

Rapid Evolution and Gene-Specific Patterns of Selection for Three Genes of Spermatogenesis in *Drosophila*

Alberto Civetta,* Sujeetha A. Rajakumar,*†¹ Barb Brouwers,* and John P. Bacik*²

*Department of Biology, University of Winnipeg, Winnipeg, Manitoba, Canada; and †Department of Biochemistry and Medical Genetics, University of Manitoba, Winnipeg, Manitoba, Canada

Hybrid males resulting from crosses between closely related species of *Drosophila* are sterile. The F1 hybrid sterility phenotype is mainly due to defects occurring during late stages of development that relate to sperm individualization, and so genes controlling sperm development may have been subjected to selective diversification between species. It is also possible that genes of spermatogenesis experience selective constraints given their role in a developmental pathway. We analyzed the molecular evolution of three genes playing a role during the sperm developmental pathway in *Drosophila* at an early (*bam*), a mid (*aly*), and a late (*dj*) stage. The complete coding region of these genes was sequenced in different strains of *Drosophila melanogaster* and *Drosophila simulans*. All three genes showed rapid divergence between species, with larger numbers of nonsynonymous to synonymous differences between species than polymorphisms. Although this could be interpreted as evidence for positive selection at all three genes, formal tests of selection do not support such a conclusion. Departures from neutrality were detected only for *dj* and *bam* but not *aly*. The role played by selection is unique and determined by gene-specific characteristics rather than site of expression. In *dj*, the departure was due to a high proportion of neutral synonymous polymorphisms in *D. simulans*, and there was evidence of purifying selection maintaining a high lysine amino acid protein content that is characteristic of other DNA-binding proteins. The earliest spermatogenesis gene surveyed, which plays a role in both male and female gametogenesis, was *bam*, and its significant departure from neutrality was due to an excess of nonsynonymous substitutions between species. *Bam* is degraded at the end of mitosis, and rapid evolutionary changes among species might be a characteristic shared with other degradable transient proteins. However, the large number of nonsynonymous changes between *D. melanogaster* and *D. simulans* and a phylogenetic comparative analysis among species confirms evidence of positive selection driving the evolution of *Bam* and suggests an yet unknown germ cell line developmental adaptive change between these two species.

Introduction

Hybrid males between closely related species of *Drosophila* are usually sterile, and such a form of postzygotic reproductive isolation may be at least partially a consequence of rapid evolution of male reproductive genes (Civetta and Singh 1998; Presgraves et al. 2003; Orr, Masly, and Presgraves 2004). A genomic comparative approach has found higher amino acid sequence divergence between *Drosophila melanogaster* and *Drosophila pseudoobscura* for genes expressed in the testes than the genome-wide average (Richards et al. 2005). Other studies point at reproductive traits and genes, more specifically genes with a role in sperm differentiation and maturation, as targets of drastic change during species differentiation. For example, microscopy analyses of spermatogenesis in F1 hybrids from crosses between *Drosophila simulans*, *Drosophila sechellia*, and *Drosophila mauritiana* show that hybrid male sterility results from a combination of premeiotic and postmeiotic defects (Kulathinal and Singh 1998). The expression of genes in late stages of sperm development are downregulated in interspecific male *Drosophila* hybrids (Michalak and Noor 2003). Highly divergent genes involved in sex-related regulatory functions are also good candidates for genes that promote speciation (Ting et al. 1998; Orr and Presgraves 2000). For example,

the *Odysseus* locus of *Drosophila* contains a homeodomain thought to play a role in regulation of gene transcription in spermiogenesis that shows a very high rate of replacement substitutions among closely related species (Ting et al. 1998; Nei and Zhang 1998).

While genes with a role in sperm development might be targets of rapid change, developmental genes in general are supposed to be widely conserved during evolution. In particular, genes that control early stages of differentiation and therefore influence a cascade of events are supposed to be under stronger selective constraints than genes with a role in later stages of development (Richter et al. 1997; Wilkins 2002). Spermatogenesis in *Drosophila* is a relatively well-understood developmental process both in terms of cellular and subcellular changes that define stages and mutations at specific genes that affect them (Fuller 1993, 1998; Wakimoto, Lindsley, and Herrera 2004). Therefore, spermatogenesis provides an excellent model system to study the molecular evolution of a group of developmental genes that can impact male fertility and be important in determining sex-specific adaptations among species. Although several genes have been identified to play different roles during sperm development, mutagenesis identifies genes that act at key regulatory stages in the spermatogenesis pathway of *Drosophila*: the transition from mitotic division producing spermatocytes into meiosis, the progression of the meiotic cell cycle and the onset of spermatid differentiation, and a final step of spermatid differentiation (spermiogenesis) marked by remodeling of subcellular sperm components such as growth and organization of the long flagellar axoneme, elongation of specialized mitochondrial derivative for motility, and DNA condensation and nuclear shaping (Fuller 1998).

¹ Present address: Lady Davis Institute for Medical Research, McGill University, Montreal, Quebec, Canada.

² Present address: Department of Medical Microbiology and Immunology, University of Alberta, Edmonton, Alberta, Canada.

Key words: population genetics, spermatogenesis, *Drosophila*, selection.

E-mail: a.civetta@uwinnipeg.ca.

Mol. Biol. Evol. 23(3):655–662. 2006

doi:10.1093/molbev/msj074

Advance Access publication December 15, 2005

In this study, we targeted three genes with roles in the mitosis—meiosis transition (*bag of marbles*), meiotic progression (*always early*), and spermiogenesis (*don juan*) (see fig. 4, Fuller 1998). *bag of marbles* (*bam*) encodes a protein needed to cease mitotic division and begin meiotic division of both the male and female germ-line cell lineage. However, in females, *bam* is also required for the germ-line stem cells (GSCs) to differentiate into progenitor cells of oogenesis (McKearin and Spradling 1990; McKearin and Ohlstein 1995). *always early* (*aly*) is a meiotic arrest gene required for the accumulation of other meiotic arrest class gene products during meiosis, so *aly* plays a central role in regulating the progress of meiosis (Fuller 1998; Jiang and White-Cooper 2003). *aly* also plays an important role in modifying chromatin structure and triggering transcription of genes required in spermiogenesis (Lin, Viswanathan, and Wood 1996; White-Cooper et al. 1998; Jiang and White-Cooper 2003). Finally, *don juan* (*dj*) encodes a basic lysine-rich protein expressed in the tail of mature sperm cells and involved in maturation of elongated spermatids during spermiogenesis (Santel et al. 1997). Given the basic nature of the protein, it is thought to also function as a DNA-binding protein with a possible role during the final phase of chromatin condensation in spermatids (Santel et al. 1997, 1998).

We analyzed patterns of sequence polymorphism and interspecies sequence divergence using a North American and an African population of *D. melanogaster* together with a *D. simulans* population to test for signs of selection shaping the evolution of three sex-related genes that control different stages of sperm differentiation and maturation in *Drosophila*. We find high sequence divergence in all three genes but departures from neutrality only for *bam* and *don juan*. In *don juan* the departure is due to an excess of synonymous polymorphism in *D. simulans* that are mainly localized within a lysine-rich domain under purifying selection. The gene *bam* shows an excess of amino acid substitutions between *D. melanogaster* and *D. simulans*, which are indicative of adaptive diversification between species, and signs of lineage-specific positive selection in a phylogenetic comparative analysis including other species of *Drosophila*.

Materials and Methods

DNA Samples and Sequencing

Flies were maintained on a cornmeal molasses media, and DNA was extracted from pools of approximately 10 adult flies using a standard extraction protocol. We sampled 16 North American isofemale lines of *D. melanogaster* (collected in Winnipeg and established in 2001 by A.C.), nine *D. melanogaster* lines from Zimbabwe (Africa) and 6 *D. simulans* lines from California. The entire coding region of *bam*, *aly*, and *dj* genes were polymerase chain reaction (PCR) amplified from DNA extractions using oligonucleotide primers designed from published *D. melanogaster* sequence information (GenBank accession numbers: NM_057452, NM_080118 and NM_169163). PCR products were cleaned using the Wizard SV gel and PCR Clean-up system kit (Promega, Madison, Wisc.) and sequenced on a Beckman Coulter CEQ 2000XL automated sequencer using protocols provided by the manu-

facturers. Primer sequences used for PCR amplifications and PCR sequencing reactions along with conditions for each gene analyzed are available upon request. All sequencing reactions were done using both forward and reverse primers (both strands) from two independent PCR products, and a consensus sequence for each strain was obtained from alignments using the program ClustalX 1.81 (Thompson et al. 1997) with manual adjustments. All sequences can be found in GenBank under accession numbers DQ315948–DQ316036.

Sequence Data Analyses

Sequences from all *Drosophila* strains were aligned using ClustalX 1.81 (Thompson et al. 1997), and analysis of polymorphism and divergence was carried out using the program DnaSP 4.0 (Rozas et al. 2003) and MEGA 3.0 (Kumar, Tamura and Nei 2004). DNA polymorphism was quantified using estimates of π , the average pairwise difference between haplotypes (Nei 1987) and θ , the expected nucleotide diversity under neutrality determined from the number of segregating sites (Watterson 1975). The tests of Tajima (1989) and Fu and Li (1993) were used to detect signs of selection from the distribution of polymorphisms. Tajima's *D* test measures the differences between the number of segregating sites and the average number of nucleotide differences. The Fu and Li tests are based on the differences between the number of singletons and the total number of mutations (D^*) or the average number of nucleotide differences between pairs of sequences (F^*). The *H* test of Fay and Wu (2000) was used to distinguish between selective sweep and background selection, and the test is based on the differences between the average number of nucleotide differences between pairs of sequences and an estimator of the frequency of derived variants. Statistical significance was tested using a coalescence simulation with 10,000 replications. The rate of nonsynonymous-synonymous substitutions per nonsynonymous-synonymous site (Nei and Gojobori 1986) was calculated for both polymorphism and interspecies comparisons. Overall and codon nucleotide composition as well as GC content were estimated using both DnaSP and MEGA. The test of McDonald and Kreitman (1991), which compares interspecific and intraspecific fixed and polymorphic synonymous and replacement changes, was used to detect selection. Protein sequence domains and signals were identified using the MotifScan tool available at PROSITE (<http://ca.expasy.org/prosite/>) or predicted using the computer program PSORT II (<http://psort.nibb.ac.jp>).

We tested for positive selection for the three genes sequenced by using phylogenetic analysis by maximum likelihood (PAML v3.14) (Yang 1997) among species of *Drosophila*. For this purpose, our *D. melanogaster* and *D. simulans* gene sequences were used to Blast the *Drosophila* species trace files available from GenBank (<http://www.ncbi.nlm.nih.gov>), and traces were assembled into a single gene contig using Sequence Manager (DNASTar). Aligned sequences were tested using different models within the program codonml available in the PAML software package. The site-specific models test for signs of positive selection by allowing the ω (d_N/d_S) ratio to vary among

Table 1
Number of Silent and Replacement Polymorphisms within
***Drosophila melanogaster* (Winnipeg and Zimbabwe Samples)**
and *Drosophila simulans*

Gene	Strain/ Species	Sample Size	Number of Sites	Replacement	Silent	Total
<i>bam</i>	Wpg	16	1449	3	7	10
	Zim	8	1452	3	12	15
	<i>D.sim</i>	5	1452	4	6	10
<i>aly</i>	Wpg	15	1664	10	22	32
	Zim	9	1652	21	40	61
	<i>D.sim</i>	6	1658	12	13	25
<i>dj</i>	Wpg	16	803	3	4	7
	Zim	9	805	0	3	3
	<i>D.sim</i>	5	802	6	34	40

NOTE.—Wpg, Winnipeg; Zim, Zimbabwe; *D.sim*, *Drosophila simulans*.

codon sites but not among lineages (Yang et al. 2000). Positive selection is inferred from comparing log-likelihood estimates of a tree topology under models that make different assumptions in terms of ω variation over codon sites. We compared log-likelihood results from models 1a and model 2a. Model 1a (M1a) is a nearly neutral model in which ω is estimated from the data and it is assumed to freely vary between 0 and 1, and model 2a (M2a) is a positive selection model that considers the possibility of codon sites evolving under positive selection ($\omega > 1$) (Wong et al. 2004). We also compared log-likelihood estimates from model M7, in which ω is assumed to range between 0 and 1 following a beta distribution, and model M8 which allows codon sites with $\omega > 1$ (Yang et al. 2000; Wong et al. 2004).

It is possible that positive selection is restricted to specific lineages and few amino acid sites within a protein, so we also examined a set of models that allow ω to vary among codon sites and lineages within a phylogeny (Yang and Nielsen 2002). This is branch-site model A (model = 2; NSsites = 2), in which the background lineages vary in ω between 0 and 1 and the lineage or foreground branch to test is allowed $\omega > 1$. The log likelihood of the branch-site model A is compared to the log likelihood of the nearly neutral model M1a, and if significant, the posterior proba-

bility that a codon site within the tested branch is under positive selection is estimated using the Bayes empirical Bayes procedure implemented in codonml (Yang, Wong, and Nielsen 2005). The log-likelihood of model A is also compared to the same model but fixing the ω value of the foreground branch to 1 so that any significant variation in ω between foreground and background branches can be attributed to positive selection as opposed to differences in selective constraints (Yang and Nielsen 2002).

Results and Discussion

Approximately 1.5 kb including the entire coding region of the *bag of marble* (*bam*) gene, 900 bp of *don juan* (*dj*), and 1.6 kb of *always early* (*aly*) were sequenced and aligned among different strains of *D. melanogaster* and *D. simulans* (Appendix 1, Supplementary Material online). Hudson's permutation based nearest-neighbor test statistic (*Snn*) (Hudson 2000) shows significant levels of genetic differentiation in *bam* (*Snn* = 0.80; $P = 0.004$), *aly* (*Snn* = 0.78; $P < 0.001$), and *dj* (*Snn* = 0.95; $P < 0.001$) between the Zimbabwe and Winnipeg samples of *D. melanogaster*. Subsequent analysis considers *D. melanogaster* Zimbabwe and Winnipeg as separate populations.

Within-Species Polymorphism

Within-species polymorphism was estimated for the coding region of Zimbabwe and Winnipeg samples of *D. melanogaster* and the *D. simulans* California strains. The *dj* gene shows very few polymorphic sites in *D. melanogaster* compared to *bam* and *aly* and there seems to be a higher number of polymorphic sites in the *D. melanogaster* African sample (table 1). There is a higher proportion of silent polymorphisms at the *dj* gene in *D. simulans* than in *D. melanogaster* (table 1). π (π) and θ (θ) estimates of total sequence variation are within those obtained for other *Drosophila* genes in previous studies (Moriyama and Powell 1996; Andolfatto 2001). Two exceptions are *aly* for the Zimbabwe sample and *don juan* for *D. simulans* (table 2), which have experienced an elevated proportion of silent site polymorphisms compared to values reported

Table 2
Summary Statistics of Sequence Polymorphism and Tests of Neutrality

Gene	Pop	N ^a	S ^b	θ^c	π^c	D ^d	D* ^d	F* ^d	H ^e	P(nr) ^e	P(fr) ^e
<i>aly</i>	Wpg	15	32	0.0059	0.0060	0.035	−0.532	−0.431	−7.200	0.085	0.005
	Zim	9	61	0.0136	0.0114	−0.862	−1.023	−1.106	3.889	0.819	0.864
	Sim	6	25	0.0066	0.0059	−0.661	−0.633	−0.698	−1.067	0.276	0.294
<i>bam</i>	Wpg	16	10	0.0017	0.0021	−0.708	−0.786	−0.879	1.433	0.760	0.843
	Zim	8	15	0.0040	0.0040	−0.321	−0.210	−0.262	2.929	0.886	0.973
	Sim	5	10	0.0028	0.0033	−1.193	−1.193	−1.258	−0.300	0.304	0.401
<i>dj</i>	Wpg	16	7	0.0026	0.0025	−0.127	−0.439	−0.407	0.483	0.461	0.583
	Zim	9	3	0.0014	0.0012	−0.552	−0.726	−0.759	0.639	0.751	0.794
	Sim	5	40	0.0239	0.0236	−0.235	−0.235	−0.254	−4.400	0.200	0.080

NOTE.—Wpg, Winnipeg; Zim, Zimbabwe; Sim, *Drosophila simulans*.

^a Sample size.

^b Number of polymorphic positions excluding alignment gaps.

^c Estimates of DNA sequence polymorphism (see *Materials and Methods*).

^d Tajima's and Fu and Li's tests of neutrality (see *Materials and Methods*). All these tests were nonsignificant.

^e *H* test (see *Materials and Methods*) and *P* values for the test under the assumption of no recombination (nr) or free recombination (fr). Significant results ($P < 0.05$) are given in bold.

Table 3
Polymorphism and Divergence in Gene-Coding Regions and the McDonald-Kreitman Test of Selection at Genes of Spermatogenesis

	Locus	P _S ^a	P _R ^a	F _S ^a	F _R ^a	G-test ^b	P Value ^b	P Value ^c
wpg-sim ^d	<i>aly</i>	29 (21, 10)	21 (10, 12)	31	40	1.88	0.171	0.142
	<i>bam</i>	10 (4, 6)	7 (3, 4)	31	66	4.35	0.037*	0.053
	<i>dj</i>	34 (4, 30)	8 (3, 5)	13	9	2.43	0.119	0.078
zim-sim ^d	<i>aly</i>	42 (35, 10)	33 (21, 12)	30	40	2.01	0.156	0.136
	<i>bam</i>	14 (8, 6)	7 (3, 4)	31	68	8.97	0.003**	0.005**
	<i>dj</i>	31 (1, 30)	6 (0, 6)	12	10	4.48	0.034*	0.032*

^a Counts of synonymous and replacement polymorphisms and fixed differences. Numbers in parentheses are counts in *D. melanogaster* and *D. simulans*, respectively.

^b McDonald-Kreitman G-test and P value. * $P < 0.05$; ** $P < 0.01$.

^c Fisher exact test.

^d Comparisons between wpg (*Drosophila melanogaster* Winnipeg strains) or zim (*D. melanogaster* Zimbabwe strains) and sim (*Drosophila simulans*).

for other genes ($\theta_S = 0.036$ and $\theta_S = 0.076$, respectively) (see table 1 in Andolfatto 2001).

Tests of neutrality based on comparisons of heterozygosity found within populations were nonsignificant (table 2). These tests have been found to have low power to demonstrate departures from neutrality, particularly when sample sizes are below 20 (Richter et al. 1997). The negative sign of the tests suggest a tendency toward a larger than expected number of polymorphic sites in low frequency. Rare polymorphisms are expected under either recent selection removing variation (purifying selection) or a positive selective sweep driving sites to fixation (Tajima 1989). The H test of Fay and Wu is used to distinguish between a selective sweep, which will lead to neutral variants in either high frequency (hitchhike to fixation) or low frequency (newly arisen mutations after the selection event), and background selection that is expected to remove polymorphisms leaving only new low-frequency variants that have arisen after the selective event. The test uses the different expectations in terms of frequency of derived mutations since the selective event (Fay and Wu 2000). A negative H test indicates an excess of derived mutations, and it supports a recent selective sweep as responsible for a higher than expected proportion of rare polymorphism. The H test was only negative and significant under the assumption of free recombination for *aly* in the *D. melanogaster* Winnipeg sample (table 2), and the significant result held when using an estimate of recombination obtained from the sample ($R = 59$; $P = 0.019$). However, it is worth noting that the H test can lead to a false detection of selection when loci are chosen without independent evidence of a recent selective sweep or simply as a consequence of sampling (Przeworski 2002).

Nuclear Localization and Mitochondrial Targeting Signals

Aly and Dj contain short amino acid sequences in the amino end of the protein that are predicted signals for transport to the mitochondria (Appendix 1, Supplementary Material online). A mitochondrial localization signal (IRPHH) in exon 1 of *dj* shows no polymorphisms or substitutions both within and between species, and the absence of nucleotide changes within these domains suggests strong

selective constraints most likely related to the fact that Dj is found along the flagellum of the sperm mitochondrial derivatives (Santel et al. 1997, 1998). Despite the presence of a PRKMV mitochondrial targeting signal in Aly, two nonsynonymous polymorphisms were found suggesting poor amino acid sequence conservation. The poor conservation of the mitochondrial targeting signal motif in Aly is likely related to relaxed selective constraints, as Aly is synthesized in the cytoplasm and transported to the nuclei of primary spermatocytes with no localization to the mitochondria (White-Cooper et al. 2000; Jiang and White-Cooper 2003).

All three genes have recognizable nuclear localization signals (NLS) within their amino acid sequences. The Dj protein is found in the nuclei of spermatids during chromatin condensation (Santel et al. 1997, 1998), and the protein contains an eight times repeated motif (DPCKKK) that serves as an NLS (Appendix 1, Supplementary Material online). There is strong conservation of this domain, with 17 synonymous polymorphisms in *D. simulans* but no nonsynonymous polymorphisms or substitutions between species (Appendix 1, Supplementary Material online). The transportation of Aly from the cytoplasm to the nucleus, where it triggers chromatin remodeling and transcription of genes needed for the progression of spermatogenesis (Lin, Viswanathan, and Wood 1996), can be aborted by mutations in the NLS (Jiang and White-Cooper 2003). There is sequence conservation of the NLS regions found in the Aly protein. One region, present in exon 1, has one nonsynonymous polymorphism in *D. simulans* that results in the replacement between two positively charged amino acids (lysine and arginine). Another NLS comprising 17 amino acids is present in exon 2 and has only experienced synonymous polymorphisms and substitutions but no nonsynonymous changes. An NLS is found in Bam, but the domain is poorly conserved with similar numbers of synonymous and nonsynonymous changes between *D. simulans* and *D. melanogaster* (Appendix 1, Supplementary Material online). Such poor conservation of an NLS in Bam suggests no selective constraints and may be due to the fact that the protein is not transported to cell nucleus but found in a germ cell-specific organelle, the fusome, and the cytoplasm (McKearin and Ohlstein 1995).

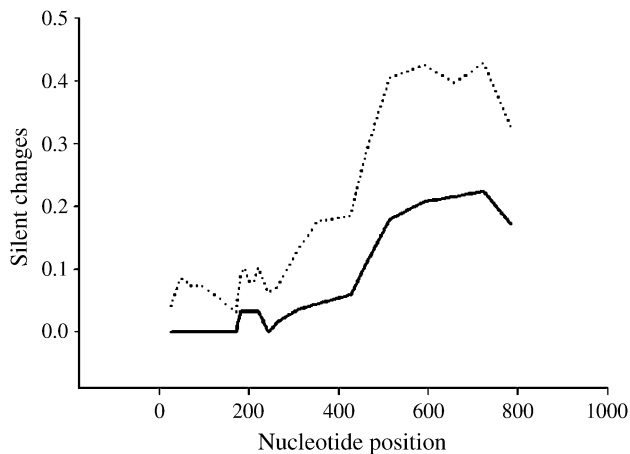


FIG. 1.—Slide window calculation using a 50-bases window and a 10-bases step of the proportion of silent polymorphisms (solid line) in *Drosophila simulans* and silent changes between *D. simulans* and *Drosophila melanogaster* (dotted line) for the *don juan* gene.

Purifying Selection Maintains High Lysine Content in *don juan* While Positive Selection Drives the Evolution of *bag of marble*

All three genes analyzed in this study show the same trend of an excess of nonsynonymous to synonymous fixed differences when compared with nonsynonymous to synonymous polymorphisms (table 3). However, the departure is only significant for the *dj* and *bam* genes in comparisons between *D. melanogaster* Zimbabwe and *D. simulans*. These two genes show also marginally significant departures in *D. melanogaster* Winnipeg–*D. simulans* comparisons (table 3). However, the reasons for the departures from neutrality are different for the two genes. While *bam* shows an excess of nonsynonymous substitutions between species, *dj* shows an excess of synonymous polymorphisms in *D. simulans* (table 3). A comparison of estimates of nonsynonymous and synonymous between species substitutions for the three genes analyzed in this study with other *Drosophila* genes also depicts the rather large proportion of nonsynonymous substitutions at *bam* and of synonymous changes at *dj* ($d_N = 0.072$ and $d_S = 0.229$, respectively; see table 1 in Begun and Whitley [2000] and table 2 in Begun [2002]).

A window analysis of the proportion of silent and synonymous changes shows an elevated proportion of polymorphisms and substitutions in the second exon of the *dj* gene (fig. 1). There is no sign of selection for specific codons driving the accumulation of synonymous changes between *D. simulans* and *D. melanogaster* (Unpreferred: 10/115 = 0.09; Preferred: 13/81 = 0.16; $P = 0.175$) but rather a high proportion of synonymous polymorphisms that are concentrated within a lysine-rich domain in the second exon of *D. simulans* (fig. 1 and Appendix 1, Supplementary Material online). The large proportion of synonymous polymorphisms in *D. simulans* equally involves C-T and A-G changes, and an average 53% AT nucleotide content at synonymous third codon position does not lend support for mutational bias. The high proportion of synonymous polymorphisms appears to be a consequence of the accumu-

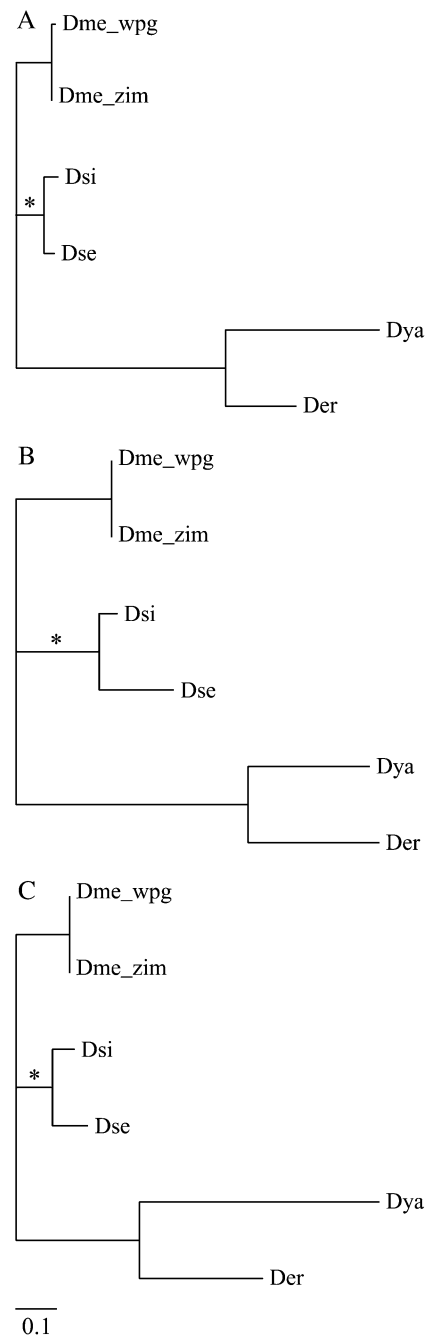


FIG. 2.—Phylogeny of *Drosophila* species for genes *aly* (A), *bam* (B), and *dj* (C). Branch lengths measure the number of nucleotide substitutions per codon. The branch leading to the simulans clade species (asterisk) is the foreground branch for test of positive selection.

lation of neutral mutations that are tolerated as far as the high lysine-rich content is maintained within the second exon of the *dj* gene, where there is a content bias of A nucleotides compared to exon 1 (43% and 29%, respectively). Such bias is particularly strong at the first (45%) and second (54%) codon position that defines AAR triplets for lysine but not at the third more neutral nucleotide position (34%). This form of purifying selection is similar to a previously reported case of purifying selection maintaining high arginine content in the protamine sperm protein among

Table 4
Results from PAML Analysis of the *bam* Gene in *Drosophila*

Model	Parameters	df ^a	L ^b	2Δi ^c	Positively Selected Sites ^d
M1a (nearly neutral)	$p_0 = 0.699$ ($p_1 = 0.301$), $\omega_0 = 0.125$	2	−3547.04		
Model A (test2)	$p_0 = 0.534$, $p_1 = 0.208$, ($p_2 + p_3 = 0.258$)	2	−3545.40		
Model A (test1)	$p_0 = 0.699$, $p_1 = 0.252$, ($p_2 + p_3 = 0.049$), $\omega_2 = 14.605$	3	−3541.15	Versus test 2, 8.5**	7 V 0.51, 23 Q 0.58, 33 E 0.94 , 62 S 0.56
			Versus M1a, 11.8***		99 F 0.59, <i>154 L 0.86</i> , 164 V 0.92 , 198 M 0.59, <i>207 L 0.83</i> , 336 R 0.54, 349 F 0.57, 354 Y 0.58, 355 K 0.53, 395 A 0.61, 406 E 0.57, 423 G 0.66

^a Degrees of freedom are given by the number of independent parameters in each model.

^b Log likelihood of tree topology under model assumption.

^c Likelihood ratio test is two times the difference in log likelihood between models A (test 1 vs. test 2), and model A (test 1) vs. model M1a. Significance is tested using a chi-square test with degree of freedom given by the difference in degrees of freedom (df) between models compared. ** $P < 0.01$; *** $P < 0.001$.

^d Codon sites showing posterior probabilities $P > 0.90$ (bold underlined), $P > 0.80$ (italics underlined) and $P > 0.50$ (plain text) of being under positive selection.

mammalian and nonmammalian vertebrates (Clark and Civetta 2000; Rooney, Zhang, and Nei 2000).

The *bam* gene shows a high proportion of replacement to synonymous fixed differences between species indicative of positive selection (table 3). The comparison of log-likelihood values from site models in codonml for alignments of our consensus *D. melanogaster* African and Winnipeg population plus *D. simulans* with *D. sechellia*, *Drosophila yakuba*, and *Drosophila erecta* sequences retrieved and assembled from GenBank trace files (Appendix 2, Supplementary Material online) showed no evidence of positive selection among sites for either *bam*, *aly*, or *dj*. However, the site models detect positive selection at individual codon sites only if the average d_N over lineages is higher than d_S . Therefore, we used *D. yakuba* sequences as an ancestral sequence to both *D. melanogaster* and *D. simulans* and assigned replacement and synonymous substitutions to either the melanogaster or simulans branch of the three species tree. We found the highest proportion of nonsynonymous to synonymous changes in the simulans branch of *bam* (36/14) compared to melanogaster (27/15) and the ratios for *bam* were higher than for *aly* (simulans: 18/16; melanogaster: 17/12) and *dj* (simulans: 4/3; melanogaster: 5/7). The sign of positive selection detected by the McDonald-Kreitman test and the high proportion of nonsynonymous changes in the simulans lineage prompted us to use the branch-site model of codonml to formally test for selection in the branch leading to the simulans clade (fig. 2). Evidence of positive selection along this branch was detected only for *bam* with approximately 5% of codon sites under positive selection (table 4).

Seven amino acid replacements and only one synonymous substitution between *D. melanogaster* and *D. simulans* are found within a PEST domain in the carboxy-terminal region of Bam and two show a weak sign of positive selection (table 4 and Appendix 1, Supplementary Material online). This domain is typical of unstable proteins that are sensitive to proteases and targeted for degradation (Rogers, Wells and Rechsteiner 1986; Fukuda and Itoh 2004). Bam is needed for switching from self-renewal of GSCs into differentiation, and its cytoplasmic form is

a transiently expressed protein that begins to accumulate at the start of cytotblast differentiation and disappears after four rounds of mitosis (McKearin and Ohlstein 1995; Ohlstein and McKearin 1997; Szakmary et al. 2005). Gain of function screens in lines with defects at early stages of spermatogenesis show that repression of *bam* expression in the male GSC population is important for male GSC survival (Schulz et al. 2004). It is possible that the rapid evolution of the *bam* gene, and perhaps the detected signal of positive selection, is common among transiently expressed genes or even restricted to proteins degraded during mitosis. Comparison of the pattern of molecular evolution between orthologs in *Caenorhabditis elegans* and *Caenorhabditis briggsae* has shown that transiently expressed embryonic genes and modulated larval and adult genes show elevated rates of protein evolution (Cutter and Ward 2005). However, the sign of positive selection is indicative of interspecies adaptations and seems to be restricted to the branch leading to species of the simulans clade. This suggests that positive selection might relate to an unknown developmental adjustment unique to these species. One possibility is that proteins regulated through proteolysis during mitosis are under well-defined species-specific interactions and can be driven by positive selection.

Our study shows a pattern of high sequence divergence between closely related species for three genes with roles during different stages of sperm development in *Drosophila*. Rapid evolution is expected for sex-related genes but not consistent with views of selective constraints on early developmental genes and relaxed selection in more laterally expressed genes. In fact, our results show that the most downstream gene (*dj*) in the sperm developmental cascade is the only one showing signs of purifying selection, while an early expressed gene (*bam*) shows signs of adaptive divergence between species. The form of purifying selection operating on *dj* is similar to what has been reported for protamine, a DNA-binding sperm surface protein in mammals. Our results suggest that the selective mechanisms for rapid evolution of genes of reproduction are dictated by specific gene functional characteristics rather than site of expression. The combined analysis of

polymorphism and divergence allowed us to reach conclusions about selection that would have gone unnoticed in large interspecies gene comparisons, and our findings place constraints on broad conclusions about selection derived from the pooled analysis of genes on the basis of site of expression.

Supplementary Material

Appendices 1 and 2 are available at *Molecular Biology and Evolution* online (<http://www.mbe.oxfordjournals.org/>).

Acknowledgments

This work was supported by grants from the Natural Sciences and Engineering Research Council of Canada, the Canada Foundation for Innovation, and the Manitoba Health Research Council to A.C.

Literature Cited

- Andolfatto, P. 2001. Contrasting patterns of X-linked and autosomal nucleotide variation in *Drosophila melanogaster* and *Drosophila simulans*. *Mol. Biol. Evol.* **18**:279–290.
- Begun, D. J. 2002. Protein variation in *Drosophila simulans*, and comparison of genes from centromeric versus noncentromeric regions of chromosome 3. *Mol. Biol. Evol.* **19**:201–203.
- Begun, D. J., and P. Whitley. 2000. Reduced X-linked nucleotide polymorphism in *Drosophila simulans*. *Proc. Natl. Acad. Sci. USA* **97**:5960–5965.
- Civetta, A., and R. S. Singh. 1998. Sex-related genes, directional sexual selection, and speciation. *Mol. Biol. Evol.* **15**:901–909.
- Clark, A. G., and A. Civetta. 2000. Evolutionary biology. Protein wars. *Nature* **403**:261–263.
- Cutter, A. D., and S. Ward. 2005. Sexual and temporal dynamics of molecular evolution in *C. elegans* development. *Mol. Biol. Evol.* **22**:178–188.
- Fay, J. C., and C. I. Wu. 2000. Hitchhiking under positive Darwinian selection. *Genetics* **155**:1405–1413.
- Fu, Y. X., and W. H. Li. 1993. Statistical tests of neutrality of mutations. *Genetics* **133**:693–709.
- Fukuda, M., and T. Itoh. 2004. Slac2-a/melanophilin contains multiple PEST-like sequences that are highly sensitive to proteolysis. *J. Biol. Chem.* **279**:22314–22321.
- Fuller, M. T. 1993. Spermatogenesis. Pp. 71–147 in M. Bate and A. Martinez Arias, eds. *The development of Drosophila*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
- . 1998. Genetic control of cell proliferation and differentiation in *Drosophila* spermatogenesis. *Semin. Cell Dev. Biol.* **9**:433–444.
- Hudson, R. R. 2000. A new statistic for detecting genetic differentiation. *Genetics* **155**:2011–2014.
- Jiang, J., and H. White-Cooper. 2003. Transcriptional activation in *Drosophila* spermatogenesis involves the mutually dependent function of *aly* and a novel meiotic arrest gene *cookie monster*. *Development* **130**:563–573.
- Kulathinal, R., and R. S. Singh. 1998. Cytological characterization of premeiotic versus postmeiotic defects producing hybrid male sterility among sibling species of the *Drosophila melanogaster* complex. *Evolution* **52**:1067–1079.
- Kumar, S., K. Tamura, and M. Nei. 2004. MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief. Bioinform.* **5**:150–163.
- Lin, T. Y., S. Viswanathan, and C. Wood. 1996. Coordinate developmental control of the meiotic cell cycle and spermatid differentiation in *Drosophila* males. *Development* **122**:1331–1341.
- McDonald, J. H., and M. Kreitman. 1991. Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature* **351**:652–654.
- McKearin, D., and B. Ohlstein. 1995. A role for the *Drosophila* Bag-of-marbles protein in the differentiation of cystoblasts from germline stem cells. *Development* **121**:2937–2947.
- McKearin, D. M., and A. C. Spradling. 1990. *bag-of-marbles*: a *Drosophila* gene required to initiate both male and female gametogenesis. *Genes Dev.* **4**:2242–2251.
- Michalak, P., and M. A. Noor. 2003. Genome-wide patterns of expression in *Drosophila* pure species and hybrid males. *Mol. Biol. Evol.* **20**:1070–1076.
- Moriyama, E. N., and J. R. Powell. 1996. Intraspecific nuclear DNA variation in *Drosophila*. *Mol. Biol. Evol.* **13**:261–277.
- Nei, M. 1987. *Molecular evolutionary genetics*. Columbia University Press, New York.
- Nei, M., and T. Gojobori. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.* **3**:418–426.
- Nei, M., and J. Zhang. 1998. Molecular origin of species. *Science* **282**:1428–1429.
- Ohlstein, B., and M. McKearin. 1997. Ectopic expression of the *Drosophila* Bam protein eliminates oogenic germline stem cells. *Development* **124**:3651–3662.
- Orr, H. A., J. P. Masly, and D. C. Presgraves. 2004. Speciation genes. *Curr. Opin. Genet. Dev.* **14**:675–679.
- Orr, H. A., and D. C. Presgraves. 2000. Speciation by postzygotic isolation: forces, genes and molecules. *Bioessays* **22**:1085–1094.
- Presgraves, D. C., L. Balagopalan, S. M. Abmayr, and H. A. Orr. 2003. Adaptive evolution drives divergence of a hybrid inviability gene between two species of *Drosophila*. *Nature* **423**:715–719.
- Przeworski, M. 2002. The signature of positive selection at randomly chosen loci. *Genetics* **160**:1179–1189.
- Richards, S., Y. Liu, B. R. Bettencourt et al. (49 co-authors). 2005. Comparative genome sequencing of *Drosophila pseudoobscura*: chromosomal, gene, and cis-element evolution. *Genome Res* **15**:1–18.
- Richter, B., M. Long, R. C. Lewontin, and E. Nitasaka. 1997. Nucleotide variation and conservation at the *dpp* locus, a gene controlling early development in *Drosophila*. *Genetics* **145**:311–323.
- Rogers, S., R. Wells, and M. Rechsteiner. 1986. Amino acid sequences common to rapidly degraded proteins: the PEST hypothesis. *Science* **234**:364–368.
- Rooney, A. P., J. Zhang, and M. Nei. 2000. An unusual form of purifying selection in a sperm protein. *Mol. Biol. Evol.* **17**:278–83.
- Rozas, J., J. C. Sánchez-DelBarrio, X. Messeguer, and R. Rozas. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* **19**:2496–2497.
- Santel, A., N. Blumer, M. Kampfer, and R. Renkawitz-Pohl. 1998. Flagellar mitochondrial association of the male-specific Don Juan protein in *Drosophila* spermatozoa. *J. Cell Sci.* **111**:3299–3309.
- Santel, A., T. Winhauer, N. Blumer, and R. Renkawitz-Pohl. 1997. The *Drosophila* don juan (dj) gene encodes a novel sperm specific protein component characterized by an unusual domain of a repetitive amino acid motif. *Mech. Dev.* **64**:19–30.
- Schulz, C., A. A. Kiger, S. I. Tazuke, Y. M. Yamashita, L. C. Pantalena-Filho, D. L. Jones, C. G. Wood, and M. T. Fuller. 2004. A misexpression screen reveals effects of *bag-of-marbles*

- and TGF beta class signaling on the *Drosophila* male germline stem cell lineage. *Genetics* **167**:707–723.
- Szakmary, A., D. N. Cox, Z. Wang, and H. Lin. 2005. Regulatory relationship among piwi, pumilio, and bag-of-marbles in *Drosophila* germline stem cell self-renewal and differentiation. *Curr. Biol.* **15**:171–178.
- Tajima, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**:585–595.
- Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins. 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* **24**:4876–4882.
- Ting, C. T., S. C. Tsaur, M. L. Wu, and C. I. Wu. 1998. A rapidly evolving homeobox at the site of a hybrid sterility gene. *Science* **282**:1501–1504.
- Wakimoto, B. T., D. L. Lindsley, and C. Herrera. 2004. Toward a comprehensive genetic analysis of male fertility in *Drosophila melanogaster*. *Genetics* **167**:207–216.
- Watterson, G. A. 1975. On the number of segregating sites in genetical models without recombination. *Theor. Popul. Biol.* **7**:256–276.
- White-Cooper, H., M. A. Schafer, L. S. Alpey, and M. T. Fuller. 1998. Transcriptional and post-transcriptional control mechanisms coordinate the onset of spermatid differentiation with meiosis I in *Drosophila*. *Development* **125**:125–134.
- White-Cooper, H., D. Leroy, A. MacQueen, and M. T. Fuller. 2000. Transcription of meiotic cell cycle and terminal differentiation genes depends on a conserved chromatin associated protein, whose nuclear localisation is regulated. *Development* **127**:5463–5473.
- Wilkins, A. S. 2002. The evolution of developmental pathways. Sinauer Associates, Sunderland, Mass.
- Wong, W. S., Z. Yang, N. Goldman, and R. Nielsen. 2004. Accuracy and power of statistical methods for detecting adaptive evolution in protein coding sequences and for identifying positively selected sites. *Genetics* **168**:1041–1051.
- Yang, Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* **13**:555–556.
- Yang, Z., and R. Nielsen. 2002. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Mol. Biol. Evol.* **19**:908–917.
- Yang, Z., R. Nielsen, N. Goldman, and A. M. Pedersen. 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* **155**:431–449.
- Yang, Z., W. S. Wong, and R. Nielsen. 2005. Bayes empirical Bayes inference of amino acid sites under positive selection. *Mol. Biol. Evol.* **22**:1107–1118.

Marcy Uyenoyama, Associate Editor

Accepted December 12, 2005