Positive and Negative Selection in the *DAZ* Gene Family

Joseph P. Bielawski and Ziheng Yang

Department of Biology, Galton Laboratory, University College London, London, England

Because a microdeletion containing the *DAZ* gene is the most frequently observed deletion in infertile men, the *DAZ* gene was considered a strong candidate for the azoospermia factor. A recent evolutionary analysis, however, suggested that *DAZ* was free from functional constraints and consequently played little or no role in human spermatogenesis. The major evidence for this surprising conclusion is that the nonsynonymous substitution rate is similar to the synonymous rate and to the rate in introns. In this study, we reexamined the evolution of the *DAZ* gene family by using maximum-likelihood methods, which accommodate variable selective pressures among sites or among branches. The results suggest that *DAZ* is not free from functional constraints. Most amino acids in *DAZ* are under strong selective constraint, while a few sites are under diversifying selection with nonsynonymous/synonymous rate ratios (d_N/d_S) well above 1. As a result, the average d_N/d_S ratio over sites is not a sensible measure of selective pressure on the protein. Lineage-specific analysis indicated that human members of this gene family were evolving by positive Darwinian selection, although the evidence was not strong.

Introduction

Azoospermia is the most common form of infertility in human males (Shinka and Nakahori 1996). A locus on the human Y chromosome, the azoospermic factor (AZF), is believed to contain a gene, or genes, crucial to proper differentiation of male germ cells. The observation that microdeletions at three different loci of AZF occur in 5%–15% of infertile men supports this hypothesis (Ferlin et al. 1999). One of these loci (AZFc) encodes the Deleted in AZoospermia (*DAZ*) gene. Because AZFc is the most frequently observed deletion in infertile men, it was considered a strong candidate for the azoospermia factor (Ferlin et al. 1999). Genes from a number of different pathways, however, are required for normal spermatogenesis (Elliott and Cooke 1997).

DAZ is located on the Y chromosome, but it is closely related to the autosomal gene *DAZL1*. While *DAZL1* is present in all vertebrates, *DAZ* is found only in Old World Monkeys. Thus, *DAZ* [Yq11.23] is believed to have evolved via translocation of *DAZL1* [3p24] to the Y chromosome (Saxena et al. 1996; Gromoll et al. 1999) some time after the divergence of Old and New World monkeys; Kumar and Hedges (1998) dated that divergence to about 40 MYA. After the translocation event, *DAZ* underwent a series of rearrangements and a modified copy was amplified, yielding a Y gene cluster.

DAZ and *DAZL1* have a functional role in fertility. Both *DAZ* and *DAZL1* are expressed exclusively in germ cells (Cooke et al. 1996; Ruggiu et al. 1997; Gromoll et al. 1999), and in humans *DAZ* expression is highest in spermatogonia (Menke, Mutter, and Page 1997). Experimental elimination of *DAZL1* in mice results in termination of germ cell development beyond the spermatoginial stage (Ruggiu et al. 1997). Moreover, Y-encoded human *DAZ* can compliment the sterile phenotype of *DAZL1* null mice, yielding a partial recovery of spermatogenesis, which suggests the same or similar target mRNA for *DAZ* and *DAZL1* during spermatogenesis (Slee et al. 1999). Although the specific functions of *DAZ* and *DAZL1* are unknown, the presence of RNA recognition motifs suggests that these genes could be involved in controlling the cell cycle switch from mitotic to meiotic cell division (Gromoll et al. 1999); this cell cycle switch is controlled by RNA-binding proteins in yeast (Watanabe et al. 1997).

Surprisingly, a recent evolutionary analysis of the *DAZ* family (*DAZ* and *DAZL1* genes) indicated a lack of functional constraints on *DAZ*. Agulnik et al. (1998) found a high rate of nonsynonymous substitution, similar rates between exons and introns, and similar rates among the three codon positions. They concluded that there were no functional constraints on evolution of *DAZ* and that patterns of sequence divergence were due to neutral drift. They hypothesized that Y-linked *DAZ* played little role in human spermatogenesis.

The nonsynonymous-to-synonymous rate ratio (d_N/d_S) provides a sensitive measure of selective pressure on the protein. However, when selection pressure varies among amino acid sites, the average d_N/d_S ratio might not be very informative about the evolutionary processes affecting the gene. The objective of this study was to investigate the role of both purifying and positive selection on the *DAZ* gene family by using maximum-likelihood methods that accommodate differences in selective pressures among sites (Nielsen and Yang 1998; Yang et al. 2000). Our findings indicated that *DAZ* was not free of functional constraints and that other explanations for its rapid rate of nonsynonymous evolution must be considered. There has been considerable debate as to whether rapid evolution in gene families is caused by positive Darwinian selection after gene duplication (Ohta 1993) or by relaxation, but not complete loss, of functional constraints in redundant genes (Kimura 1983; Li 1985). In the latter case, a new function might evolve when formerly neutral substitutions convey a selective advantage in a novel environment or genetic background (Dykhuizen and Hartl 1980). We also examined variable selective pressures among lineages (Yang 1998; Yang and Nielsen 1998), and our findings suggest that both

Key words: DAZ, DAZL1, gene family, maximum likelihood, codon model, positive selection.

Address for correspondence and reprints: J. P. Bielawski, Department of Biology, University College London, 4 Stephenson Way, London NW1 2HE, United Kingdom. E-mail: j.bielawski@ucl.ac.uk.

Mol. Biol. Evol. 18(4):523–529. 2001

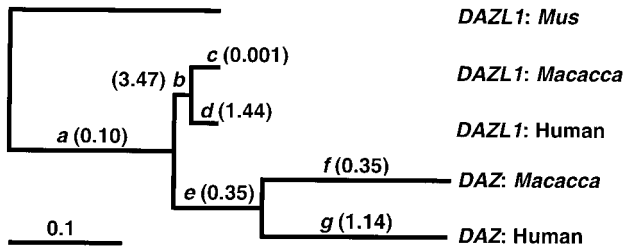

FIG. 1.—Phylogeny for the *DAZ* gene family. This topology was recovered from a maximum-likelihood analysis of nucleotide sequences and also from least-squares analysis of synonymous divergence. Branch lengths are proportional to the mean number of nucleotide substitutions per codon as inferred under model D: free ratios (Yang 1998). All analyses were conducted using unrooted topologies; this topology is rooted for convenience. Numbers in parentheses are branch-specific ω ratios estimated under model D.

models could have played a role in the evolution of the *DAZ* gene family.

Materials and Methods

Sequence Data

Two data sets were compiled, reflecting a trade-off between more characters versus more taxa. Data set 1 was composed of 618 bp of DNA sequence (after exclusion of gaps) from five representatives of the *DAZ* gene family (fig. 1). Sequences of *DAZL1* were from *Homo sapiens* (GenBank accession number U066078), *Macaca mulatta* (AF053608), and *Mus musculus* (U046694), and sequences of *DAZ* were from *H. sapiens* (NM004081) and *Macaca fascicularis* (AJ012216). Sequences included exons 1–6, A7, C8, and 10. *Macaca fascicularis DAZ* (AJ012216) contains an intragenic duplication of exons 2–6 and multiple copies of exons 7 and 8. We sampled the 3′ copy of exons 2–6, which is 99% similar to the 5′ copy. The ''A'' copy of exon 7 and the ''C'' copy of exon 8 were sampled because each predates divergence of the human and Macaca lineages (Gromoll et al. 1999). Data set 1 was used to investigate variation in selective pressure among lineages. To study variation in selective pressure among sites, however, more sequences were needed. Hence, a second data set was compiled consisting of 11 members of the *DAZ* gene family (fig. 2), but only 291 bp of DNA sequence. Included in data set 2 were exons 3–5 and portions of exons 2 and 6. This data set was composed almost entirely of the RNA recognition domain, which spans exons 2–5. Data set 2 included *DAZL1* from *Cebua apella* (AF053608), *H. sapiens* (U066078), *M. mulatta* (AF053608), and *Papio hamadryas* (AF053607); a single copy of *DAZ* from *H. sapiens* (NM004081), *Pan troglodytes* (AF072324), and *M. fascicularis* (AJ012216); and two *DAZ* clones (C1 and C2) from *P. hamadryas* (C1: AF07230; C2: AF07321) and *M. mulatta* (C1: AF072322; C2: AF072323). Clones from *P. hamadryas* and *M. mulatta* were divergent copies from a multicopy *DAZ* array on the Y chromosome (Agulnick et al. 1998). Saxena et al. (2000) recently reported that human *DAZ* genes occur as a four-copy array in the AZFc region of the Y chromosome. However, they found that the four copies differed by only a single, silent, transition in exon 7A, so only one copy was included in our analysis.

Data Analysis

Tree topologies were estimated using maximum likelihood (ML) under the general time-reversible (GTR) model with a discrete gamma model (dΓ) of rate variation among sites (Yang 1994*a*, 1994*b*). Trees also were estimated by least-squares from synonymous divergences estimated by ML under a codon model of evolution (Goldman and Yang 1994). The PAUP* computer program (Swofford 2000) was used for conducting tree searches.

We implemented four nested models of variable selective pressures among branches (Yang 1998; Yang and Nielsen 1998). Model A was the simplest and assumed the same ω ratio for all branches. Models B and C were based on the prediction that a gene family evolves under different selective pressures following gene duplication. Model B assumed two ω ratios: one for the branch predating the translocation to the Y chromosome (fig. 1; branch *a*), and a second for branches postdating the translocation (branches *b–g*). Model C assumed three ω ratios: one for branch *a*, one for *DAZL1* branches postdating the translocation (branches *b–d*), and one for all *DAZ* branches (branches *e–g*). Model D (free ratios) assumed an independent ω ratio for each branch of a topology and was employed to evaluate the potential for positive selection in any one branch of the tree.

ML models (Yang and Nielsen 2000; Yang et al. 2000) also permit testing and identification of selective pressures at individual codon sites. We implemented three such models: M3 (discrete), M7 (beta), and M8 (beta&ω). M3 assumed two site classes with the proportions f_0 and f_1 and ratios ω_0 and ω_1 estimated from the data. M7 assumed that ω ratios were distributed among sites according to a beta distribution. Depending on parameters p and q, the beta distribution can take a variety of shapes within the interval (0, 1). M8, an extension of M7, added an extra class of sites having an ω parameter freely estimated from the data. Positive selection was indicated when an ω parameter of M3 or M8 was >1. The likelihood ratio test was used to compare a one-ratio model (M0) with M3 and to compare M7 with M8. If there were sites with $\omega > 1$, Bayesian methods were used to calculate the posterior probability that a site fell into each site class; sites with high probabilities for $\omega > 1$ were likely to be under positive Darwinian selection (Yang et al. 2000).

All ML analyses of codon models were performed using the codeml program of the PAML package (Yang 1999). The models employed correction for transition/transversion rate bias and codon usage bias, features of DNA sequence evolution that have a significant effect on the estimation of substitution rates (Yang and Nielsen 1998, 2000).

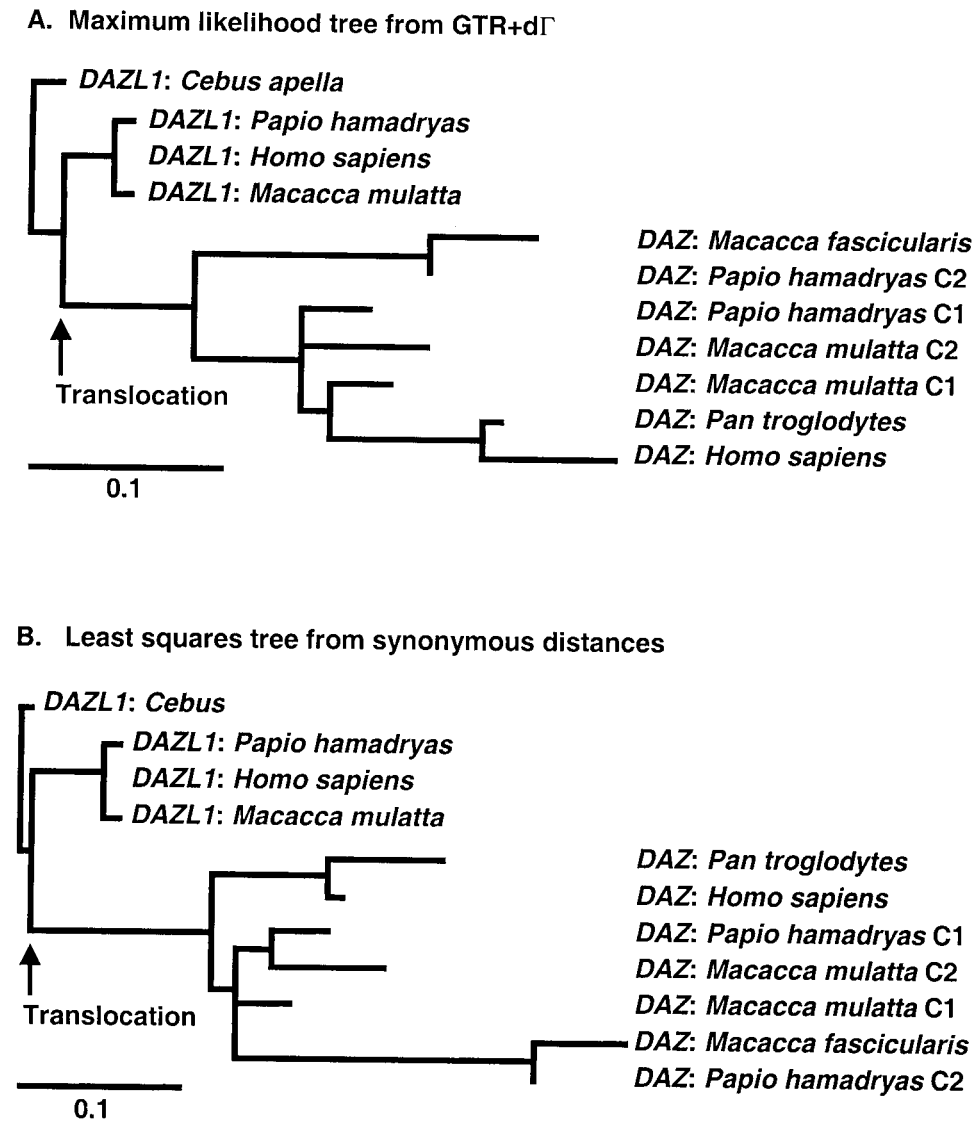

FIG. 2.—Candidate topologies for the *DAZ* gene family in primates. *A*, Tree topology recovered from a maximum-likelihood analysis under the GTR substitution matrix combined with a gamma correction for among-sites rate variation. *B*, Tree topology recovered from least-squares analysis of synonymous divergence. Branch lengths are proportional to the mean number of nucleotide substitutions per codon as inferred under model M8: beta&ω (Yang et al. 2000). All analyses were conducted using unrooted topologies; these topologies are rooted for convenience.

Results

Variable Selection Pressure Among Lineages

Phylogenetic analyses of data set 1 by different tree reconstruction methods yielded the same tree topology (fig. 1), and this topology was used to analyze variable selective pressures among lineages. Four models (A–D) were fitted by ML to data set 1 (table 1). The estimate of ω for model A ($\omega = 0.295$) represented an average over all codon sites and branches. Model A was then compared with model B, which assumed that selective constraints changed after translocation of *DAZL1* to the Y chromosome. Twice the difference in their likelihood scores (2δ) was compared with a χ^2 distribution with degrees of freedom equal to the difference between models in number of parameters. This likelihood ratio test indicated that model B provided a significantly better fit to these data ($2\delta = 17.8$, df = 1, $P = 0.000025$). The d_N/d_S ratio after the translocation, $\omega_1 = 0.51$, is significantly higher than that prior to the translocation, $\omega_0 = 0.10$.

Model B assumed that both autosomal *DAZL1* and the copy that was translocated to the Y chromosome (*DAZ*) experienced the same change in selective constraints after the translocation event. This simple model was compared with a more complex model (model C) in which changes in selective constraints after the translocation were allowed to differ between *DAZL1* and *DAZ* (table 1). Estimates under model C indicate very different ω values for *DAZ* and *DAZL1* after duplication (table 1). However, the likelihood of model C was not significantly better than that of model B ($2\delta = 1.12$, df = 1, $P = 0.29$).

Because positive selection at any one point in the phylogeny could have affected our results, we applied

Table 1
Log Likelihood Scores and Parameter Estimates Under Models of Variable Selection Pressures Among Lineages

Model	*p*	Parameters for Branches	ℓ
A: One ratio	1	$\omega_0 = 0.295$ for all branches	−1,442.44
B: Two ratios	2	$\omega_0 = 0.102$ for branch *a*	−1,433.52
		$\omega_1 = 0.513$ for branches *b, c, d, e, f,* and *g*	
C: Three ratios	3	$\omega_0 = 0.103$ for branch *a*	−1,432.96
		$\omega_1 = 0.290$ for branches *b, c,* and *d*	
		$\omega_2 = 0.574$ for branches *e, f* and *g*	
D: Free ratios	7	$\omega_0 = 0.100$ for branch *a*	−1,426.40
		$\omega_1 =$ **3.474** for branch *b*	
		$\omega_2 = 0.001$ for branch *c*	
		$\omega_3 =$ **1.444** for branch *d*	
		$\omega_4 = 0.350$ for branch *e*	
		$\omega_5 = 0.355$ for branch *f*	
		$\omega_6 =$ **1.144** for branch *g*	

NOTE.—Analyses were conducted using κ as a free parameter and the F61 model of equilibrium codon frequencies. *p* is the number of branch-specific ω parameters. ω ratios greater than 1 are in bold. Branches are defined in figure 1.

the free-ratios model (model D) to the same data. The likelihood score under model D was significantly better than that obtained for model B ($2\delta = 14.2$, df = 5, $P = 0.014$). Branches *b, d,* and *g* exhibited ω values >1 (table 1). Use of the simpler but less realistic F3×4 model, which calculates codon frequencies by using base composition at the three codon positions, produced similar results. Note that ω for branch *g* was slightly less than 1 under the F3×4 model ($\omega_6 = 0.99$), whereas it was greater than 1 under the F61 model ($\omega_6 = 1.144$).

Variable Selection Pressure Among Sites

Phylogenetic analysis of data set 2 under the nucleotide model GTR+dΓ recovered a topology in which divergent copies of *DAZ* from the same species were not monophyletic, indicating that divergent copies of *DAZ* originated in an early amplification event and persisted in multiple lineages (fig. 2*A*). This result is similar to that obtained in a previous analysis of the *DAZ* gene family (Agulnik et al. 1998). We also inferred a tree topology from synonymous divergences (fig. 2*B*). This tree, although different from the tree obtained from the nucleotide analysis, also indicated that some copies of *DAZ* originated from an early amplification event and persisted to the present day. Both trees also indicate a clear bifurcation between all *DAZL1* and *DAZ* sequences, supporting the hypothesis that a single translocation event gave rise to the Y-encoded *DAZ*. To investigate the impact of tree topology, models of variable ω values among sites were analyzed using both topologies in figure 2 (table 2). The small size of data set 2 (291 bp; 97 codons) prevented use of the parameter-rich model of empirical codon frequencies (the F61 model), and the F3×4 model was used instead.

The discrete model (M3), which allowed two site classes with independent ω ratios, provided a significant improvement over the one-ratio model (M0) regardless of the tree topology assumed (table 3). The selective pressure is not uniform among amino acid sites. Estimates of parameters under M3 suggest that most sites (95%–97%) are under selective constraint, with $\omega_0 =$ 0.35–0.37, while a few sites (3%–5%) are evolving by positive selection, with ω_1 close to 6. Both models, M3 and M8, which allowed for the presence of positively

Table 2
Log Likelihood Scores and Parameter Estimates for Four Models of Variable ω's Among Sites and Two Tree Topologies

Model	Parameter Estimates	Positively Selected Sites	ℓ
Tree 1 (fig. 2*A*)			
M0: One ratio	$\omega = 0.47$	None	−747.19
M3: Discrete	$\omega_0 = 0.37$, $f_0 = 0.97$	26, **28,** 42	−742.72
	$\omega_1 = 5.66$ ($f_1 = 0.03$)		
M7: Beta	$p = 0.63$, $q = 0.78$	Not allowed	−745.84
M8: Beta&ω	$p = 2.2$, $q = 3.1$, $f_0 = 0.98$	**28**	−742.70
	$\omega_1 = 12.47$ ($f_1 = 0.02$)		
Tree 2 (fig. 2*B*)			
M0: One ratio	$\omega = 0.47$	None	−758.47
M3: Discrete	$\omega_0 = 0.35$, $f_0 = 0.95$	**26, 28,** 42, **91**	−748.90
	$\omega_1 = 5.96$ ($f_1 = 0.05$)		
M7: Beta	$p = 0.30$, $q = 0.37$	Not allowed	−754.99
M8: Beta&ω	$p = 107$, $q = 197$, $f_0 = 0.95$	**26, 28,** 42, **91**	−748.91
	$\omega_1 = 5.70$ ($f_1 = 0.05$)		

NOTE.—*p* and *q* are parameters of the beta distribution. *f* is the proportion of sites assigned to an individual ω category or to a beta distribution with shape parameters *p* and *q*. The proportion f_1 (in parentheses) is not a free parameter. Positively selected sites are those with posterior probabilities (P) > 0.50, and those with $P > 0.95$ are in bold.

Table 3
Likelihood Ratio Statistic (2δ) for Comparing Models of Variable ω's Among Sites

	M3 vs. M0	M8 vs. M7
Tree 1 (fig. 2*A*)	8.94*	6.82*
Tree 2 (fig 2*B*)	19.14*	12.16*

NOTE.—See table 2 for model parameters.
* Significant at the 5% level ($\chi^2_{5\%} = 5.99$, df = 2).

selected sites indicated that some variation in selective pressure was due to positive selection (table 2). Likelihood ratio tests indicated that these models fit the data better than models in which positively selected sites were not allowed (table 3). It is also noteworthy that regardless of model or topology, ω values for sites not subject to positive Darwinian selection were well below 1 (table 2), indicating evolution by purifying selection.

Agulnick et al. (1998) hypothesized that there were no functional constraints on *DAZ* sequences. To test this hypothesis specifically, we reanalyzed only the *DAZ* sequences of data set 2. The results were consistent with the previous analysis of data set 2; ω values for those sites not subject to positive Darwinian selection were well below 1 (e.g., tree 1—discrete model: $\omega_0 = 0.43$, $f_0 = 0.93$, $\omega_1 = 3.4$, $f_1 = 0.07$; beta&ω model—$p = 98$, $q = 122$, $f_0 = 0.94$, $\omega_1 = 4.1$, $f_1 = 0.06$).

Discussion

Maximum-likelihood analysis of the *DAZ* gene family revealed significant variation in selective pressures among lineages and among sites. The majority of sites are clearly subject to purifying selection, with the nonsynonymous rate being well below the synonymous rate. A small fraction of sites exhibit nonsynonymous rates almost six times the synonymous rate, indicating the action of positive Darwinian selection. Lineage-specific analyses indicated that following the translocation of an autosomal copy of *DAZL1* to the Y chromosome, both loci experienced increased rates of nonsynonymous substitution. In *DAZL1* this was due, at least in part, to early evolution by positive Darwinian selection. Later, *DAZL1* of *M. fascicularis* returned to evolution by purifying selection, whereas *DAZL1* of humans continued to evolve by positive Darwinian selection. Although there was also an increase in nonsynonymous substitution in *DAZ*, its early evolution was most consistent with purifying selection. Nevertheless, more recent evolution in *DAZ* shows the same pattern as *DAZL1*, with human evolution by positive Darwinian selection and *M. fascicularis* evolution by purifying selection.

Based on an evolutionary analysis of the *DAZ* gene family, Agulnik et al. (1998) concluded that there were no functional constraints on the evolution of *DAZ* in primates and questioned the role of *DAZ* in human spermatogenesis. Our findings, however, indicate that the majority of sites in *DAZ* are subject to purifying selection. This notion is supported by the observation of intact reading frames for all the sampled exons of *DAZ*; there are no frameshift mutations or premature stop codons. Moreover, complementation of sterile-phenotype *DAZL1* mice by human *DAZ* strongly suggests a functional role for human *DAZ* in spermatogenesis (Slee et al. 1999). These observations, taken together with expression patterns of *DAZ*, lead us to conclude that *DAZ* is not free from functional constraints in primates and that the *DAZ* gene is likely to have functional importance in human spermatogenesis.

The pairwise approach used by Agulnick et al. (1998) is the most common method of computing synonymous and nonsynonymous rates. This approach, however, averages rates over all sites and also over the entire time interval that separates a pair of sequences. Agulnick et al. (1998) did not observe d_N/d_S ratios in excess of 1 in most comparisons because evolution by positive selection occurred at a subset of sites and only in certain lineages of *DAZ*. This example is not unique, as the pairwise approach also failed to detect positive selection in HIV (Leigh Brown 1997; Crandall et al. 1999; Zanotto et al. 1999). Moreover, the same effect led to the incorrect conclusion that the κ-casein gene was free from functional constraint (Ward, Honeycutt, and Derr 1997). These studies indicate that for some proteins the traditional approach to estimating the d_N/d_S ratio might not provide a sensible measure of selective pressure.

Gene duplication is considered an important mechanism for functional divergence (Ohno 1970; Ohta 1993). However, the process by which duplicated genes acquire new functions is less clear. There is often an acceleration of the rate of evolution following gene duplication (Li 1985; Ohta 1993, 1994). Accelerated rates could initially be driven by positive Darwinian selection for functional divergence (Ohta 1993, 1994) or by relaxation of selective constraints. In the latter case, it is thought that random fixation of neutral changes eventually leads to a novel function in one or both copies. This model was referred to as the ''Dykhuizen-Hartl effect'' by Zhang, Rosenberg, and Nei (1998). Our findings are consistent with both models. The elevated rate of nonsynonymous substitution in autosomal *DAZL1* following its duplication appears to result from the action of positive Darwinian selection. However, elevated rates of nonsynonymous substitution in *DAZ* immediately following its origin via the translocation event appear to result from decreased levels of purifying selection, suggesting a possible role for the Dykhuizen-Hartl effect in the early stages of *DAZ* evolution.

Our findings are consistent with several studies of gene families in which positive Darwinian selection was shown to be at least partially responsible for a rate increase following gene duplication (Ohta 1993, 1994; Zhang, Rosenberg, and Nei 1998; Duda and Palumbi 1999; Rooney and Zhang 1999; Schmidt, Goodman, and Grossman 1999). However, in the only other case in which the relative contribution of both models was investigated, the Dykhuizen-Hartl effect was ruled out (Zhang, Rosenberg, and Nei 1998). Although in this respect our findings appear to differ, it is important to point out that the method we employed to accommodate variation in selective pressures among lineages calcu-

lated ω as an average across all sites. Because an episode of positive selection at a subset of sites could elevate ω at a specific branch without causing it to exceed 1, it is not possible to completely rule out the positive-selection model. More complex models which can simultaneously accommodate rate variation among sites and lineages are under development and might be useful in distinguishing between positive selection and the Dykhuizen-Hartl effect.

Darwinian selection appears to be a relatively common feature of mammalian reproductive proteins (Karn and Nachman 1999; Rooney and Zhang 1999; Wyckoff, Wang, and Wu 2000). Our findings indicate the *DAZ* gene family represents another example of this pattern. However, not all lineages of the *DAZ* gene family are presently evolving by positive Darwinian selection; both *DAZ* and *DAZL1* of *M. fascicularis* are evolving by purifying selection. With regard to the difference between humans and *M. fascicularis*, it is interesting to note that the mature *DAZ* protein of humans has only 1 processed copy of exon 8, whereas the mature *DAZ* protein of *M. fascicularis* has 10 processed copies of exon 8. The DNA sequence for *DAZ* in both species contains multiple copies of both exons 7 and 8. However, at some point in the human lineage, a mutation disabled the splice sites of exons 8A and 8D. Because all but one copy of exon 8 in present-day human *DAZ* are descended from the disabled copies of exons 8A and 8D, the mature *DAZ* protein of humans includes only one processed copy of exon 8 (Gromoll et al. 1999). Because the open reading frame was preserved in *M. fascicularis* despite several duplication and rearrangement events, multiple copies of exons 7 and 8 must have evolved functional importance during divergence of this gene family (Gromoll et al. 1999). It is tempting to speculate that in the human lineage a loss of processing of all but one copy of exon 8 initiated adaptive evolution at other sites in both *DAZ* and *DAZL1* to maintain proper spermatogenesis. Additional sequences of *DAZ* and *DAZL1* from a variety of primate species are needed to understand the role of positive selection in functional divergence of the *DAZ* gene family.

Acknowledgments

We thank Katherine A. Dunn for helpful comments and discussion. We also thank the associate editor and two anonymous reviewers for their constructive comments. This study was supported by a Biotechnology and Biological Sciences Research Council grant (31/G10434) to Z.Y.

LITERATURE CITED

AGULNIK, A. I., A. ZHARKIKH, H. BOETTGER-TONG, T. BOURGERON, K. MCELREAVEY, and C. E. BISHOP. 1998. Evolution of the DAZ gene family suggests that Y-linked *DAZ* plays little, or a limited, role in spermatogenesis but underlines a recent African origin for human populations. Hum. Mol. Genet. **7**:1371–1377.

COOKE, H. J., M. LEE, S. KERR, and M. RUGGIU. 1996. A murine homologue of the human *DAZ* gene is autosomal and expressed only in male and female gonads. Hum. Mol. Genet. **5**:513–516.

CRANDALL, K. A., C. R. KELSEY, H. IMANICHI, H. C. LANE, and N. P. SALZMAN. 1999. Parallel evolution of drug resistance in HIV: failure of nonsynonymous/synonymous substitution rate ratio to detect selection. Mol. Biol. Evol. **16**: 372–382.

DUDA, T. F., and S. R. PALUMBI. 1999. Molecular genetics of ecological diversification: duplication and rapid evolution of toxin genes of the venomous gastropod *Conus.* Proc. Natl. Acad. Sci. USA **96**:6820–6823.

DYKHUIZEN, D., and D. L. HARTL. 1980. Selective neutrality of 6PGD allozymes in *E. coli* and the effects of genetic background. Genetics **96**:801–817.

ELLIOTT, D. J., and H. J. COOKE. 1997. The molecular genetics of male infertility. Bioessays **19**:801–809.

FERLIN, A., E. MORO, A. GAROLLA, and C. FORESTA. 1999. Human male infertility and Y chromosome deletions: role of the AZF- candidate genes *DAZ, RBM* and *DFFRY*. Hum. Reprod. **14**:1710–1716.

GOLDMAN, N., and Z. YANG. 1994. A codon based model of nucleotide substitution for protein-coding DNA sequences. Mol. Biol. Evol. **11**:725–736.

GROMOLL, J., G. F. WEINBAUER, H. SKALETSKY, S. SCHLATT, M. ROCCHIETTI-MARCH, D. C. PAGE, and E. NIESCHLAG. 1999. The Old World monkey *DAZ* (Deleted in AZoospermia) gene yields insights into the evolution of the *DAZ* gene cluster on the human Y chromosome. Hum. Mol. Genet. **8**: 2017–2024.

KARN, R. C., and M. W. NACHMAN. 1999. Reduced nucleotide variability at an androgen-binding protein locus (*Abpa*) in house mice: evidence for positive natural selection. Mol. Biol. Evol. **16**:1192–1197.

KIMURA, M. 1983. The neutral theory of molecular evolution. Cambridge University Press, Cambridge, England.

KUMAR, S., and B. HEDGES. 1998. A molecular timescale for vertebrate evolution. Nature **392**:917–920.

LEIGH BROWN, A. J. 1997. Analysis of HIV-1 *env* gene reveals evidence for a low effective number in the viral population. Proc. Natl. Acad. Sci. USA **94**:1862–1865.

LI, W.-H. 1985. Accelerated evolution following gene duplication and its implications for the neutralist-selectionist controversy. Pp. 333–352 *in* T. OHTA and K. AOKI, eds. Population genetics and molecular evolution. Japan Scientific Press, Tokyo.

MENKE, D. B., G. L. MUTTER, and D. C. PAGE. 1997. Expression of *DAZ*, an azoospermia factor candidate, in human spermatogonia. Am. J. Hum. Genet. **60**:237–241.

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

OHNO, S. 1970. Evolution by gene duplication. Springer-Verlag, Berlin.

OHTA, T. 1993. Pattern of nucleotide substitution in growth hormone-prolactin gene family: a paradigm for evolution by gene duplication. Genetics **134**:1271–1276.

———. 1994. Further examples of evolution by gene duplication revealed through DNA sequence comparisons. Genetics **138**:1331–1337.

ROONEY, A. P., and J. ZHANG. 1999. Rapid evolution of primate sperm protein: relaxation of functional constraint or positive Darwinian selection? Mol. Biol. Evol. **16**:706–710.

RUGGIU, M., R. SPEED, M. TAGGART, S. J. MCKAY, F. KILANOWSKI, P. SAUNDERS, J. DORIN, and H. J. COOKE. 1997. The mouse *DAZL1a* gene encodes a cytoplasmic protein essential for gametogenesis. Nature **389**:73–77.

SAXENA, R., L. G. BROWN, T. HAWKINS et al. (11 co-authors). 1996. The *DAZ* gene cluster on the human Y chromosome arose from an autosomal gene that was transposed, repeatedly amplified and pruned. Nat. Genet. **14**:292–299.

SAXENA, R., J. W. A. DE VRIES, S. REPPING, R. K. ALAGAPPAN, H. SKALETSKY, L. G. BROWN, P. MA, E. CHEN, J. M. N. HOOVERS, and D. C. PAGE. 2000. Four *DAZ* genes in two clusters found in the *AZFc* region of the human Y chromosome. Genomics **67**:256–267.

SCHMIDT, T. R., M. GOODMAN, and L. I. GROSSMAN. 1999. Molecular evolution of the *COX7A* gene family in primates. Mol. Biol. Evol. **16**:619–626.

SHINKA, T., and Y. NAKAHORI. 1996. The azoospermic factor on the Y chromosome. Acta Paediatr. Jpn. **38**:399–404.

SLEE, R., B. GRIMES, R. M. SPEED, M. TAGGART, S. M. MAGUIRE, A. ROSS, N. I. MCGILL, P. T. SAUNDERS, and H. J. COOKE. 1999. A human *DAZ* transgene confers partial rescue of the mouse *Dazl* null phenotype. Proc. Natl. Acad. Sci. USA **96**:8040–8045.

SWOFFORD, D. L. 2000. PAUP*. Phylogenetic analysis using parsimony (*and other methods). Version 4. Sinauer, Sunderland, Mass.

WARD, T. J., R. L. HONEYCUTT, and J. N. DERR. 1997. Nucleotide sequence evolution at the kappa-casein locus: evidence for positive selection within the family Bovidae. Genetics **147**:1863–1872.

WATANABE, Y., S. SHINOZAKI-YABANA, Y. CHIKASHIGE, Y. HIRAOKA, and M. YAMAMOTO. 1997. Phosphorylation of RNA-binding protein controls cell cycle switch from mitotic to meiotic in fission yeast. Nature **386**:187–190.

WYCOFF, G. J., W. WANG, and C.-I. WU. 2000. Rapid evolution of male reproductive genes in the descent of man. Nature **403**:304–309.

YANG, Z. 1994*a*. Estimating the pattern of nucleotide substitution. J. Mol. Evol. **39**:105–111.

———. 1994*b*. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. J. Mol. Evol. **39**:306–314.

———. 1998. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. Mol. Biol. Evol. **15**:568–573.

———. 1999. Phylogenetic analysis by maximum likelihood (PAML). Version 2. University College London, England.

YANG, Z., and R. NIELSEN. 1998. Synonymous and nonsynonymous rate variation in nuclear genes of mammals. J. Mol. Evol. **46**:409–418.

———. 2000. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. Mol. Biol. Evol. **17**:32–43.

YANG, Z., R. NIELSEN, N. GOLDMAN, and A.-M. K. PEDERSON. 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics **155**:431–449.

ZANOTTO, P. M. DE A., E. G. KALLAS, R. F. DE SOUZA, and E. C. HOLMES. 1999. Genealogical evidence for positive selection in the *nef* gene of HIV-1. Genetics **153**:1077–1089.

ZHANG, J., H. F. ROSENBERG, and M. NEI. 1998. Positive Darwinian selection after gene duplication in primate ribonuclease genes. Proc. Natl. Acad. Sci. USA **95**:3708–3713.

EDWARD HOLMES, reviewing editor

Accepted November 20, 2000