test for the presence of positively selected sites. An LRT comparing M2 (selection) to M1 (neutral) rejects the neutral model in favor of a model that allows for positively selected sites with marginal significance ($-2\ln\Lambda = 6.154$; $p = 0.046$, df = 2), but an empirical Bayesian approach with parameters estimated under M2 fails to identify any sites under positive selection. An LRT comparing the more parameter-rich models M8 (β, selection) and M7 (β, neutral) also rejects the neutral model in favor of the selection model at a similar level of significance ($-2\ln\Lambda = 6.585$; $p = 0.037$, df = 2). In contrast to M2, the Bayesian approach applied under M8 identifies several sites under positive selection with a posterior probability >50%; three of these sites have a posterior probability >95% of having $\omega > 1$ (Table 2).

These results are not surprising given the limitations of the M2 model to infer positive selection (Yang et al. 2000a). If a gene has a high proportion of slightly deleterious mutations ($0 < \omega < 1$), the free class in M2 is forced to account for these; any positively selected mutations ($\omega > 1$) are then incorporated into the class of conserved sites ($\omega = 1$) (Yang et al. 2000a). The failure of M2 to detect positively selected sites for *Zp-3* is probably due to this model's inability to account simultaneously for ω values between 0 and 1 and for those greater than 1. For *Zp-3*, model M2 identifies a large proportion of sites (82.7%) in the slightly deleterious class ($\omega = 0.065$; Table 2). However M8, which explicitly accounts for these mutations, estimates 94.7% of sites as neutral ($\omega = 1$), deleterious ($\omega = 0$), or slightly deleterious ($0 < \omega < 1$); the free parameter is therefore able to capture the small proportion of sites (5.4%) under positive selection ($\omega = 2.4$; Table 2). When these parameters are subsequently used in the Bayesian approach, several sites with a high

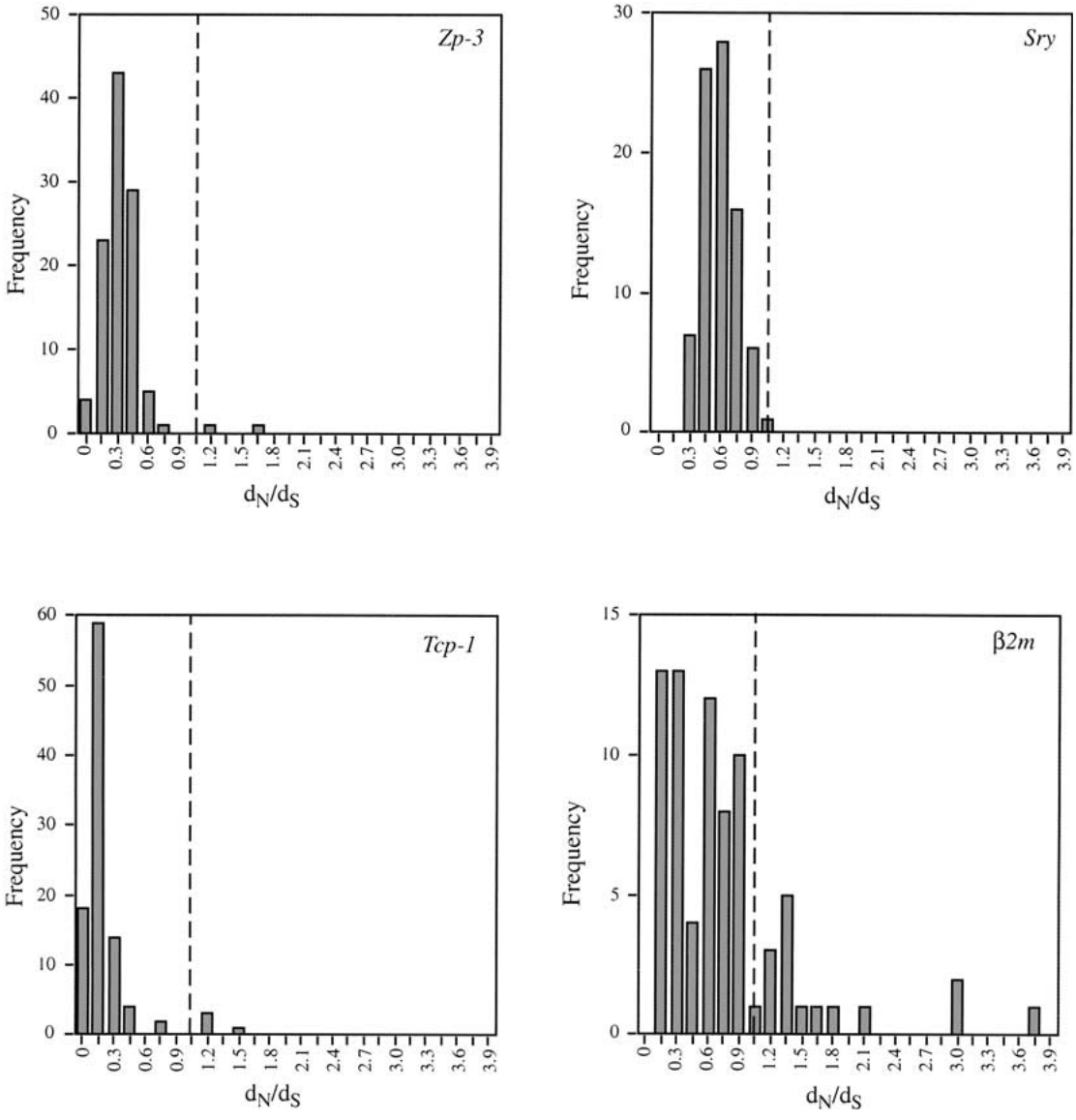

Fig. 2. Histograms of pairwise d_N/d_S values for each of the four genes examined in this study. Pairwise comparisons among all species of *Mus* and the outgroup taxa *Mastomys, Hylomyscus,* and *Rattus* are included. d_N and d_S values were calculated using the method of Yang and Nielsen (2000) as implemented in the YN00 algorithm of PAML. The *dashed vertical line* marks the threshold value ($d_N/d_S = 1$) above which positive selection is inferred.

posterior probability of being positively selected can be identified.

Sry

No pairwise comparison of the d_N/d_S ratio for *Sry* exceeds 1 for these mice (Fig. 2), and comparison of the likelihood scores under the free-ratio and one-ratio models indicates that a constant ω ratio can apply across the tree ($-2\ln\Lambda = 18.81$; $p = 0.877$, df = 27). Evidence for positive selection on *Sry* seems slight when these genewide tests are applied. However, models that allow for positively selected sites within the gene (M2 and M8) fit the *Sry* data much better than models that assume neutrality (M1 and M7) (M2 vs M1, $-2\ln\Lambda = 13.255$, $p < 0.001$, df = 2; M8 vs M7, $-2\ln\Lambda = 12.992$, $p < 0.001$, df = 2). In contrast to *Zp-3*, the *Sry* gene does not appear to have a high proportion of slightly deleterious mutations, therefore parameter estimates under the selection models (M2 and M8) are similar (e.g., 0.8% of sites have ω = 14.6 under M2 and 0.8% of sites have ω = 14.0 under M8; Table 2), and the Bayesian approach as applied under each of these models infers the same site (92) as positively selected with a similarly high posterior probability (>99%).

Tcp-1

For *Tcp-1*, pairwise comparisons of d_N/d_S detected evidence for positive selection in only one lineage (*M. spretus*). Four pairwise estimates of the d_N/d_S

Table 1. Estimated numbers of replacement and synonymous changes and estimated ω values for each of four genes along each branch of the *Mus* phylogeny (Fig. 1)[a]

	Zp-3			*Sry*			*Tcp-1*			β_2m		
Branch[b]	d_N	d_S	ω	d_N	d_S	ω	d_N	d_S	ω	d_N	d_S	ω
1	0.000	0.006	0.001	0.000	0.000	1.011	0.000	0.010	0.001	0.000	0.000	0.001
2	0.002	0.012	0.187	0.005	0.000	∞	0.000	0.000	1.006	0.006	0.019	0.316
3	0.000	0.000	0.996	0.000	0.000	0.656	0.000	0.000	1.006	0.000	0.000	0.072
4	0.000	0.000	1.224	0.000	0.000	1.011	0.000	0.000	1.006	0.998	0.000	∞
5	0.002	0.006	0.375	0.000	0.000	1.011	0.000	0.000	1.006	0.000	0.000	0.036
6	0.000	0.000	0.835	0.000	0.000	1.011	0.000	0.000	1.006	0.000	0.000	0.059
7	0.000	0.000	0.986	0.009	0.000	∞	0.011	0.021	0.525	0.018	0.000	∞
8	0.002	0.000	∞	0.000	0.011	0.001	0.000	0.011	0.001	0.006	0.000	∞
9	0.000	0.012	0.001	0.000	0.000	0.732	0.000	0.000	0.999	0.006	0.018	0.333
10	0.007	0.006	1.122	0.000	0.000	63.375	0.004	0.011	0.344	0.000	0.000	6.221
11	0.000	0.000	0.843	0.000	0.000	1.007	0.000	0.000	0.998	0.000	0.000	1.321
12	0.004	0.006	0.748	0.000	0.023	0.001	0.011	0.000	∞	0.030	0.000	∞
13	0.000	0.018	0.001	0.019	0.023	0.801	0.000	0.021	0.001	0.023	0.056	0.406
14	0.002	0.000	∞	0.009	0.011	0.808	0.004	0.010	0.348	0.006	0.037	0.159
15	0.007	0.018	0.371	0.000	0.000	2.037	0.000	0.000	0.999	0.000	0.000	25.796
16	0.000	0.006	0.001	0.005	0.000	∞	0.000	0.010	0.001	0.022	0.000	∞
17	0.000	0.024	0.001	0.005	0.011	0.405	0.000	0.010	0.001	0.007	0.019	0.381
18	0.000	0.000	0.981	0.000	0.011	0.001	0.000	0.000	0.991	0.030	0.000	∞
19	0.002	0.018	0.120	0.000	0.014	0.001	0.000	0.021	0.001	0.032	0.152	0.210
20	0.011	0.018	0.628	NA	NA	NA	0.004	0.021	0.174	0.037	0.034	1.089
21	0.000	0.000	0.001	NA	NA	NA	0.004	0.000	∞	0.007	0.009	0.803
22	0.027	0.037	0.739	0.331	0.056	0.590	0.004	0.077	0.047	0.018	0.079	0.231
23	0.000	0.000	0.965	0.000	0.000	51.768	0.000	0.021	0.001	0.020	0.048	0.413
24	0.004	0.082	0.054	0.185	0.021	0.881	0.000	0.067	0.001	0.032	0.174	0.182
25	0.009	0.041	0.224	0.010	0.015	0.662	0.000	0.049	0.001	NA	NA	NA
26	0.054	0.100	0.536	0.013	0.071	0.188	NA	NA	NA	NA	NA	NA
27	0.016	0.010	0.163	0.010	0.001	∞	NA	NA	NA	NA	NA	NA
28	0.009	0.073	0.122	0.010	0.009	0.896	0.247	0.007	0.0293	NA	NA	NA
29	0.012	0.053	0.215	0.040	0.121	0.329	0.004	0.189	0.0195	0.057	0.694	0.082

[a] Estimates of d_N, d_S, and ω were calculated under the free-radio model of Yang (1998).
[b] Branch numbers from Fig. 1.

ratio exceed 1 (Fig. 2); each of these involves comparisons with *M. spretus*. Moreover, branch length estimates from the free-ratio model show an ω ratio approaching infinity for the branch leading to *M. spretus* (branch 12; Fig. 1, Table 1), but a constant rate ratio cannot be rejected for the tree as a whole ($-2\ln\Lambda = 32.055$; $p = 0.230$, df = 27); therefore, this deviation cannot be considered statistically meaningful. When models that test for selection across sites are applied, neither of the models that allow for positively selected sites (M2 or M8) fits these sequences better than the corresponding model that assumes neutrality (M1 or M7) (M2 vs M1, $-2\ln\Lambda = 3.844$, $p = 0.146$, df = 2; M8 vs M7, $-2\ln\Lambda = 0.171$, $p = 0.918$, df = 2). Therefore, there is no statistically compelling evidence for positive selection having affected any site or any lineage during the evolution of *Tcp-1* in these mice.

β_2m

Of the genes we examined, β_2m has the highest variance in pairwise d_N/d_S estimates and exhibits a number of comparisons in which the d_N/d_S ratio greatly exceeds 1 (Fig. 2). Not surprisingly, then, a model that allows heterogeneity in the d_N/d_S ratio fits the β_2m data better than one that assumes a constant ratio across all lineages ($-2\ln\Lambda = 38.383$; $p = 0.042$, df = 25). Each of the models that allow ω to vary across sites (M1, M2, M7, and M8) assumes a constant ω ratio across all lineages (Nielsen and Yang 1998; Yang et al. 2000a). Results from an LRT between the one-ratio model and the free-ratio model suggest that the β_2m data violate this assumption. It is not known how robust these models are to violations of a constant substitution ratio across lineages; therefore, we report the results of all pertinent LRTs among these models for the β_2m gene, but note that parameter estimates may not be accurate. An LRT between M2 (selection) and M1 (neutral) does not allow a model of neutrality to be rejected in favor of one that allows for positively selected sites ($-2\ln\Lambda = 3.273$; $p = 0.195$, df = 2). However, an LRT between the more complex models M8 and M7 rejects a neutral model (M7) in favor of a positive selection model (M8) ($-2\ln\Lambda = 6.389$; $p = 0.041$,

Table 2. Likelihood values, parameter estimates, and sites under postive selection as inferred under six models as applied to each of four loci

Locus	Model	lnλ	Parameter estimate	Positively selected sites[a]
Zp-3	M0 (one ratio)	−1802.47	$\omega = 0.267$	
	Free-ratio model	−1783.57		
	M1 (neutral)	−1859.21	$p_0 = 0.709$ $p_1 = 0.290$	
	M2 (selection)	−1856.14	$p_0 = 0.000$ $p_1 = 0.173$ $p_2 = 0.827$ $\omega_2 = 0.065$	None
	M7 (β, neutral)	−1857.26	$p = 0.111$ $q = 0.366$	
	M8 (β, selection)	−1853.97	$p_0 = 0.947$ $p_1 = 0.054$ $p = 0.464$ $q = 2.591$ $\omega = 2.420$	25, 27, 36, 185, 206, <u>223</u>, 335, <u>337</u>, **342**
Sry	M0 (one ratio)	−823.60	$\omega = 0.478$	
	Free-ratio model	−814.20		
	M1 (neutral)	−1245.67	$p_0 = 0.477$ $p_1 = 0.523$	
	M2 (selection)	−1239.04	$p_0 = 0.452$ $p_1 = 0.540$ $p_2 = 0.008$ $\omega_2 = 14.576$	**92**
	M7 (β, neutral)	−1245.73	$p = 0.016$ $q = 0.015$	
	M8 (β, selection)	−1239.23	$p_0 = 0.992$ $p_1 = 0.008$ $p = 0.005$ $q = 0.005$ $\omega = 13.978$	**92**
Tcp-1	M0 (one ratio)	−870.22	$\omega = 0.090$	
	Free-ratio model	−854.19		
	M1 (neutral)	−914.35	$p_0 = 0.833$ $p_1 = 0.167$	
	M2 (selection)	−912.43	$p_0 = 0.000$ $p_1 = 0.056$ $p_2 = 0.944$ $\omega_2 = 0.061$	None
	M7 (β, neutral)	−912.51	$p = 0.192$ $q = 0.146$	
	M8 (β, selection)	−912.43	$p_0 = 0.957$ $p_1 = 0.043$ $p = 70.065$ $q = 998.891$ $\omega = 1.224$	22, 131
β_2m	M0 (one ratio)	−857.60	$\omega = 0.316$	
	Free-ratio model	−838.41		
	M1 (neutral)	−1070.64	$p_0 = 0.509$ $p_1 = 0.491$	
	M2 (selection)	−1069.00	$p_0 = 0.395$ $p_1 = 0.315$ $p_2 = 0.290$ $\omega_2 = 0.256$	None
	M7 (β, neutral)	−1069.37	$p = 0.135$ $q = 0.203$	
	M8 (β, selection)	−1066.17	$p_0 = 0.971$ $p_1 = 0.021$ $p = 0.177$ $q = 0.296$ $\omega = 6.354$	77, 88, **94**

[a] Sites with a posterior probability > 50% of having $\omega > 1$. Boldface indicates a posterior probability > 99%; an underline indicates a posterior probability > 95%.

df = 2). Moreover, a Bayesian approach applied under M8 identifies three sites under positive selection with a posterior probability >50%; one of these has a posterior probability >99% (Table 2).

Discussion

Statistical tests that examine the pattern of nucleotide substitution are only the first step in inferring whether particular sites in a gene have been subject to positive selection. The tests we have used assume that positive selection has occurred when a particular site shows a d_N/d_S ratio greater than 1; however, an elevated d_N/d_S ratio is also consistent with an explanation of relaxed functional constraints. It is therefore critical to couple the results of statistical tests such as these with an examination of molecular function. Sites that are functionally significant and show evidence of an elevated d_N/d_S ratio have most likely been driven by positive selection.

In practice, many studies of molecular evolution proceed without sufficient information about gene function to make inferences about the relative functional importance of different nucleotide sites. In these cases, statistical tests of positive selection can be predictive: particular sites showing an elevated d_N/d_S ratio can be targeted for future studies of molecular function and adaptive significance. If sites with an elevated d_N/d_S ratio exhibit no demonstrable function, it may be more appropriate to conclude that replacement substitutions have been allowed to accumulate due to processes other than positive selection. In the following discussion, we examine our results in light of what is known about the function of each gene to determine whether sites have functional importance and can be reasonably inferred to have evolved under positive selection.

Positive Selection on the Gamete Recognition Locus Zp-3

Genes involved in gamete recognition have some of the highest rates of replacement substitutions ever reported (Lee et al. 1995; Swanson and Vacquier 1995; Metz and Palumbi 1996; Vacquier et al. 1997; Yang et al. 2000b). It has been suggested that these genes accumulate amino acid changes between species in response to selection preventing cross-species fertilization, but other evolutionary processes have been implicated (Vacquier et al. 1997; Yang et al. 2000b). The majority of studies of gamete recognition genes has been on free-spawning marine organisms where gamete recognition is a critical step in maintaining interspecific reproductive barriers. However, experimental evidence suggests that gamete recognition in mammals is also largely species-specific. Studies of different species of rodents (Maddock and Dawson 1974; Hanada and Chang 1978; Fukuda et al. 1979; Roldan et al. 1985; Roldan and Yanagimachi 1989), rabbits (Chang and Hancock 1967), and mustelids (Chang and Hancock 1967) show that cross-species fertilization is rare in vitro and that gametes strongly prefer to bind conspecifically. Specific studies of *Mus* gametes in vitro show that several of the species included in our study cannot successfully cross-fertilize (West et al. 1977; Lambert 1984) and that this failure is due to changes in gamete recognition molecules among different species (Lambert 1984).

The *Zp-3* gene codes for the primary sperm receptor in mammals and appears to be responsible for controlling species-specific gamete interactions (Bleil and Wassarman 1980a, b; Ringuette et al. 1988; Wassarman 1990). Sperm bind to the *Zp-3* molecule via oligosaccharide side chains that are attached to the protein backbone at serine and threonine residues (Florman and Wassarman 1985; Bleil and Wassarman 1988). It has been suggested that differences among species in the structure and presentation of these oligosaccharide side chains may be responsible for determining species-specific gamete recognition (Skutelsky et al. 1994; Wassarman and Litscher 1995). In particular, a region containing five serine residues (Ser 329, Ser 331–334) that are essential for *Zp-3* to bind sperm successfully has been identified in *Mus musculus* (Rossiere and Wassarman 1992; Kinloch et al. 1995; Wassarman and Litscher 1995; Chen et al. 1998). None of the sites we identify as having an elevated ω ratio affect these serine residues, and this region is conserved across all murine species sequenced (Fig. 3). However, we find two sites (337 and 342) immediately outside this region that have a posterior probability >95% of having ω>1 and an additional site in this region (335) that has a posterior probability >75% of having ω>1 (Fig. 3, Table 2).

Swanson et al. (2001) recently examined the *Zp-3* protein among distantly related mammalian species and identified several sites that show a high posterior probability of being under positive selection. Many of the changes affected the serine residues in the sperm binding region (sites 331 and 333) or adjacent to it (sites 340, 341, 345, 347, 348). Intriguingly, while we also find sites adjacent to the sperm binding region with ω>1, none of the sites we identify among these species of *Mus* overlap with those identified by Swanson et al. (2001). It is possible that changes in residues adjacent to the sperm-binding region affect the glycosylation pattern for the serine residues in that region, which may in turn affect interspecific sperm recognition (as Swanson et al. [2001] suggest). Moreover, we identify an additional site in *Zp-3* (site 223) with a posterior probability >95% of having evolved under positive selection. This site is of particular interest as it is adjacent to a potential N-linked

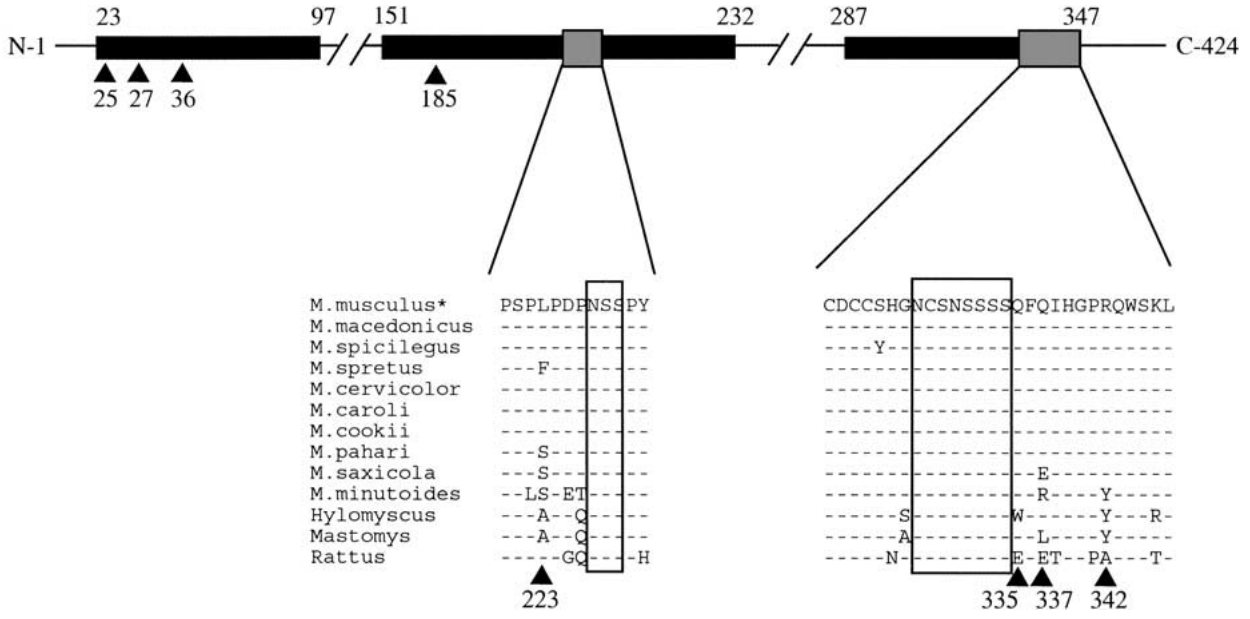

Fig. 3. Schematic representation of the *Zp-3* protein. *Black bars* represent regions of the gene sequenced for this study; *broken lines* show regions for which the sequence is not available. *Gray bars* indicate regions that are shown in greater detail below the schematic. *Triangles* identify specific sites that have a posterior probability >50% of being under positive selection (Table 2); amino acid site numbers are given beneath the triangles. The *box* surrounding residues 327–334 includes the sperm recognition site and the five serine residues essential for successful sperm binding; the *box* surrounding residues 227–229 identifies a potential N-glycosylation site. *M. musculus** includes all the subspecies we sequenced; all subspecies have identical *Zp-3* sequences.

oligosaccharide-binding site (Asn–Ser–Ser) that is conserved across all murine species included in this study (Fig. 3) and contains a Ser residue in three species that may function as an O-linked oligosaccharide-binding site and affect sperm's ability to bind *Zp-3*.

While studies of the evolution of the *Zp-3* protein may inform studies of its function, it should be noted that the molecular biology of sperm–egg interaction in mammals is still poorly understood. Recently, Rankin et al. (1998) showed that transgenic mice carrying a human *Zp-3* molecule bind mouse sperm but fail to bind human sperm. This result suggests that factors other than *Zp-3* protein sequence are important for sperm recognition. Ultimately, the functional importance of the specific sites Swanson et al. (2001) and our study have identified can be tested using directed mutagenesis and other experimental methods.

Rapid Evolution of the Male Reproductive Gene Sry *but Not* Tcp-1

Genes involved in male reproduction have been hypothesized to evolve under positive selection, though in many cases the forces driving their evolution remain unclear (Ferris et al. 1997; Tsaur and Wu 1997; Tsaur et al. 1998; Civetta and Singh 1998; Wycoff et al. 2000; Clark and Civetta 2000). There appear to be two primary modes of selection that could promote diversification of male reproductive proteins. The first involves the creation and maintenance of postzygotic isolating mechanisms between closely related species. According to this theory, sterility of male hybrids might result from incompatibility among loci involved in gamete production (Ting et al. 1998; Nei and Zhang 1998). Rapid diversification of male reproductive proteins between potentially interbreeding species could then be selectively advantageous. The second theory relies on male–male competition as the driving force for positive selection (Eberhard, 1985). This theory posits that variation in the structure of sperm proteins may affect sperm morphology and ultimately determine competitiveness in fertilizing the egg (Wycoff et al. 2000; Clark and Civetta 2000). Competition among sperm (either inter- or intraspecific) could then drive rapid diversification of reproductive proteins. Teasing these two modes of selection apart may be all but impossible; nonetheless, either may cause genes involved in male reproduction to exhibit the signature of positive directional selection. Here, we examine two genes involved in male reproduction in mice for evidence of positive selection.

Sry. In mammals, male sex determination is defined by the initiation of testis development. There is now considerable evidence that *Sry* is the testis-determining gene (Gubbay et al. 1990; Sinclair et al. 1990): *Sry* is male-specific in all mammal species examined thus far (Sinclair et al. 1990; Foster et al. 1992; Bianchi and Bianchi 1993; but see Bianchi et al. 1993),

the gene is expressed in the genital ridge at the time of testis development (Gubbay et al. 1990; Koopman et al. 1991b), and a 14.5-kb fragment of mouse DNA containing *Sry* is capable of inducing testis development when introduced into chromosomally female mouse embryos (Koopman et al. 1991a).

Using distance-based comparisons of synonymous and nonsynonymous changes between *Mus musculus* and six other species of murine rodents, Tucker and Lundrigan (1993) demonstrated a high rate of nonsynonymous substitution in *Sry* and suggested that this gene may have evolved under positive selection in murine rodents. The likelihood tests performed here include more species than those examined by Tucker and Lundrigan (1993) and explicitly use a phylogenetic framework to estimate ω ratios. These tests confirm that *Sry* has an elevated rate of replacement substitution in these rodents. Despite the fact that no pairwise comparison of the d_N/d_S ratio exceeds 1 for these mice (Fig. 2), both models of positive selection (M2 and M8) fit this gene much better than the corresponding models of neutrality (M1 and M7). In addition, we identify a single site (site 92) with a high posterior probability (>99.9%) of being under positive selection when the empirical Bayesian approach is applied (Table 2).

Sry is a single-exon gene with an open reading frame consisting of a highly conserved DNA-binding domain (the HMG box), flanked by N- and C-terminal regions of variable length (Sinclair et al. 1990; Gubbay et al. 1990). Site 92 is in the C-terminal region of the gene (see Tucker and Lundrigan [1993] for *Sry* sequence and position numbering). In murine rodents, this region contains a CAG repeat motif and is highly length variable, ranging from 92 amino acids in *Hylomyscus alleni* to 313 amino acids in *Mus m. musculus* (Tucker and Lundrigan 1993). Length variation in the C-terminal region is evident even between subspecies of house mice: this region is only 153–155 amino acids long in *M. m. domesticus*, having been shortened from the 313 amino acids in *M. m. musculus* by mutation to a stop codon (Coward et al. 1994; Tucker and Lundrigan 1995). These observations suggest that the C-terminal region of *Sry* is of little functional importance (Sinclair et al. 1990). It is puzzling that a site that shows such strong evidence of positive selection should occur in such a region. We suggest that the rapid rate of evolution for site 92 is more consistent with an explanation of relaxed functional constraints (but see Tucker and Lundrigan 1995) than one of positive selection. However, the paradox of these results should also prompt additional mutational and transgenic studies of this region and of site 92 in particular.

Tcp-1. The *Tcp-1* locus has been mapped to the region of chromosome 17 responsible for transmission-ratio distortion and male sterility produced by the mouse t complex (Silver et al. 1979; Lyon 1991). The *Tcp-1* gene codes for a nonglycosylated protein that is part of the testicular cell surface matrix (Silver and White 1982). While the role of *Tcp-1* in t-complex effects remains unknown, certain observations argue for its role in normal spermatogenesis. First, the gene is expressed at higher levels in mouse testes than in any other tissue examined, and the translation of *Tcp-1* mRNA is up regulated during spermatogenesis (Silver and White 1982; Dudley et al. 1984; Willison et al. 1986; Kubota et al. 1992). Second, while information on the function of *Tcp-1* is scarce, it has been suggested that that mutant forms of *Tcp-1* may be involved in the complicated effects of the t haplotype on spermatogenesis (Silver 1981; Silver and White 1982; Willison et al. 1986). If true, then the presence of the wild-type *Tcp-1* allele may be required for normal spermatogenesis and fertility in male mice.

Prior comparative studies of the *Tcp-1* gene in rodents have not revealed any evidence of positive selection acting on this gene despite its potential role in spermatogenesis. Morita et al. (1992) examined pairwise divergence in *Tcp-1* between *Rattus* and nine species of *Mus* and found that the gene exhibits an unusually high synonymous substitution rate but highly conserved amino acid sequences. Our study includes four additional species of *Mus*. Only the lineage leading to *M. spretus* (branch 12; Fig. 1) shows an unusually high ω ratio (Table 1), but a constant rate ratio cannot be rejected for the tree as a whole and the elevated ω value for the *M. spretus* lineage cannot be considered statistically meaningful. Moreover, models that allow for selection on specific sites (M8 and M2) do not fit the *Tcp-1* data better than corresponding neutral models (M7 and M1); therefore, evidence of positive selection affecting specific sites as identified under M8 (Table 2) is not compelling. These results suggest that the pattern of evolution for *Tcp-1* can best be described as purifying selection.

Selection on the Immune System Locus β_2m

The β_2m protein is a small polypeptide that has been implicated in a number of functional roles. The protein appears to be responsible for inducing collagenase synthesis in fibroblasts (Brinckerhoff et al. 1989) and can stimulate collagen, protein, and DNA synthesis in osteoblast cultures (Canalis et al. 1987). However, the principal role of the β_2m protein appears to be in the immune response. β_2m comprises part of the Fc receptor that mediates the uptake of IgG from milk in neonatal rat intestinal cells (Simister and Mostov 1989), and it is a critical component in ensuring the correct presentation of cell surface antigens (Bjorkman et al. 1987; Hansen et al. 1989;

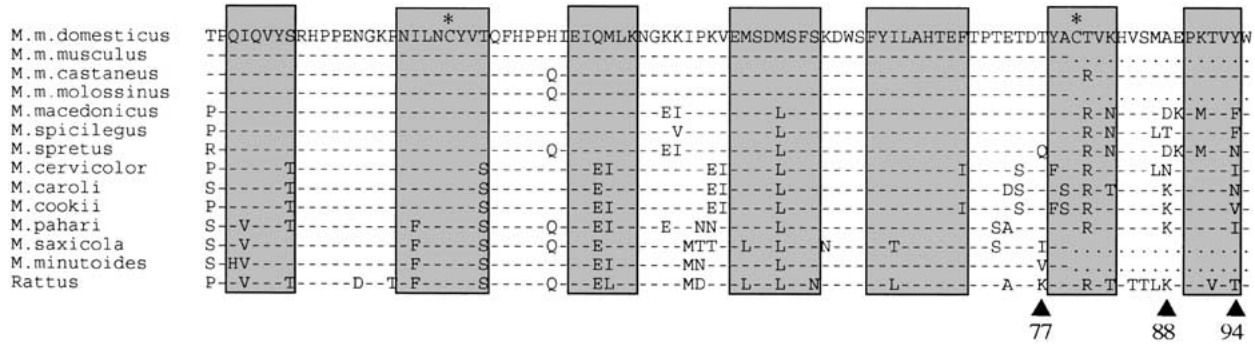

Fig. 4. Aligned amino acid sequences of the $\beta_2 m$ locus for the species included in this study. *Boxes* identify the seven β strands that fold to determine the tertiary structure of the protein (Bjorkman et al. 1987). *Asterisks* indicate the two conserved Cys residues that form a disulfide bond in the mature $\beta_2 m$ molecule. *Triangles* identify specific sites that have a posterior probability >50% of being under positive selection (Table 2); amino acid site numbers are given beneath the triangles. Missing data are shown as *dots*.

Williams et al. 1989; Vitiello et al. 1990; Hermel et al. 1993). Mice that are deficient for the $\beta_2 m$ gene develop normally but fail to assemble MHC class I proteins or to present them on the cell surface (Zijlstra et al. 1990; Koller et al. 1990); moreover, intestinal cells from $\beta_2 m$-deficient neonates fail to bind IgG (Zijlstra et al. 1990). While $\beta_2 m$-deficient mice can clear viral infections under laboratory conditions (Spriggs et al. 1992), they would undoubtedly be severely disadvantaged in the wild by their immunodeficiency (Bender et al. 1992; Tarleton et al. 1992; Hermel et al. 1993).

Prior studies of $\beta_2 m$ sequence divergence in mice suggest that this gene did indeed evolve in response to positive selection. Hermel et al. (1993) found an exceptionally high ratio of replacement-to-silent changes in pairwise comparisons between *M. spretus* and *M. domesticus* and attributed this to positive selection during the evolution of *M. spretus*. Our study expands the taxon sampling to include several additional species of *Mus* and suggests that positive selection on $\beta_2 m$ is not restricted to the *M. spretus* lineage. The free-ratio likelihood model assigns an ω ratio greater than 1 to several branches in the *Mus* phylogeny (Table 1). Many of these branches have one (or fewer) inferred replacement changes, which may not be statistically meaningful; however, three lineages have sufficient numbers of changes to warrant the suggestion of positive selection (Table 1). Specifically, $\beta_2 m$ appears to have been subject to three recent independent episodes of positive selection: one in the lineage leading to *M. musculus* (branch 7; Fig. 1), one in the *M. spretus* lineage (branch 12), and one in the lineage leading to *M. cookii* and *M. cervicolor* (branches 16 and 18). It is not clear why $\beta_2 m$ should be subject to positive selection for these three lineages; nonetheless, it appears that recent periodic episodes of positive selection have punctuated the evolution of the $\beta_2 m$ gene during the evolution of these mice.

In addition, we identify three sites in the $\beta_2 m$ molecule that have a high probability of having evolved under positive selection (Table 2, Fig. 4). The crystal structure for the $\beta_2 m$ molecule has been determined for cow and human (Becker and Reeke 1985; Bjorkman et al. 1987), and the particular sites where $\beta_2 m$ interacts with the α chains of the MHC molecule have been identified for human (Bjorkman et al. 1987). Only two of the sites that interact with the α chains (sites 11 and 22) show amino acid differences among the species of mice we examined, and neither of these shows a high probability of being under positive selection. In contrast, we find three sites (sites 77, 88, and 94) toward the carboxy terminal of the protein that have a high posterior probability of being under positive selection (Fig. 4).

In an intriguing study, Pérarnau et al. (1990) found that polymorphism in site 88 of $\beta_2 m$ influences the ability of MHC class I molecules to present antigens to the immune system (Pérarnau et al. refer to the polymorphic site as 85. They do not include the three-residue signal peptide in numbering their sequence; therefore, their site 85 is our site 88). Specifically, *Mus* MHC molecules comprised of heavy chains complexed with an alternate form of $\beta_2 m$ ($\beta_2 m^b$) which differs from the normal allele ($\beta_2 m^a$) only at site 88, fail to present specific peptides to T cells. Therefore, it appears that while $\beta_2 m$ is considered primarily a structural molecule, changes in specific sites that do not necessarily interact with the α heavy chains can alter ligand recognition for the MHC molecule. In our study, we identify site 88 as a site with a high posterior probability (77%) of being under positive selection, suggesting that selection on this site of the $\beta_2 m$ molecule may have acted to promote the ability of MHC class I molecules to recognize a diversity of antigens. While the role of the two additional sites in this region that show evidence of positive selection (sites 77 and 94) is unknown, they should be the focus of further functional studies of the $\beta_2 m$ molecule.

Conclusions

Perhaps the most powerful application of the phylogeny-based, maximum-likelihood methods applied in this study lies in their ability to identify particular sites within a gene that show the signature of positive selection. We identified a number of such sites in the male sex-determining locus *Sry*, the mammalian gamete recognition gene *Zp-3*, and the immune response gene $\beta_2 m$. Inferences of positive selection resulting from molecular evolution studies are most powerful when functional regions of a gene are well characterized. For example, of the three sites we identified in $\beta_2 m$, one is known to influence the MHC class I molecule's role in the immune response; therefore, selection on this site is consistent with evolution promoting the diversity of immune system proteins. When sites are found in regions of little functional importance, as is the case for the single site in *Sry*, relaxation of functional constraints, rather than positive selection, may better explain the high replacement rate. In those instances where the function of a gene is incompletely characterized, evolutionary studies can identify good candidate sites for careful examination. For example, all of the rapidly evolving sites in *Zp-3* lie outside of the well-characterized sperm-binding region; however, the functional role of these sites has yet to be examined. Studies of gene evolution can therefore predict regions of functional importance and direct future experimental studies of gene function.

Acknowledgments. We thank Keith Barker for help in navigating PAML and for critiquing the manuscript. Phil Myers, Scott Stanley, and two anonymous reviewers provided useful comments on early versions of this paper. S.A.J. was supported by a Smithsonian Institution Scholarly Studies Grant during the later stages of this project. This research was supported in part by National Science Foundation Grant DEB-9209950.

References

Becker JW, Reeke GN (1985) Three-dimensional structure of b2-microglobulin. Proc Natl Acad Sci USA 82:4225–4229

Bender BS, Croghan T, Zhang L, Small PA (1992) Transgenic mice lacking class I major histocompatibility complex-restricted T cells have delayed viral clearance and increased mortality after influenza virus challenge. J Exp Med 175:1143–1145

Bianchi NO, Bianchi MS (1993) Male specific *Sry* patterns in marsupials. J Mammal 74:531–534

Bianchi NO, Bianchi MS, Bailliet G, de la Chapelle A (1993) Characterization and sequencing of the sex determining region Y gene (*Sry*) in *Akodon* (Cricetidae) species with sex reversed females. Chromosoma 102:389–395

Bielawski JP, Dunn KA, Yang Z (2000) Rates of nucleotide substitution and mammalian nuclear gene evolution: Approximate and maximum-likelihood methods lead to different conclusions. Genetics 156:1299–1308

Bjorkman PJ, Saper MA, Samraoui B, Bennett WS, Strominger JL, Wiley DC (1987) Structure of the human class I histocompatibility antigen, HLA-A2. Nature 329:506–512

Bleil JD, Wassarman PM (1980a) Structure and function of the zona pellucida: Identification and characterization of the proteins of the mouse oocyte's zona pellucida. Dev Biol 76:185–202

Bleil JD, Wassarman PM (1980b) Mammalian sperm-egg interaction: Identification of a glycoprotein in mouse egg zona pellucida possessing receptor activity for sperm. Cell 20:873–882

Bleil JD, Wassarman PM (1988) Galactose at the nonreducing terminus of O-linked oligosaccharides of mouse egg zona pellucida glycoprotein ZP3 is essential for the glycoprotein's sperm receptor activity. Proc Natl Acad Sci USA 85:6778–6782

Brinckerhoff CE, Mitchell TI, Karmilowicz MJ, Kluve-Beckerman B, Benson MD (1989) Autocrine induction of collagenase by serum amyloid A-like and β_2-microglobulin-like proteins. Science 243:655–657

Canalis E, McCarthy T, Centrella M (1987) A bone-derived growth factor isolated from rat calvariae is beta2 microglobulin. Endocrinology 121:1198–1200

Chang MC, Hancock JL (1967) Experimental hybridization. In: Benirschke K (ed) Comparative aspects of reproductive failure. Springer-Verlag, New York, pp 206–217

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 95:6196–6197

Civetta A, Singh RS (1998) Sex-related genes, directional sexual selection and speciation. Mol Biol Evol 15:901–909

Clark A Civetta A (2000) Protamine wars. Nature 403:261–263

Coward P, Nagai K, Chen D, Thomas HD, Nagamine CM, Lau Y-FC (1994) Polymorphism of a CAG trinucleotide repeat within Sry correlates with B6Ydom sex reversal. Nature Genet 6:245–250

Dudley K, Potter J, Lyon MF, Willison KR (1984) Analysis of male sterile mutations in the mouse using haploid stage expressed cDNA probes. Nucl Acids Res 12:4281–4293

Eberhard WG (1985) Sexual selection and animal genitalia. Harvard University Press, Cambridge, MA

Endo T, Ikeo K, Gojobori T (1996) Large-scale search for genes on which positive selection may operate. Mol Biol Evol 13:685–690

Ferris PJ, Pavlovic C, Pabry S, Goodenough UW (1997) Rapid evolution of sex-related genes in Chlamydomonas. Proc Natl Acad Sci USA 94:8634–8639

Florman HM, Wassarman PM (1985) O-linked oligosaccharides of mouse egg ZP3 account for its sperm receptor activity. Cell 41:313–324

Foster JW, Brennan FE, Hampikian FK, Goodfellow PM, Sinclair AH, Lovell-Badge R, Selwood L, Renfree MB, Cooper DW, Marshall-Graves JA (1992) Evolution of sex determination and the Y chromosome: SRY-related sequences in marsupials. Nature 359:531–533

Fukuda Y, Maddock MB, Chang MC (1979) In vitro fertilization of two species of deer mouse eggs by homologous or heterologous sperm and penetration of laboratory mouse eggs by deer mouse sperm. J Exp Zool 207:481–490

Goldman N, Yang Z (1994) A codon-based model of nucleotide substitution for protein-coding DNA sequences. Mol Biol Evol 11:725–736

Gubbay J, Colignon J, Koopman P, Capel B, Economou A, Munsterberg A, Vivian N, Goodfellow P, Lovell-Badge R (1990) A gene mapping to the sex-determining region of the mouse Y chromosome is a member of a novel family of embryonically expressed genes. Nature 346:245–250

Hanada A, Chang MC (1978) Penetration of the zona-free or intact eggs by foreign spermatozoa and the fertilization of deer mouse eggs in vitro. J Exp Zool 203:277–286

Hansen TH, Myers NB, Lee DR (1989) Studies of two antigenic forms of L^d with disparate β2-microglobulin ($\beta_2 m$) associations suggest that $\beta_2 m$ facilitates the folding of the α1 and α2 domains during de novo synthesis. J immunol 140:3522–3527

Hermel E, Robinson PJ, She J-X, Fischer-Lindahl K (1993) Sequence divergence of $\beta_2 m$ alleles of wild *Mus musculus* and *Mus spretus* implies positive selection. Immunogenetics 38:106–116

Hughes AL, Nei M (1988) Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. Nature 355:167–170

Hughes AL, Nei M (1989) Nucleotide substitution at major histocompatibility complex class II loci: Evidence for overdominant selection. Proc Natl Sci USA 86:958–962

Ina Y (1995) New methods for estimating the numbers of synonymous and nonsynonymous substitutions. J Mol Evol 40:190–226

Kimura M (1983) The neutral theory of molecular evolution. Cambridge University Press, Cambridge

Kinloch RA, Yukata S, Wassarman PM (1995) Mapping the mouse ZP3 combining site for sperm by exon swapping and site-directed mutagenesis. Proc Natl Acad Sci USA 92:263–267

Koller BH, Marrack P, Kappler JW, Smithies O (1990) Normal development of mice deficient in b2M, MHC class I proteins, and CD8+ T cells. Science 248:1227–1230

Koopman P, Gubbay J, Vivian N, Goodfellow P, Lovell-Badge R (1991a) Male development of chromosomally female mice transgenic for *Sry*. Nature 351:117–121

Koopman P, Munsterberg A, Capel B, Vivian N, Lovell-Badge R (1991b) Expression of a candidate sex-determining gene during mouse testis differentiation. Nature 348:450–452

Kubota H, Willison K, Ashworth A, Nozaki M, Miyamoto H, Yamamoto H, Matsushiro A, Morita T (1992) Structure and expression of the gene encoding mouse t-complex polypeptide (*Tcp-1*). Gene 120:207–215

Lambert H (1984) Role of sperm-surface glycoproteins in gamete recognition in two mouse species. J Reprod Fert 70:281–284

Lee YH, Ota T, Vacquier VD (1995) Positive selection is a general phenomenon in the evolution of abalone sperm lysin. Mol Biol Evol 12:231–238

Li W-H (1993) Unbiased estimation of the rates of synonymous and nonsynonymous substitution. J Mol Evol 36:96–99

Li W-H, Wu C-I, Luo CC (1985) A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. Mol Biol Evol 2:150–174

Lundrigan BL, Jansa SA, Tucker PK (2002). Phylogenetic relationships in the genus *Mus*, based on paternally, maternally, and bi-parentally inherited characters. Syst Biol 51:410–431

Lyon MF (1991) The genetic basis of transmission-ratio distortion and male sterility due to the t complex. Am Nat 147:349–358

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

Messier W, Stewart C-B (1997) Episodic adaptive evolution of primate lysozymes. Nature 385:151–154

Metz EC, Palumbi SR (1996) Positive selection and sequence rearrangements generate extensive polymorphism in the gamete recognition protein bindin. Mol Biol Evol 13:397–406

Metz EC, Robles-Sikisaka R, Vacquier VD (1998) Nonsynonymous substitution in abalone sperm fertilization genes exceeds substitution in introns and mitochondrial DNA. Proc Natl Acad Sci USA 95:10676–10681

Morita T, Kubota H, Murata K, Nozaki M, DeLarbre C, Willison K, Satta Y, Sakaizumi M, Takahata N, Gachelin G, Matsushiro A (1992) Evolution of the mouse t-haplotype: Recent and worldwide introgression to *Mus musculus*. Proc Natl Acad Sci USA 89:6851–6855

Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol 3:418–426

Nei M, Zhang J (1998) Molecular origin of species. Science 282:1428–1429

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

Pérarnau B, Siegrist C-A, Gillet A, Vincent C, Kimura S, Lemonnier F (1990) B2-Microglobulin restriction of antigen presentation. Nature 346:751–754

Rankin TL, Tong Z-B, Castle PE, Lee E, Gore-Langton R, Nelson LM, Dean J (1998) Human ZP3 restores fertility in *Zp-3* null mice without affecting order-specific sperm binding. Development 125:2415–2424

Ringuette MJ, Chamberlin ME, Baur AW, Sobieski DA, Dean J (1988) Molecular analysis of cDNA coding for ZP3, a sperm binding protein of the mouse zona pellucida. Dev Biol 127:287–295

Roldan ERS, Yanagimachi R (1989) Cross-fertilization between Syrian and Chinese hamsters. J Exp Zool 250:321–328

Roldan ERS, Vitullo AD, Merani MS, vonLawzewitsch I (1985) Cross fertilization in vivo and in vitro between three species of vesper mice, *Calomys* (Rodentia, Cricetidae). J Exp Zool 233:433–442

Rossiere TK, Wassarman PM (1992) Identification of a region of mouse zona pellucida glycoprotein mZP3 that possesses sperm receptor activity. Dev Biol 154:309–317

Silver LM (1981) A structural gene (*Tcp-1*) within the mouse t complex is separable from effects on tail length and lethality but may be involved in effects on spermatogenesis. Genet Res 38:115–123

Silver LM, White M (1982) A gene product of the mouse t complex with chemical properties of a cell surface-associated component of the extracellular matrix. Dev Biol 91:423–430

Silver LM, Arzt K, Bennett D (1979) A major testicular cell protein specified by a mouse T/t complex gene. Cell 17:275–284

Simister NE, Mostov KE (1989) An Fc receptor structurally related to MHC class I antigens. Nature 337:184–187

Sinclair AH, Berta P, Palmer MS, Hawkins JR, Griffiths BL, Smith MJ, Foster JW, Frischauf A, Lovell-Badge R, Goodfellow PN (1990) A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif . Nature 346:240–244

Skutelsky E, Ranen E, Shlagi R (1994) Variations in the distribution of sulfur residues in the zona pellucida as possible species-specific determinants of mammalian oocytes. J Reprod Fertil 100:35–41

Spriggs MK, Koller BH, Sato T, Morrisset PJ, Fanslow WC, Smithies O, Voice RF, Widmer MB, Maliszewski CR (1992) β2-Microglobulin-, CD8+ T-cell deficient mice survive inoculation with high doses of vaccinia virus and exhibit altered IgG responses. Proc Natl Acad Sci USA 89:6070–6074

Suzuki Y, Gojobori T (1999) A method for detecting positive selection at single amino acid sites. Mol Biol Evol 16:1315–1328

Swanson WJ, Vacquier VD (1995) Extraordinary divergence and positive Darwinian selection in a fusagenic protein coating the acrosomal process of abalone spermatozoa. Proc Natl Acad Sci USA 92:4957–4961

Swanson WJ, Yang Z, Wolfner MF, Aquadro CF (2001) Positive Darwinian selection drives the evolution of several female re-

productive proteins in mammals. Proc Natl Acad Sci USA 98:2509–2514

Tarleton RL, Koller BH, Latour A, Pstan M (1992) Susceptibility of β2-microglobulin-deficient mice to *Trypanosoma cruzi* infection. Nature 356:338–340

Ting C-T, Tsaur S-C, Wu M-L, Wu C-I (1998) A rapidly evolving homeobox at the site of a hybrid sterility gene. Science 282:1501–1504

Tsaur S-C, Wu C-I (1997) Positive selection and the molecular evolution of a gene of male reproduction, Acp26Aa of Drosophila. Mol Biol Evol 14:544–549

Tsaur S-C, Ting C-T, Wu C-I (1998) Positive selection driving the evolution of a gene of male reproduction, Acp26Aa, of Drosophila: II. Divergence versus polymorphism. Mol Biol Evol 15:1040–1046

Tucker PK, Lundrigan BL (1993) Rapid evolution of the sex determining locus in old world mice and rats. Nature 364:715–717

Tucker PK, Lundrigan BL (1995) The nature of gene evolution on the mammalian Y chromosome: Lessons from Sry. Phil Trans R Soc Lond B 350:221–227

Vacquier VD, Swanson WJ, Lee Y-H (1997) Positive Darwinian selection on two homologous fertilization proteins: What is the selective pressure driving their divergence? J Mol Evol 44 (Suppl 1):S15–S22

Vitiello A, Potter TA, Sherman LA (1990) The role of b2-microglobulin in peptide binding by class I molecules. Science 250:1423–1426

Wassarman PM (1990) Profile of a mammalian sperm receptor. Development 108:1–17

Wassarman PM, Litscher ES (1995) Sperm-egg recognition mechanisms in mammals. Curr Topics Dev Biol 30:1–19

West JD, Frels WI, Papaioannou VE, Karr JP, Chapman VM (1977) Development of interspecific hybrids of *Mus*. J. Embryol Exp Morphol 41:233–243

Willet C (2000) Evidence for directional selection acting on pheromone-binding proteins in the genus *Choristoneura*. Mol Biol Evol 17:553–562

Williams DB, Barber BH, Flavell RA, Allen H (1989) Role of β2-microglobulin in the intracellular transport and surface expression of murine class I histocompatibility molecules. J Immunol 142:2796–2806

Willison KR, Dudley K, Potter J (1986) Molecular cloning and sequence analysis of a haploid expressed gene encoding t complex polypeptide 1. Cell 44:727–738

Wycoff GJ, Wang W, Wu C-I (2000) Rapid evolution of male reproductive genes in the descent of man. Nature 403:304–309

Yang Z (1997) PAML: A program package for phylogenetic analysis by maximum likelihood. CABIOS 13:555–556

Yang Z (1998) Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. Mol Biol Evol 15:568–573

Yang Z (2000) Phylogenetic analysis using maximum likelihood (PAML). Version 3.0. University College London, London

Yang Z, Bielawski JP (2000) Statistical methods for detecting molecular adaptation. Trends Ecol Evol 15:496–502

Yang Z, Nielsen R (2000) Estimating synonymous and nonsynonympus substitution rates under realistic evolutionary models. Mol Biol Evol 17:32–43

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

Yang Z, Swanson WJ, Vacquier VD (2000b) Maximum-likelihood analysis of molecular adaptation in abalone sperm lysin reveals variable selective pressures among lineages and sites. Mol Biol Evol 17:1446–1455

Zijlstra M, Bix M, Simister NE, Loring JM, Raulet DH, Jaenisch R (1990) B2-Microglobulin deficient mice lack CD4-8+ cytolytic T cells. Nature 344:742–746