

Adaptive Evolution of Genes Involved in the Regulation of Germline Stem Cells in *Drosophila melanogaster* and *D. simulans*

Heather A. Flores,^{1,2} Vanessa L. Bauer DuMont,¹ Aalya Fadoo, Diana Hubbard, Mohammed Hijji, Daniel A. Barbash, and Charles F. Aquadro³

Department of Molecular Biology and Genetics, Cornell University, Ithaca, New York 14853

ABSTRACT Population genetic and comparative analyses in diverse taxa have shown that numerous genes involved in reproduction are adaptively evolving. Two genes involved in germline stem cell regulation, *bag of marbles* (*bam*) and *benign gonial cell neoplasm* (*bgn*), have been shown previously to experience recurrent, adaptive evolution in both *Drosophila melanogaster* and *D. simulans*. Here we report a population genetic survey on eight additional genes involved in germline stem cell regulation in *D. melanogaster* and *D. simulans* that reveals all eight of these genes reject a neutral model of evolution in at least one test and one species after correction for multiple testing using a false-discovery rate of 0.05. These genes play diverse roles in the regulation of germline stem cells, suggesting that positive selection in response to several evolutionary pressures may be acting to drive the adaptive evolution of these genes.

KEYWORDS

germline stem cells
adaptive evolution
positive selection
oogenesis
spermatogenesis

Reproduction and fertility are among the most important traits for organismal fitness. Many models and theoretical studies have proposed that germline and fertility-related genes will be targeted for selection, and empirical evidence has documented rapid evolution and in many cases positive selection on numerous genes known or proposed to be involved in male fertility (Tsaur *et al.* 1998; Begun *et al.* 2000; Swanson *et al.* 2001b, 2004; Clark and Swanson 2005; Haerty *et al.* 2007), female reproductive tract function (Lawniczak and Begun 2007; Prokupek *et al.* 2008; Kelleher and Markow 2009), host defense against segregation distorters (Presgraves 2007; Phadnis and Orr 2009), and sperm-egg interactions (Swanson and Vacquier 1995; Swanson *et al.* 2001a; Aagaard *et al.* 2010). Most of these genes are expressed at the

latter stages of gametogenesis and often are associated with meiosis or interactions between gametes. However, Civetta *et al.* (2006) and Bauer DuMont *et al.* (2007) independently discovered that two genes expressed in the earliest stages of gametogenesis, specifically germline stem cell (GSC) regulation, also show evidence of adaptive evolution. One of these genes, *bag of marbles* (*bam*), is under intensely strong positive selection with an astonishing 59 nonsynonymous substitutions among 442 codons between two closely related fruit fly species, *Drosophila melanogaster* and *D. simulans* (Civetta *et al.* 2006; Bauer DuMont *et al.* 2007). A second gene, *benign gonial cell neoplasm* (*bgn*), which acts together with *bam* as a key “switch” to initiate GSC differentiation, is also evolving under positive selection in these two species (Bauer DuMont *et al.* 2007). These discoveries raise a fundamental question: what is the selective pressure(s) driving these adaptive changes at early gametogenesis loci?

There have been several genome-wide, next-generation sequencing surveys of variation in *D. melanogaster* and *D. simulans* that have reported departures from an equilibrium neutral model in directions consistent with natural selection for GSC-related gene ontology categories or at/near several GSC genes (Begun *et al.* 2007; Langley *et al.* 2012; Pool *et al.* 2012). It remains informative to examine specific genes, particularly using parallel assays on population data from both *D. melanogaster* and *D. simulans*. Here, we report high-quality Sanger resequencing from population samples of both species for eight genes involved in GSC regulation (*cyclin A*, *mei-P26*, *nanos*, *P-element induced wimpy testis* (aka *piwi*), *pumilio*, *stonewall*, *fs(1)Yb*, and *zero population growth*), test for evidence of selection using polymorphism-based

Copyright © 2015 Flores *et al.*

doi: 10.1534/g3.114.015875

Manuscript received November 24, 2014; accepted for publication February 5, 2015; published Early Online February 9, 2015.

This is an open-access article distributed under the terms of the Creative Commons Attribution Unported License (<http://creativecommons.org/licenses/by/3.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Supporting information is available online at <http://www.g3journal.org/lookup/suppl/doi:10.1534/g3.114.015875/-/DC1>

Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under accession nos. JX647382–JX647689.

¹These authors contributed equally to this work.

²Present address: School of Biological Sciences, Building 53, Room 117, Monash University, Clayton, VIC, 3800, Australia.

³Corresponding author: Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY 14853. E-mail: cfa1@cornell.edu

methods and reanalyze longer-term sequence evolution at these genes using phylogenetic analysis by maximum likelihood (PAML). These eight genes include those whose products genetically and/or physically interact with *bam* and/or *bgn* and are likely to have shared functions, and those that appear to have non-*bam/bgn*-related roles in GSC regulation. Figure 1 illustrates the roles of these loci within the female germline, wherein the functions and interactions of these genes are more thoroughly understood. We note that several of these genes function somewhat differently in the male germline (Fuller and Spradling 2007; Gilboa and Lehmann 2004; Gonczy *et al.* 1997; Insko *et al.* 2009; Kawase *et al.* 2004; Song *et al.* 2004).

GSCs produce the cells that will develop to form either eggs or sperm throughout a fly's life. GSCs are maintained in a microenvironment called the stem cell niche that is located in the proximal end of the *Drosophila* ovary or the apical end of the testis (Figure 1). *bam* acts, together with *bgn*, as a switch to allow for female GSC differentiation, and therefore its expression is repressed in the GSCs (McKearin and Ohlstein 1995; Lavoie *et al.* 1999; Ohlstein *et al.* 2000) by extrinsic signals from the stem cell niche (Song *et al.* 2004). However, this signaling quickly dissipates and thus repression only occurs in cells that are in physical contact with the stem cell niche (Wong *et al.* 2005; Xia *et al.* 2010). To receive these extrinsic signals, GSCs remain physically attached to the niche through adherens junctions (Song *et al.* 2002). The gap junction protein Zero population growth (*Zpg*) is present in the cytoplasmic membrane of both GSCs and niche cells

and is required for the maintenance of GSCs through the sharing of small molecules and signals between the niche and GSC (Tazuke *et al.* 2002; Gilboa *et al.* 2003). Repression of *bam* expression in the GSC is also controlled by the genes *female-sterile(1)Yb* (also abbreviated as *Yb*) and *P-element induced wimpy testis (piwi)* (King *et al.* 2001; Szakmary *et al.* 2005).

Intrinsic mechanisms within the GSC play an important role in its maintenance as well, at the levels of transcription and translation. The chromatin-associated protein Stonewall (*Stwl*) represses genes that promote differentiation (Maines *et al.* 2007), whereas *Mei-P26* antagonizes the miRNA pathway and represses transcripts that will promote differentiation in the cystoblast (Neumuller *et al.* 2008; Li *et al.* 2012). At the translational level, *Nanos* (*Nos*) and *Pumilio* (*Pum*) bind to mRNAs that promote differentiation and inhibit their translation (Lin and Spradling 1997; Wang and Lin 2004). *zpg* is also required to promote cystoblast differentiation (Tazuke *et al.* 2002; Gilboa *et al.* 2003). So depending on the context, *zpg* and *mei-P26* both inhibit and promote GSC differentiation. Finally, the cystoblast will undergo four mitotic divisions. *bam* is thought to regulate the number of mitotic divisions, and genetic interaction assays have suggested that *bam* interacts with the cell cycle factor, *cyclin A (cycA)* in this process (Lilly *et al.* 2000).

We report here that all eight genes show a statistically significant departure from an equilibrium neutral model for at least one polymorphism-based statistical test. Additionally, *Yb* and *stwl* also

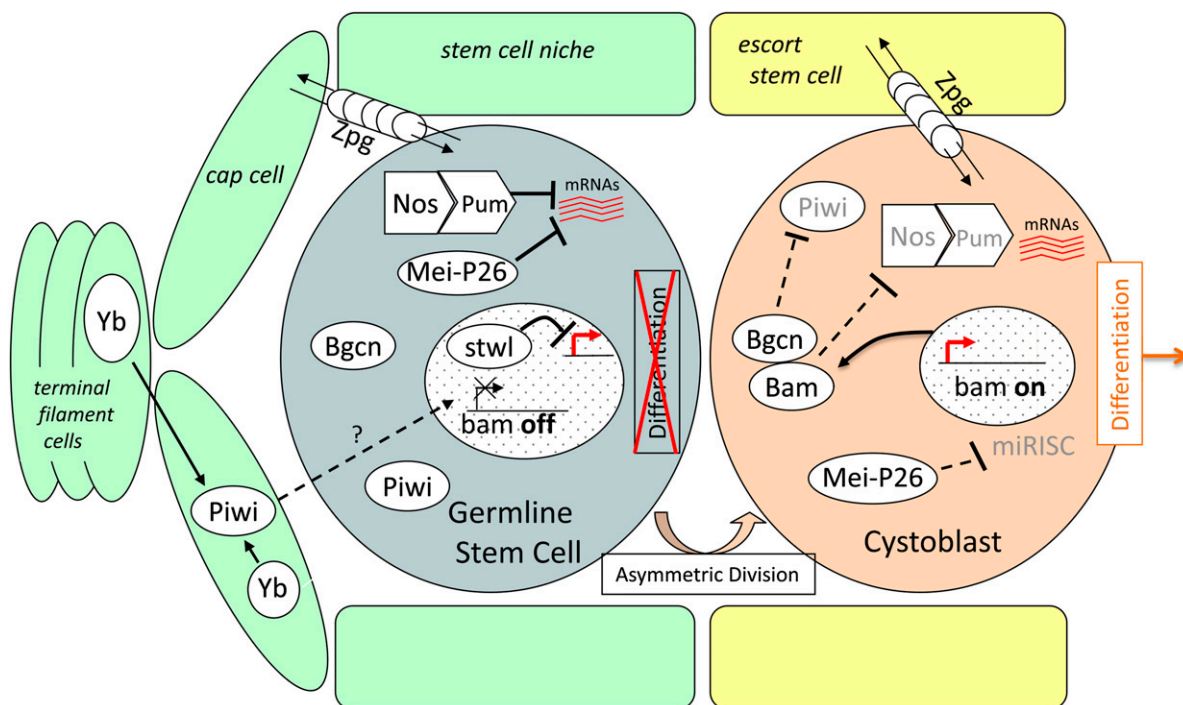


Figure 1 Schematic of the *Drosophila* ovarian germline stem cell (GSC) niche with genes analyzed. Adapted from Wong *et al.* (2005). The GSC (light blue cell) is present in a niche environment (green cells are somatic cap and terminal filament cells, yellow cells are escort stem cells) required to maintain its stem cell state. *Bam* is repressed in the GSC. Only when the GSC moves away from the niche is *Bam* expressed and this cell starts to differentiate (tan cell). *Yb* is involved in the maintenance of GSCs and regulating their division. *Piwi* acts cell nonautonomously to help in the repression of *Bam* in the GSC. *Zpg* is an adherens junction protein that functions in cell signaling. *Nos* and *Pum* act as translational repressors of genes that will promote differentiation. *Mei-p26* acts in concert with the miRNA machinery (miRISC in the figure) to also repress transcripts (indicated by red squiggly lines), some of which are shared with *Nos* and *Pum*. *Bgcn* is required for *Bam* to cause GSCs to differentiate. *Bam* and *Bgcn* antagonize the *Nos/Pum* complex. *Stwl* represses *Bam*-independent differentiation pathways and thus maintains GSC self-renewal. The cystoblast (tan cell) will undergo four mitotic divisions. *CycA* participates in the regulation of these mitotic divisions but is not shown in this diagram.

reject neutrality by the McDonald-Kreitman (MK) test, suggesting an excess of nonsynonymous fixations between species consistent with positive selection. These eight genes together with *bam* and *bgn* have various molecular functions and are expressed in a range of cell types including GSCs, cysts, and surrounding somatic cells suggesting that multiple evolutionary forces are acting throughout the early germline to drive the adaptive evolution of these genes.

MATERIALS AND METHODS

Fly stocks

When possible, African populations of *Drosophila melanogaster* and *D. simulans* were used to minimize the effects of demography in our ability to detect selection (Begun and Aquadro 1993). In some cases, different populations were used for different genes because of the availability of stocks with extracted chromosomes, which allowed us to sequence homozygous lines in *D. melanogaster* for the X, second, or third chromosomes. For *D. simulans* populations, inbred lines were used. For *stwl*, *zpg*, *piwi*, and *pum* a *D. melanogaster* population from Uganda, Africa (Pool and Aquadro 2006) and a *D. simulans* population from Lake Kariba, Zimbabwe, Africa (Pool and Aquadro 2006) were used. For *Yb* and *mei-P26*, a *D. melanogaster* population collected from Sengwa Wildlife Research Institute in Zimbabwe, Africa (Begun and Aquadro 1994) and a *D. simulans* population from Lake Kariba, Zimbabwe (Pool and Aquadro 2006) were used. For *cyclin A* and *nanos*, a *D. melanogaster* population sample collected from Lake Kariba, Zimbabwe, Africa (Pool and Aquadro 2006) and an inbred *D. simulans* population sample from North Carolina (Aquadro *et al.* 1988) were used.

Sequencing

Genomic DNA was extracted from approximately 20 adult flies using Puregene Core Kit A DNA isolation kits (QIAGEN). Polymerase chain reaction and sequencing primer sequences for each gene are listed in Supporting Information, Table S1. Sanger sequencing (both strands) was performed by the Cornell University Genomics Core DNA Sequencing Facility (<http://cores.lifesciences.cornell.edu/brcinfo/?f=1>) using ABI chemistry and 3730XL DNA Analyzers. Sequences were assembled and edited using Sequencher 4.9 (Gene Codes) and aligned using MEGA 4 (Tamura *et al.* 2007) using default settings, and checked manually to assure the reading frame was retained. Sequences have been deposited in GenBank under accession numbers JX647382–JX647689. For *piwi*, a single 4.8-kb sequence that includes all exons was amplified. This large fragment was problematic for direct sequencing, so it was cloned into the pCR-BluntII-TOPO plasmid (Invitrogen). Two clones of each sample were sequenced to control for PCR errors. If there was ambiguity between the two clones, a third was sequenced and the majority nucleotide was used. The *pum* locus spans over 160 kb, so four separate products were sequenced that include most of the exons (Figure S1A). The *stwl* locus was amplified in two separate products that included both exons (Figure S1B). The *cycA* locus also amplified in two separate products that include two groups of exons in the 5' and 3' region of the gene (Figure S1C). For *mei-P26*, only exons 3–6 were amplified. Our results based on this region are consistent with other reports that *mei-P26* has not been subject to recurrent, positive selection (Anderson *et al.* 2009).

Polymorphism analysis

DnaSP (Librado and Rozas 2009) was used to estimate basic summary statistics of variation within and between species. To detect signatures of recent selection from polymorphism data we applied two quite

different tests: OmegaPlus (Pavlidis *et al.* 2010a), which focuses on the linkage disequilibrium signature of selective sweeps, and SweeD (Pavlidis *et al.* 2013), which assesses the fit of the site frequency spectrum to a particular neutral model (it is a faster extension of the widely used SweepFinder method; Nielsen *et al.* 2005).

Statistical significance of OmegaPlus (dependent on linkage disequilibrium) and SweeD (dependent on SFS) test results was determined using neutral simulations with or without demography. We considered a region to be a significant outlier if it fell within the 5% quantile of the simulated datasets. These simulations were done using the program msABC (Pavlidis *et al.* 2010b). We surveyed variation from an African population of *D. melanogaster* which is within this species' presumed ancestral range. There is mounting evidence that even African populations of this species have experienced changes in effective population size over time (Glinka *et al.* 2003; Haddrill *et al.* 2008; Hutter *et al.* 2007; Li and Stephan. 2006; Duchon *et al.* 2013; Singh *et al.* 2013) and/or migration (Pool *et al.* 2012). Because inferring demographic parameters is challenging, we simulated three different scenarios: standard neutral model with constant population size, standard neutral model with exponential growth as estimated by Hutter *et al.* (2007), or standard neutral model with a 3-phase ("3 epoch") bottleneck as estimated by Duchon *et al.* (2013). We supplied msABC with uniform prior distributions for theta and all demographic parameters. The theta prior distribution for *D. melanogaster* was obtained from Pool *et al.* (2012) and ranged between 0.006 and 0.009 per site. Figure S2 shows the basic model of the demographic scenarios we considered and the demographic priors used in the simulations. To date, there are no comparable estimates of demographic parameters available for *D. simulans*. Given that the ancestral range of both of these species is in Africa and they are sympatric, we used the *D. melanogaster* demographic parameters as an approximation for *D. simulans*. For *D. simulans*, we used the theta range we observed across the eight GSC loci in this study, which ranged between 0.003 and 0.04 per site.

The MK test (McDonald and Kreitman 1991) was used to test for recurrent, historical positive selection by contrasting pooled polymorphism for *D. melanogaster* and *D. simulans* to fixed differences between species using *D. yakuba* as an outgroup. We used the program DoFE (<http://www.sussex.ac.uk/lifesci/eyre-walkerlab/resources>) from Eyre-Walker and Keightley (2009) to calculate the proportion of amino acid fixations predicted to be due to positive selection (α). This method uses the site frequency spectrum to jointly estimate the selective effects of new deleterious mutations and the number of adaptive substitution for a selected class of mutations while also incorporating a generalized model of effective population size. For our analysis, we used the site frequency spectrum of fourfold (neutral class) and 0-fold (selected class) codon positions for both *D. melanogaster* and *D. simulans*. The sample size for each locus in our analysis varied for each species. We randomly selected nine and six alleles in *D. melanogaster* and *D. simulans*, respectively. These values correspond to the smallest sample size in each species.

Divergence analysis

The *Drosophila* 12 Genomes Consortium (2007) has previously reported tests of long-term recurrent positive selection using PAML (Yang 1997, 2007) for *nos*, *pum*, *zpg*, *cycA*, and *mei-P26* and found none departed from a neutral model. Three genes (*stwl*, *piwi*, and *Yb*), had not been included in this previous study due to their strict criteria that ruled out genes with alignment ambiguities. We generated new multiple-sequence alignments using PRANK alignment software (Löytynoja and Goldman 2005) from single sequences of

D. melanogaster, *D. simulans*, *D. sechellia*, *D. yakuba*, *D. erecta*, and *D. ananassae* downloaded from FlyBase. We did not use more divergent species due to the problems of saturation of synonymous site divergence (The *Drosophila* 12 Genomes Consortium *et al.* 2007). *Yb* from *D. ananassae* has a large number of indels relative to the other five species (and has a much larger coding sequence and an additional intron). Therefore, we analyzed these *Yb* alignments with PAML in two ways: 1) excluding any region with an indel, and 2) excluding any region with an indel as well as with one codon on either side (to reduce the chance calling of “false” substitutions associated with alignment problems). For *Yb*, we also used the recently published improved reference genome sequence of *D. simulans* from Hu *et al.* (2013). The models M0 vs. M3, M7 vs. M8, and M8 vs. M8a were compared. Consistent with the analyses from the *Drosophila* 12 Genomes Consortium (2007), each run was performed using three tree topologies: Tree 1, *D. yakuba* and *D. erecta* as sister species; Tree 2, *D. yakuba* as an outgroup and Tree 3, *D. erecta* as an outgroup. Each model comparison was run under three different initial ω values to assure that convergence was to a global and not local maximum.

Adjusting for multiple testing: We adjusted our criteria for statistical significance by estimating the appropriate *P*-value threshold assuming an experiment-wide Benjamini and Hochberg (1995) false-discovery rate (FDR) of 0.05 using the *p.adjust* function in the R Project (www.r-project.org). The *P*-values of SweeD and MK tests were combined for correction for each species separately as both tests use the frequency or counts of each polymorphism. OmegaPlus only uses patterns of linkage disequilibrium across sites, and thus those *P*-values were corrected separately (again for each species alone).

RESULTS

Polymorphism-based analyses

Gene function and sample size data from African populations of *D. melanogaster* and either African or North American *D. simulans* are reported in Table 1, and standard summary statistics for each gene

in Table 2. We find that *D. simulans* levels of nucleotide variability are generally higher than those seen in *D. melanogaster*, consistent with previous results (Aquadro *et al.* 1988).

Analysis of the polymorphism site frequency data using SweeD reveals significant departures from neutrality at 15 of 16 gene/species comparisons after multiple-testing correction (Table 3). For this tabulation, we consider a gene to be showing a significant departure from neutrality if at least one of the gene regions analyzed shows a significant departure (after multiple test correction) for all three demographic scenarios (standard neutral, exponential growth, and 3-epoch bottleneck). Only *piwi* in *D. simulans* fits a neutral model under all three demographic scenarios.

OmegaPlus rejected the standard neutral model only for *Yb* in *D. melanogaster* after multiple test correction at the 0.05 FDR level (Table 3). The generally short size of the regions analyzed may have limited the statistical power of the OmegaPlus method, which relies on a unique structure of linkage disequilibrium generated by recent selective sweeps.

Polymorphism and divergence-based tests

The McDonald-Kreitman (MK) test rejected neutrality for both *Yb* and *stwl* after correction for multiple testing (Table 4). The method of Bauer DuMont *et al.* (2004) suggests that these MK test rejections are not due to selection on synonymous sites for either gene. High d_N/d_S ratios between species (0.627 for *Yb*, and 0.502 for *stwl*) compared with the genome-wide average of 0.0125 (Larracunte *et al.* 2008), yet normal levels of d_S for both genes (0.132 and 0.119, respectively), suggests that the MK test rejections are due to excesses of fixed non-synonymous differences between species consistent with positive selection.

Using the DoFE program of Eyre-Walker and Keightley (2009), we estimated, in both *D. melanogaster* and *D. simulans*, the overall proportion of amino acid substitutions fixed due to positive selection (α), and the 95% credibility interval around this estimate (supplemental method presented in Eyre-Walker and Keightley 2009). This analysis uses the site frequency spectrum across the eight GSC loci to estimate

■ Table 1 Genes surveyed and sample sizes

Gene	Function	Number of Alleles Sampled	
		<i>D. melanogaster</i>	<i>D. simulans</i>
<i>cyc A</i> ^a	Regulation of cyst mitotic divisions		
Segment 1		9	10
Segment 2		9	10
<i>Yb</i>	GSC maintenance and cytotblast differentiation	19	9
<i>mei-P26</i> ^a	GSC maintenance	19	10
<i>nos</i> ^a	GSC maintenance	9	10
<i>pum</i> ^a	GSC maintenance		
Segment 1		17	9
Segment 2		11	10
Segment 3		19	9
Segment 4		18	7
<i>piwi</i>	GSC maintenance	10	6
<i>stwl</i>	Chromatin factor, GSC maintenance		
Segment 1		18	8
Segment 2		15	9
<i>zpg</i> ^a	GSC adherens junction and cytotblast differentiation	18	10

^a Indicates that gene has a genetic and/or physical interaction reported with *bam*. For *pumilio*, four separate regions were amplified and analyzed, labeled as 1-4. For *stonewall* and *cycA* two separate regions were amplified, labeled as 1 and 2. GSC, germline stem cell.

■ **Table 2 Nucleotide polymorphism estimates for GSC genes**

Gene	Species	S	θ	π_{Tot}	π_{Syn}	π_{Non}
cycA 1	<i>D. melanogaster</i>	14	0.0051	0.0046	0.0025	0.0054
	<i>D. simulans</i>	10	0.0035	0.0044	0.0277	0.0000
cycA 2	<i>D. melanogaster</i>	15	0.0085	0.0074	0.0157	0.0010
	<i>D. simulans</i>	11	0.0056	0.0061	0.0164	0.0000
mei-P26	<i>D. melanogaster</i>	26	0.0062	0.0061	0.0181	0.0000
	<i>D. simulans</i>	21	0.0055	0.0031	0.0110	0.0000
nos	<i>D. melanogaster</i>	21	0.0044	0.0045	0.0090	0.0009
	<i>D. simulans</i>	35	0.0097	0.0093	0.0150	0.0042
piwi	<i>D. melanogaster</i>	103	0.0079	0.0074	0.0196	0.0024
	<i>D. simulans</i>	196	0.0222	0.0204	0.0368	0.0025
pumilio 1	<i>D. melanogaster</i>	26	0.0040	0.0046	0.0033	0.0012
	<i>D. simulans</i>	103	0.0202	0.0169	0.0142	0.0003
pumilio 2	<i>D. melanogaster</i>	10	0.0052	0.0040	0.0072	0.0000
	<i>D. simulans</i>	33	0.0142	0.0144	0.0346	0.0005
pumilio 3	<i>D. melanogaster</i>	10	0.0040	0.0046	0.0172	0.0020
	<i>D. simulans</i>	72	0.0400	0.0388	0.0685	0.0014
pumilio 4	<i>D. melanogaster</i>	13	0.0021	0.0020	0.0070	0.000
	<i>D. simulans</i>	50	0.0095	0.0089	0.0207	0.0011
stwl 1	<i>D. melanogaster</i>	21	0.0092	0.0058	0.0000	0.0001
	<i>D. simulans</i>	17	0.0088	0.0064	0.0119	0.0000
stwl 2	<i>D. melanogaster</i>	49	0.0051	0.0048	0.0123	0.0025
	<i>D. simulans</i>	43	0.0053	0.0050	0.0097	0.0033
Yb	<i>D. melanogaster</i>	88	0.0079	0.0060	0.0129	0.0028
	<i>D. simulans</i>	111	0.0128	0.0128	0.0259	0.0085
zpg	<i>D. melanogaster</i>	41	0.0095	0.0113	0.0413	0.0001
	<i>D. simulans</i>	60	0.0164	0.0148	0.0286	0.0015

Each amplified region of *cycA*, *pumilio*, and *stwl* was analyzed separately; see the section *Materials and Methods* and Figure S1 for locations of each amplicon. S, segregating sites; θ , nucleotide diversity; π_{Tot} , total diversity; π_{Syn} , synonymous diversity; π_{Non} , nonsynonymous diversity. GSC, germline stem cell.

the distribution of fitness effects acting on new deleterious mutations, while incorporating a general model of effective population size change. The distribution of fitness effects is then used to determine the proportion of amino acid fixations that are due to positive selection. For the eight loci in our study, we estimate α to be 0.814 (95% credibility interval: 0.698–0.896) and 0.790 (95% credibility interval: 0.681–0.881) for *D. melanogaster* and *D. simulans*, respectively. We also analyzed the X and autosomal loci separately. For *D. melanogaster* we observe a α of 0.934 (95% credibility interval: 0.852–0.979) for the X chromosome and 0.672 (95% credibility interval: 0.413–0.836) for the autosomes. For *D. simulans* we observe a α of 0.856 (95% credibility interval: 0.695–0.957) for the X chromosome and 0.743 (95% credibility interval: 0.579–0.876) for the autosomes. The autosomal 95% credibility interval estimated for α from our *D. melanogaster* data encompasses the α estimate obtained from sequence data from 419 autosomal loci chosen randomly (0.52; Keightley and Eyre-Walker 2012). To date, this method to estimate α has not been applied to another *D. simulans* dataset. However, α has been calculated by other methods for *D. simulans* and estimates have ranged from 0.43 to 0.94 (reviewed in Eyre-Walker 2006), which is similar to the estimates we present here.

Divergence-based analyses

No evidence of recurrent, adaptive evolution at the same subset of codons across *D. melanogaster*, *D. simulans*, *D. sechellia*, *D. yakuba*, *D. erecta*, and *D. ananassae* was detected using PAML (Yang 1997, 2007) for seven of the eight genes *cycA*, *mei-P26*, *nos*, *piwi*, *pum*, *stwl*, *zpg* (our analyses and those presented in The *Drosophila* 12 Genomes Consortium 2007). However, we do find evidence of recurrent, positive selection at specific codons for *Yb*. Using both models M7 vs. M8,

and M8 vs. M8a, we find that the data fit a model of selection significantly better than a neutral null model (likelihood ratio test statistics of 16.068 with $P < 0.0003$, and 6.321 with $P < 0.01$, respectively). This result is robust to alignment with this highly diverged protein, including reanalysis removing all codons adjacent to predicted INDELS. 19 of the aligned codons at *Yb* are predicted by Bayes Empirical Bayes analysis to be in the selective class with an average codon-specific d_N/d_S ($= \omega$) of 1.88. However, only two codons in this class have predicted posterior probabilities greater than 0.90, and they do not fall in areas of known domains.

DISCUSSION

Previous genome-wide next-generation sequencing studies using both site frequency-based and MK tests of neutrality have reported an enrichment of putative adaptive evolution in Gene Ontology categories such as germ-cell development, cystoblast division, and germline-derived oocyte fate determination (Begun *et al.* 2007; Langley *et al.* 2012; Mackay *et al.* 2012). In this study, we performed high-quality Sanger sequencing of population samples from both *D. melanogaster* and *D. simulans* and found that all eight genes involved in GSC regulation studied here reject a neutral model of evolution in at least one test and species (Tables 3 and 4). Most of these rejections are due to the polymorphism-based SweeD analysis for which every locus, except *piwi*, rejects the neutral model in both *D. melanogaster* and *D. simulans*. The *piwi* locus only rejects neutrality by the SweeD test in *D. melanogaster*. Rejecting the neutral model with SweeD is suggestive of positive selection, but it could also be due to demographic history (Pavlidis *et al.* 2010a). We attempted to take the demographic history of these species into account by using simulated replicates of estimates of *D. melanogaster* African population dynamics (Hutter *et al.* 2007;

■ Table 3 Site frequency tests of departures from neutral models for eight GSC genes in *D. melanogaster* and *D. simulans*

Test Details	SweeD Test of Recent Selection				OmegaPlus Test of Recent Selection			
	<i>D. melanogaster</i>		<i>D. simulans</i>		<i>D. melanogaster</i>		<i>D. simulans</i>	
	Orig P-value	FDR adj P-Value	Orig P-Value	FDR adj P-Value	Orig P-Value	FDR adj P-Value	Orig P-Value	FDR adj P-Value
CycA1.SN	0.0007	0.0023	0.0015	0.0032	0.367	0.4037	0.4756	0.4756
CycA1.Ex	0.0007	0.0023	0.0009	0.0023	0.4323	0.4390	0.4049	0.4711
CycA1.3Ep	0.0007	0.0023	0.0012	0.0027	0.439	0.4390	0.4094	0.4711
CycA2.SN	0.6203	0.7030	0.9933	0.9933	0.3726	0.4037	0.2594	0.4214
CycA2.Ex	0.3460	0.4334	0.7363	0.8024	0.2809	0.3894	0.2622	0.4214
CycA2.3Ep	0.3467	0.4334	0.7863	0.8460	0.2866	0.3894	0.2595	0.4214
meiP26.SN	0.0010	0.0023	0.0009	0.0023	0.1894	0.3894	0.0333	0.4214
meiP26.Ex	0.0008	0.0023	0.0010	0.0023	0.2081	0.3894	0.0999	0.4214
meiP26.3Ep	0.0008	0.0023	0.0006	0.0023	0.2108	0.3894	0.0648	0.4214
nano.SN	0.0090	0.0150	0.0008	0.0023	0.4055	0.4274	0.2278	0.4214
nano.Ex	0.0020	0.0040	0.0010	0.0023	0.2560	0.3894	0.2329	0.4214
nano.3Ep	0.0050	0.0092	0.0009	0.0023	0.3403	0.4037	0.2204	0.4214
piwi.SN	0.0009	0.0023	0.9680	0.9795	0.1267	0.3894	0.1899	0.4214
piwi.Ex	0.0070	0.0124	0.9370	0.9713	0.1096	0.3894	0.3899	0.4711
piwi.3Ep	0.0010	0.0023	0.9500	0.9729	0.1030	0.3894	0.3909	0.4711
pum1.SN	0.0008	0.0023	0.4174	0.4997	0.2302	0.3894	0.2198	0.4214
pum1.Ex	0.0009	0.0023	0.1640	0.2213	0.2537	0.3894	0.2126	0.4214
pum1.3Ep	0.0009	0.0023	0.1736	0.2306	0.2532	0.3894	0.2293	0.4214
pum2.SN	0.0008	0.0023	0.0919	0.1281	0.3631	0.4037	0.2649	0.4214
pum2.Ex	0.0009	0.0023	0.0278	0.0407	0.3630	0.4037	0.2599	0.4214
pum2.3Ep	0.0008	0.0023	0.0227	0.0339	0.3389	0.4037	0.2605	0.4214
pum3.SN	0.2073	0.2711	0.0153	0.0245	0.1728	0.3894	0.1664	0.4214
pum3.Ex	0.1633	0.2213	0.0010	0.0023	0.0612	0.3894	0.2701	0.4214
pum3.3Ep	0.0800	0.1133	0.0019	0.0039	0.0800	0.3894	0.2131	0.4214
pum4.SN	0.0186	0.0287	0.8249	0.8765	0.2995	0.3894	0.4296	0.4711
pum4.Ex	0.0067	0.0121	0.7174	0.7919	0.2694	0.3894	0.3909	0.4711
pum4.3Ep	0.0080	0.0139	0.6710	0.7505	0.2690	0.3894	0.3414	0.4711
stwlReg1.SN	0.0159	0.0250	0.0219	0.0332	0.0885	0.3894	0.0444	0.4214
stwlReg1.Ex	0.0009	0.0023	0.0021	0.0042	0.0630	0.3894	0.1062	0.4214
stwlReg1.3Ep	0.0043	0.0081	0.0010	0.0023	0.0516	0.3894	0.0827	0.4214
stwlReg2.SN	0.0010	0.0023	0.8780	0.9214	0.2987	0.3894	0.4099	0.4711
stwlReg2.Ex	0.0006	0.0023	0.4943	0.5756	0.2851	0.3894	0.4637	0.4756
stwlReg2.3Ep	0.0007	0.0023	0.4688	0.5534	0.2869	0.3894	0.4349	0.4711
yb.SN	0.0009	0.0023	0.0014	0.0031	0.0009	0.0351	0.4469	0.4711
yb.Ex	0.0006	0.0023	0.0040	0.0077	0.0059	0.0767	0.4212	0.4711
yb.3Ep	0.0008	0.0023	0.0007	0.0023	0.0033	0.0644	0.3702	0.4711
zpg.SN	0.0008	0.0023	0.0009	0.0023	0.2129	0.3894	0.2093	0.4214
zpg.Ex	0.0090	0.0150	0.0006	0.0023	0.2991	0.3894	0.1461	0.4214
zpg.3Ep	0.0010	0.0023	0.0009	0.0023	0.2807	0.3894	0.1692	0.4214

Simulations to establish *P*-values were from the standard neutral model (SN), exponential growth model (Ex), or a 3-epoch model (3Ep; large, small, large population size) as described in the section *Materials and Methods*. FDR-adjusted *P*-values were determined as described in text. Significant results ($P < 0.05$) are in bold. FDR, false-discovery rate.

Duchen *et al.* 2013) to determine our significant SweeD cutoff points. However, the true demographic history of these species is unknown. So, we stress that our SweeD rejections are restricted to the demographic scenarios we considered.

The detection of outliers of a test statistic's genomic distribution is another method used to determine statistical significance. Recently Pool *et al.* (2012) applied SweeD (labeled SweepFinder in their manuscript) genome-wide for an African population of *D. melanogaster* and they list regions containing genomic outliers, assumed to be due to positive selection. As an attempt to determine if our SweeD rejections are more likely due to demography *vs.* selection, we checked to see if the eight GSC loci we analyzed fell within or near the Pool *et al.* (2012) outliers. The protein coding regions (CDS) for three GSC loci (*Yb*, *piwi* and *mei-P26*) are within an outlier region, suggesting that for these loci our SweeD rejections are due to positive selection. The CDS

for two other GSC loci (*zpg* and *nano*) are within 50 kb of an outlier region. Simulations have shown that SweeD's ability to pinpoint the target of selection is compromised if both selection and demographic perturbations have occurred (Pavlidis *et al.* 2010a) with the predicted target being tens of kilobases away from the actual location of selection. To determine if by chance one would expect to observe three of eight loci within an outlier regions, or five of eight loci 50 kb from an outlier region, we randomly picked eight loci from the *D. melanogaster* genome. The loci were picked such that we obtained a random sample with the same distribution across the X, 2nd, or 3rd chromosomes as observed across the GSC loci. For both cases our observation is significant with only 36 of 1000 bootstrapped samples having 3 and greater or 5 and greater loci within or 50 Kb from an outlier region, respectively (thus *P*-value = 0.036 for our observation). Therefore, for 5 of the 8 GSC loci we analyzed, two different datasets (using two different

■ **Table 4** MK tests of departures from a neutral model for eight GSC genes using polymorphism within both *D. melanogaster* and *D. simulans* and fixed differences between species

Gene	Synon Poly	Synon Div	Nonsyn Poly	Nonsyn Div	P-Value	FDR adj P-Value
<i>cyc A</i>	17	14	4	6	0.414	0.4997
<i>mei-P26</i>	24	44	0	0	NA	NA
<i>nos</i>	11	11	7	23	0.046	0.0663
<i>piwi</i>	84	54	26	22	0.416	0.4997
<i>pum</i>	77	51	11	5	0.506	0.5812
<i>stwl</i>	45	62	48	124	0.015	0.0245
<i>Yb</i>	86	62	80	149	0.00001	0.0010
<i>zpg</i>	53	14	6	4	0.230	0.2962

FDR adjusted *P*-values were determined as described in text. Significant results after FDR ($P < 0.05$) are in bold. MK, McDonald-Kreitman; GSC, germline stem cell; FDR, false-discovery rate.

methods for determining the significant cutoffs) suggest that their frequency spectra do not match the neutral model in *D. melanogaster*. For *D. simulans*, making a distinction between demography and selection is more tenuous, especially given that there are no comparable estimates of the demographic history within Africa for this species.

Yb is the only gene to show significant departures from neutrality consistent with natural selection for the site frequency test SweeD as well as for both the MK and PAML tests that can detect recurrent historical selection. This combination of test results suggests that the recent sweeps at *Yb* detected by SweeD are just the latest of many selective fixations of nonsynonymous substitutions that have occurred among these six species.

Using the method of Eyre-Walker and Keightley (2009), we estimate that 81% of the amino acid differences fixed in these eight genes in the *D. melanogaster* lineage and 79% of the amino acid differences fixed in *D. simulans* lineage have been driven by positive selection. Estimates for X-linked genes were slightly, although not significantly, larger than those for autosomes. This proportion is on the upper end of that estimated for other groups of genes in these species.

The pattern of evidence for recent or recurrent positive selection that we observe and the diverse functions and expression patterns of these genes suggest that there are likely multiple selective pressures driving the adaptive evolution in genes important in GSC regulation. For example, three of the eight genes examined that reject the neutral model have no known interaction or dependence on *bam* function (*stwl*, *piwi*, and *Yb*; Chen and McKearin 2005; Li *et al.* 2009). *Yb* is expressed in the stem cell niche (King and Lin 1999; King *et al.* 2001), whereas *stwl* binds chromatin (Clark and McKearin 1996; Maines *et al.* 2007), making it less likely that the same specific selective pressures act on both.

The hypothesis of sexual selection and sexual conflict (Swanson and Vacquier 2002) cannot be formally rejected but seem implausible for genes functioning in GSCs. For example, most theories of sexual selection predict strong effects on premating traits, which are highly unlikely to be influenced by the genes we have examined. Likewise, sexual conflict, whereby one sex manipulates the reproductive fitness of the other sex, is much more likely to occur for molecules that are transmitted between males and females, a function that is implausible for any of the GSC regulatory genes in this study.

Several other mechanistic and evolutionary hypotheses have been proposed to explain the evolutionary causes of positive selection inferred for *bam* and *bgcn*. Some of these selective pressures also may drive the adaptive evolution of other genes involved in GSC regulation. Civetta *et al.* (2006) proposed that species-specific changes in rates of proteolysis could drive protein sequence divergence. This proposal was supported by the observation that *bam*'s expression is transient and by previous studies in *C. elegans* that have shown that

transiently expressed genes have elevated rates of protein evolution (Cutter and Ward 2005). Although this could influence the molecular evolution of *bam*, and potentially *bgcn* which is also transiently expressed (Ohlstein *et al.* 2000), it is unlikely to explain all selection acting on GSC gene evolution since *piwi*, *Yb*, *stwl*, and *nanos* have much broader patterns and timings of expression (Clark and McKearin 1996; Forbes and Lehmann 1998; Cox *et al.* 2000; Szakmary *et al.* 2009).

We had previously hypothesized that coevolution with external pathogens infecting the germline could underlie the elevated nonsynonymous divergence in *bam* and *bgcn* along the *D. melanogaster* and *D. simulans* lineages (Bauer DuMont *et al.* 2007). Two maternally-inherited bacterial endosymbionts (*Wolbachia* and *Spiroplasma*) have been detected in some but not all species of *Drosophila* (Mateos *et al.* 2006; Watts *et al.* 2009). Infection by *Wolbachia* can have beneficial effects in some species by increasing resistance to viral infections, which may explain their widespread presence (Chrostek *et al.* 2013; Hedges *et al.* 2008; Teixeira *et al.* 2008). However, *Wolbachia* infection can also reduce fecundity due to cytoplasmic incompatibilities in crosses between infected and uninfected individuals (Fry *et al.* 2004). Overreplication of *Wolbachia* also has been linked to shortening life-span and rupture of host cells (Min and Benzer 1997). There is likely to be a delicate balance in controlling endosymbiont proliferation within a cell so that the host can receive benefits from the endosymbiont but minimize any deleterious effects (Chrostek *et al.* 2013). Maintaining such a balance could contribute to an “arms race” between GSC regulatory genes and endosymbionts (*e.g.*, Werren 2005; Bauer DuMont *et al.* 2007).

The expression patterns and known pleiotropic functions of *Yb*, *piwi*, *stwl* (Aravin *et al.* 2007; Brennecke *et al.* 2008; Clark and McKearin. 1996; Maines *et al.* 2007) suggest that other pressures may be acting on them. One possible selective pressure is intracellular parasites such as transposons. Transposons are selfish genetic elements that can propagate throughout the genome, resulting in deleterious effects on their host. Recent studies demonstrated that many taxa, including *Drosophila*, have a small RNA silencing pathway, termed the piRNA pathway, that is active in the germline and provides an adaptive defense against transposons (Aravin *et al.* 2007). Many piRNA pathway genes also have been shown to adaptively evolve (Obbard *et al.* 2009; Kolaczowski *et al.* 2011). *piwi* and *Yb* are required for the proper silencing of transposons (Aravin *et al.* 2007; Olivieri *et al.* 2010; Saito *et al.* 2010). Therefore, the adaptive evolution seen in these two proteins may reflect their involvement in silencing transposons as previously suggested for *piwi* (Obbard *et al.* 2009; Kolaczowski *et al.* 2011).

Additionally, it is possible that selective pressure to repress transposons may be driving the adaptive evolution of *stwl* since some

other chromatin-associated proteins are involved in transposon silencing (Klattehoff *et al.* 2009; Rangan *et al.* 2011). Species-specific changes in life history and the timing of reproduction could also pose changing selective pressures on the germline (Schmidt and Paaby 2008), though our limited knowledge of the ages of reproduction for natural populations of *Drosophila* limits our ability to test this hypothesis.

In the future, it will be important to test whether these positively selected GSC genes function in the specific biological processes that we hypothesize are driving their adaptive evolution. For example, do *bam* or *bgen* play a role in regulating the transmission of bacterial endosymbionts, or does *stonewall* act in the repression of transposons? Additional insight may come from sampling these genes from additional *Drosophila* species to determine whether they have experienced a long-term selective pressure across many *Drosophila* or whether it is specific to *D. melanogaster* and *D. simulans*.

ACKNOWLEDGMENTS

We thank Pavlos Pavlidis for his help in performing the SweeD and OmegaPlus analyzes, and Jae Young Choi for scripting help. We also thank Jae Young Choi and Jaclyn Bubnell for useful comments and discussion, and two reviewers for their thoughtful and constructive criticisms. This research was supported by National Institutes of Health grant R01 GM095793 to C.F.A. and D.A.B. and an National Science Foundation Graduate Research Fellowship to H.A.F.

LITERATURE CITED

- Aagaard, J. E., V. D. Vacquier, M. J. MacCoss, and W. J. Swanson, 2010 ZP domain proteins in the abalone egg coat include a paralog of VERL under positive selection that binds lysin and 18-kDa sperm proteins. *Mol. Biol. Evol.* 27: 193–203.
- Anderson, J. A., W. D. Gilliland, and C. H. Langley, 2009 Molecular population genetics and evolution of *Drosophila* meiosis genes. *Genetics* 181: 177–185.
- Aquadro, C. F., K. M. Lado, and W. A. Noon, 1988 The rosy region of *Drosophila melanogaster* and *Drosophila simulans*. I. Contrasting levels of naturally occurring DNA restriction map variation and divergence. *Genetics* 119: 875–888.
- Aravin, A. A., G. J. Hannon, and J. Brennecke, 2007 The piwi-piRNA pathway provides an adaptive defense in the transposon arms race. *Science* 318: 761–764.
- Bauer DuMont, V. L., H. A. Flores, M. H. Wright, and C. F. Aquadro, 2007 Recurrent positive selection at *bgen*, a key determinant of germ line differentiation, does not appear to be driven by simple coevolution with its partner protein *bam*. *Mol. Biol. Evol.* 24: 182–191.
- Begun, D. J., and C. F. Aquadro, 1993 African and North American populations of *Drosophila melanogaster* are very different at the DNA level. *Nature* 365: 548–550.
- Begun, D. J., and C. F. Aquadro, 1994 Evolutionary inferences from DNA variation at the *6-phosphogluconate dehydrogenase* locus in natural populations of *Drosophila*: selection and geographic differentiation. *Genetics* 136: 155–171.
- Begun, D. J., P. Whitley, B. L. Todd, H. M. Waldrip-Dail, and A. G. Clark, 2000 Molecular population genetics of male accessory gland proteins in *Drosophila*. *Genetics* 156: 1879–1888.
- Begun, D. J., A. K. Holloway, K. Stevens, L. W. Hillier, Y. P. Poh *et al.*, 2007 Population genomics: whole-genome analysis of polymorphism and divergence in *Drosophila simulans*. *PLoS Biol.* 5: e310.
- Benjamini, Y., and Y. Hochberg, 1995 Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc., B* 57: 289–300.
- Brennecke, J., C. D. Malone, A. A. Aravin, R. Sachidanandam, A. Stark *et al.*, 2008 An epigenetic role for maternally inherited piRNAs in transposon silencing. *Science* 322: 1387–1392.
- Chen, D., and D. McKearin, 2005 Gene circuitry controlling a stem cell niche. *Curr. Biol.* 15: 179–184.
- Chrostek, E., M. S. Marialva, S. S. Esteves, L. A. Weinert, J. Martinez *et al.*, 2013 *Wolbachia* variants induce differential protection to viruses in *Drosophila melanogaster*: a phenotypic and phylogenomic analysis. *PLoS Genet.* 9: e1003896.
- Civetta, A., S. A. Rajakumar, B. Brouwers, and J. P. Bacik, 2006 Rapid evolution and gene-specific patterns of selection for three genes of spermatogenesis in *Drosophila*. *Mol. Biol. Evol.* 23: 655–662.
- Clark, K. A., and D. M. McKearin, 1996 The *Drosophila stonewall* gene encodes a putative transcription factor essential for germ cell development. *Development* 122: 937–950.
- Clark, N. L., and W. J. Swanson, 2005 Pervasive adaptive evolution in primate seminal proteins. *PLoS Genet.* 1: e35.
- Cox, D. N., A. Chao, and H. Lin, 2000 Piwi encodes a nucleoplasmic factor whose activity modulates the number and division rate of germline stem cells. *Development* 127: 503–514.
- Cutter, A. D., and S. Ward, 2005 Sexual and temporal dynamics of molecular evolution in *C. elegans* development. *Mol. Biol. Evol.* 22: 178–188.
- Drosophila* 12 Genomes Consortium, Clark, A. G., M. B. Eisen, D. R. Smith, C. M. Bergman *et al.*, 2007 Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* 450: 203–218.
- Duchen, P., D. Zivkovic, S. Hutter, W. Stephan, and S. Laurent, 2013 Demographic inference reveals African and European admixture in the North American *Drosophila melanogaster* population. *Genetics* 193: 291–301.
- DuMont, V. B., J. C. Fay, P. P. Calabrese, and C. F. Aquadro, 2004 DNA variability and divergence at the *Notch* locus in *Drosophila melanogaster* and *D. simulans*: a case of accelerated synonymous site divergence. *Genetics* 167: 171–185.
- Eyre-Walker, A., 2006 The genomic rate of adaptive evolution. *Trends Ecol. Evol.* 21: 569–575.
- Eyre-Walker, A., and P. D. Keightley, 2009 Estimating the rate of adaptive molecular evolution in the presence of slightly deleterious mutations and population size change. *Mol. Biol. Evol.* 26: 2097–2108.
- Forbes, A., and R. Lehmann, 1998 Nanos and pumilio have critical roles in the development and function of *Drosophila* germline stem cells. *Development* 125: 679–690.
- Fry, A. J., M. R. Palmer, and D. M. Rand, 2004 Variable fitness effects of *wolbachia* infection in *Drosophila melanogaster*. *Heredity* (Edinb) 93: 379–389.
- Fuller, M. T., and A. C. Spradling, 2007 Male and female *Drosophila* germline stem cells: Two versions of immortality. *Science* 316: 402–404.
- Gilboa, L., and R. Lehmann, 2004 How different is Venus from Mars? The genetics of germ-line stem cells in *Drosophila* females and males. *Development* 131: 4895–4905.
- Gilboa, L., A. Forbes, S. I. Tazuke, M. T. Fuller, and R. Lehmann, 2003 Germ line stem cell differentiation in *Drosophila* requires gap junctions and proceeds via an intermediate state. *Development* 130: 6625–6634.
- Glinka, S., L. Ometto, S. Mousset, W. Stephan, and D. De Lorenzo, 2003 Demography and natural selection have shaped genetic variation in *Drosophila melanogaster*: a multi-locus approach. *Genetics* 165: 1269–1278.
- Gonczy, P., E. Matunis, and S. DiNardo, 1997 Bag-of-marbles and benign gonial cell neoplasm act in the germline to restrict proliferation during *Drosophila* spermatogenesis. *Development* 124: 4361–4371.
- Hadrill, P. R., D. Bachtrog, and P. Andolfatto, 2008 Positive and negative selection on noncoding DNA in *Drosophila simulans*. *Mol. Biol. Evol.* 25: 1825–1834.
- Haerty, W., S. Jagadeeshan, R. J. Kulathinal, A. Wong, K. Ravi Ram *et al.*, 2007 Evolution in the fast lane: rapidly evolving sex-related genes in *Drosophila*. *Genetics* 177: 1321–1335.
- Hedges, L. M., J. C. Brownlie, S. L. O’Neill, and K. N. Johnson, 2008 *Wolbachia* and virus protection in insects. *Science* 322: 702.
- Hu, T. T., M. B. Eisen, K. R. Thornton, and P. Andolfatto, 2013 A second-generation assembly of the *Drosophila simulans* genome provides new insights into patterns of lineage-specific divergence. *Genome Res.* 23: 89–98.

- Hutter, S., H. Li, S. Beisswanger, D. De Lorenzo, and W. Stephan, 2007 Distinctly different sex ratios in African and European populations of *Drosophila melanogaster* inferred from chromosome-wide single nucleotide polymorphism data. *Genetics* 177: 469–480.
- Insko, M. L., A. Leon, C. H. Tam, D. M. McKearin, and M. T. Fuller, 2009 Accumulation of a differentiation regulator specifies transit amplifying division number in an adult stem cell lineage. *Proc. Natl. Acad. Sci. USA* 106: 22311–22316.
- Kawase, E., M. D. Wong, B. C. Ding, and T. Xie, 2004 Gbb/Bmp signaling is essential for maintaining germline stem cells and for repressing *bam* transcription in the *Drosophila* testis. *Development* 131: 1365–1375.
- Keightley, P. D., and A. Eyre-Walker, 2012 Estimating the rate of adaptive molecular evolution when the evolutionary divergence between species is small. *J. Mol. Evol.* 74: 61–68.
- Kelleher, E. S., and T. A. Markow, 2009 Duplication, selection and gene conversion in a *Drosophila mojavensis* female reproductive protein family. *Genetics* 181: 1451–1465.
- King, F. J., and H. Lin, 1999 Somatic signaling mediated by *fs(1)yb* is essential for germline stem cell maintenance during *Drosophila* oogenesis. *Development* 126: 1833–1844.
- King, F. J., A. Szakmary, D. N. Cox, and H. Lin, 2001 *Yb* modulates the divisions of both germline and somatic stem cells through *piwi*- and *hh*-mediated mechanisms in the *Drosophila* ovary. *Mol. Cell* 7: 497–508.
- Klattenhoff, C., H. Xi, C. Li, S. Lee, J. Xu *et al.*, 2009 The *Drosophila* HP1 homolog rhino is required for transposon silencing and piRNA production by dual-strand clusters. *Cell* 138: 1137–1149.
- Kolaczowski, B., D. N. Hupaló, and A. D. Kern, 2011 Recurrent adaptation in RNA interference genes across the *Drosophila* phylogeny. *Mol. Biol. Evol.* 28: 1033–1042.
- Langley, C. H., K. Stevens, C. Cardeno, Y. C. G. Lee, D. R. Schrider *et al.*, 2012 Genomic variation in natural populations of *Drosophila melanogaster*. *Genetics* 192: 533–598.
- Larracuente, A. M., T. B. Sackton, A. J. Greenberg, A. Wong, N. D. Singh *et al.*, 2008 Evolution of protein-coding genes in *Drosophila*. *Trends Genet.* 24: 114–123.
- Lavoie, C. A., B. Ohlstein, and D. M. McKearin, 1999 Localization and function of bam protein require the *benign gonial cell neoplasm* gene product. *Dev. Biol.* 212: 405–413.
- Lawniczak, M. K., and D. J. Begun, 2007 Molecular population genetics of female-expressed mating-induced serine proteases in *Drosophila melanogaster*. *Mol. Biol. Evol.* 24: 1944–1951.
- Li, H., and W. Stephan, 2006 Inferring the demographic history and rate of adaptive substitution in *Drosophila*. *PLoS Genet.* 2: e166.
- Li, Y., N. T. Minor, J. K. Park, D. M. McKearin, and J. Z. Maines, 2009 Bam and bcn antagonize nanos-dependent germ-line stem cell maintenance. *Proc. Natl. Acad. Sci. USA* 106: 9304–9309.
- Li, Y., J. Z. Maines, O. Y. Tastan, D. M. McKearin, and M. Buszczak, 2012 Mei-P26 regulates the maintenance of ovarian germline stem cells by promoting BMP signaling. *Development* 139: 1547–1556.
- Librado, P., and J. Rozas, 2009 DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25: 1451–1452.
- Lilly, M. A., M. de Cuevas, and A. C. Spradling, 2000 Cyclin A associates with the fusome during germline cyst formation in the *Drosophila* ovary. *Dev. Biol.* 218: 53–63.
- Lin, H., and A. C. Spradling, 1997 A novel group of pumilio mutations affects the asymmetric division of germline stem cells in the *Drosophila* ovary. *Development* 124: 2463–2476.
- Löytynoja, A., and N. Goldman, 2005 An algorithm for progressive multiple alignment of sequences with insertions. *Proc. Natl. Acad. Sci. USA* 102: 10557–10562.
- Mackay, T. F., S. Richards, E. A. Stone, A. Barbadilla, J. F. Ayroles *et al.*, 2012 The *Drosophila melanogaster* genetic reference panel. *Nature* 482: 173–178.
- Maines, J. Z., J. K. Park, M. Williams, and D. M. McKearin, 2007 Stonewalling *Drosophila* stem cell differentiation by epigenetic controls. *Development* 134: 1471–1479.
- Mateos, M., S. J. Castrezana, B. J. Nankivell, A. M. Estes, T. A. Markow *et al.*, 2006 Heritable endosymbionts of *Drosophila*. *Genetics* 174: 363–376.
- 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.
- Min, K. T., and S. Benzer, 1997 *Wolbachia*, normally a symbiont of *Drosophila*, can be virulent, causing degeneration and early death. *Proc. Natl. Acad. Sci. USA* 94: 10792–10796.
- Neumuller, R. A., J. Betschinger, A. Fischer, N. Bushati, I. Poernbacher *et al.*, 2008 Mei-P26 regulates microRNAs and cell growth in the *Drosophila* ovarian stem cell lineage. *Nature* 454: 241–245.
- Nielsen, R., S. Williamson, Y. Kim, M. J. Hubisz, A. G. Clark *et al.*, 2005 Genomic scans for selective sweeps using SNP data. *Genome Res.* 15: 1566–1575.
- Obbard, D. J., K. H. Gordon, A. H. Buck, and F. M. Jiggins, 2009 The evolution of RNAi as a defence against viruses and transposable elements. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364: 99–115.
- Ohlstein, B., C. A. Lavoie, O. Vef, E. Gateff, and D. M. McKearin, 2000 The *Drosophila* cystoblast differentiation factor, *benign gonial cell neoplasm*, is related to DEXH-box proteins and interacts genetically with *bag-of-marbles*. *Genetics* 155: 1809–1819.
- Olivieri, D., M. M. Sykora, R. Sachidanandam, K. Mechtler, and J. Brennecke, 2010 An in vivo RNAi assay identifies major genetic and cellular requirements for primary piRNA biogenesis in *Drosophila*. *EMBO J.* 29: 3301–3317.
- Pavlidis, P., J. D. Jensen, and W. Stephan, 2010a Searching for footprints of positive selection in whole-genome SNP data from nonequilibrium populations. *Genetics* 185: 907–922.
- Pavlidis, P., S. Laurent, and W. Stephan, 2010b msABC: a modification of Hudson's ms to facilitate multi-locus ABC analysis. *Mol. Ecol. Resour.* 10: 723–727.
- Pavlidis, P., D. Zivkovic, A. Stamatakis, and N. Alachiotis, 2013 SweeD: likelihood-based detection of selective sweeps in thousands of genomes. *Mol. Biol. Evol.* 30: 2224–2234.
- Phadnis, N., and H. A. Orr, 2009 A single gene causes both male sterility and segregation distortion in *Drosophila* hybrids. *Science* 323: 376–379.
- Pool, J. E., and C. F. Aquadro, 2006 History and structure of sub-saharan populations of *Drosophila melanogaster*. *Genetics* 174: 915–929.
- Pool, J. E., R. B. Corbett-Detig, R. P. Sugino, K. A. Stevens, C. M. Cardeno *et al.*, 2012 Population genomics of sub-saharan *Drosophila melanogaster*: African diversity and non-African admixture. *PLoS Genet.* 8: e1003080.
- Presgraves, D. C., 2007 Does genetic conflict drive rapid molecular evolution of nuclear transport genes in *Drosophila*? *BioEssays* 29: 386–391.
- Prokupek, A., F. Hoffmann, S. I. Eyun, E. Moriyama, M. Zhou *et al.*, 2008 An evolutionary expressed sequence tag analysis of *Drosophila* spermatheca genes. *Evolution* 62: 2936–2947.
- Rangan, P., C. D. Malone, C. Navarro, S. P. Newbold, P. S. Hayes *et al.*, 2011 piRNA production requires heterochromatin formation in *Drosophila*. *Curr. Biol.* 21: 1373–1379.
- Saito, K., H. Ishizu, M. Komai, H. Kotani, Y. Kawamura *et al.*, 2010 Roles for the Yb body components Armitage and Yb in primary piRNA biogenesis in *Drosophila*. *Genes Dev.* 24: 2493–2498.
- Schmidt, P. S., and A. B. Paaby, 2008 Reproductive diapause and life-history clines in North American populations of *Drosophila melanogaster*. *Evolution* 62: 1204–1215.
- Singh, N. D., J. D. Jensen, A. G. Clark, and C. F. Aquadro, 2013 Inferences of demography and selection in an African population of *Drosophila melanogaster*. *Genetics* 193: 215–228.
- Song, X., C. H. Zhu, C. Doan, and T. Xie, 2002 Germline stem cells anchored by adherens junctions in the *Drosophila* ovary niches. *Science* 296: 1855–1857.
- Song, X., M. D. Wong, E. Kawase, R. Xi, B. C. Ding *et al.*, 2004 Bmp signals from niche cells directly repress transcription of a differentiation-promoting

- gene, *bag of marbles*, in germline stem cells in the *Drosophila* ovary. *Development* 131: 1353–1364.
- Swanson, W. J., and V. D. Vacquier, 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, W. J., and V. D. Vacquier, 2002 The rapid evolution of reproductive proteins. *Nat. Rev. Genet.* 3: 137–144.
- Swanson, W. J., Z. Yang, M. F. Wolfner, and C. F. Aquadro, 2001a Positive Darwinian selection drives the evolution of several female reproductive proteins in mammals. *Proc. Natl. Acad. Sci. USA* 98: 2509–2514.
- Swanson, W. J., A. G. Clark, H. M. Waldrip-Dail, M. F. Wolfner, and C. F. Aquadro, 2001b Evolutionary EST analysis identifies rapidly evolving male reproductive proteins in *Drosophila*. *Proc. Natl. Acad. Sci. USA* 98: 7375–7379.
- Swanson, W. J., A. Wong, M. F. Wolfner, and C. F. Aquadro, 2004 Evolutionary expressed sequence tag analysis of *Drosophila* female reproductive tracts identifies genes subjected to positive selection. *Genetics* 168: 1457–1465.
- 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.
- Szakmary, A., M. Reedy, H. Qi, and H. Lin, 2009 The Yb protein defines a novel organelle and regulates male germline stem cell self-renewal in *Drosophila melanogaster*. *J. Cell Biol.* 185: 613–627.
- Tamura, K., J. Dudley, M. Nei, and S. Kumar, 2007 MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* 24: 1596–1599.
- Tazuke, S. I., C. Schulz, L. Gilboa, M. Fogarty, A. P. Mahowald *et al.*, 2002 A germline-specific gap junction protein required for survival of differentiating early germ cells. *Development* 129: 2529–2539.
- Teixeira, L., A. Ferreira, and M. Ashburner, 2008 The bacterial symbiont *Wolbachia* induces resistance to RNA viral infections in *Drosophila melanogaster*. *PLoS Biol.* 6: e2.
- Tsaur, S. C., C. T. Ting, and C. I. Wu, 1998 Positive selection driving the evolution of a gene of male reproduction, *Acp26Aa*, of *Drosophila*: II. divergence vs. polymorphism. *Mol. Biol. Evol.* 15: 1040–1046.
- Wang, Z., and H. Lin, 2004 Nanos maintains germline stem cell self-renewal by preventing differentiation. *Science* 303: 2016–2019.
- Watts, T., T. S. Haselkorn, N. A. Moran, and T. A. Markow, 2009 Variable incidence of *Spiroplasma* infections in natural populations of *Drosophila* species. *PLoS ONE* 4: e5703.
- Werren, J. H., 2005 Heritable microorganisms and reproductive parasitism, pp. 290–315 in *Microbial Phylogeny and Evolution. Concepts and Controversies*, edited by J. Sapp. Oxford University Press, Oxford.
- Wong, M. D., Z. Jin, and T. Xie, 2005 Molecular mechanisms of germline stem cell regulation. *Annu. Rev. Genet.* 39: 173–195.
- Xia, L., S. Jia, S. Huang, H. Wang, Y. Zhu *et al.*, 2010 The Fused/Smurf complex controls the fate of *Drosophila* germline stem cells by generating a gradient BMP response. *Cell* 143: 978–990.
- Yang, Z., 1997 PAML: A program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* 13: 555–556.
- Yang, Z., 2007 PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 24: 1586–1591.

Communicating editor: R. Kulathinal