

Relocation Facilitates the Acquisition of Short *Cis*-Regulatory Regions that Drive the Expression of Retrogenes during Spermatogenesis in *Drosophila*

Mehran Sorourian,^{†,1} Mansi M. Kunte,^{†,1} Susana Domingues,¹ Miguel Gallach,² Fulya Özdil,^{‡,1} Javier Río,¹ and Esther Betrán^{*,1}

¹Department of Biology, University of Texas at Arlington

²Center for Integrative Bioinformatics Vienna (CIBIV), Max F Perutz Laboratories, University of Vienna and Medical University of Vienna, Vienna, Austria

[‡]Present address: Department of Agricultural Biotechnology, Faculty of Agriculture, Namık Kemal University, Tekirdağ, Turkey

[†]These authors contributed equally to this work.

*Corresponding author: E-mail: betran@uta.edu.

Associate editor: Patricia Wittkopp

Abstract

Retrogenes are functional processed copies of genes that originate via the retrotranscription of an mRNA intermediate and often exhibit testis-specific expression. Although this expression pattern appears to be favored by selection, the origin of such expression bias remains unexplained. Here, we study the regulation of two young testis-specific *Drosophila* retrogenes, *Dntf-2r* and *Pros28.1A*, using genetic transformation and the enhanced green fluorescent protein reporter gene in *Drosophila melanogaster*. We show that two different short (<24 bp) regions upstream of the transcription start sites (TSSs) act as testis-specific regulatory motifs in these genes. The *Dntf-2r* regulatory region is similar to the known $\beta 2$ tubulin 14-bp testis motif ($\beta 2$ -tubulin gene upstream element 1 [$\beta 2$ -UE1]). Comparative sequence analyses reveal that this motif was already present before the *Dntf-2r* insertion and was likely driving the transcription of a noncoding RNA. We also show that the $\beta 2$ -UE1 occurs in the regulatory regions of other testis-specific retrogenes, and is functional in either orientation. In contrast, the *Pros28.1A* testes regulatory region in *D. melanogaster* appears to be novel. Only *Pros28.1B*, an older paralog of the *Pros28.1* gene family, seems to carry a similar regulatory sequence. It is unclear how the *Pros28.1A* regulatory region was acquired in *D. melanogaster*, but it might have evolved de novo from within a region that may have been preprimed for testes expression. We conclude that relocation is critical for the evolutionary origin of male germline-specific *cis*-regulatory regions of retrogenes because expression depends on either the site of the retrogene insertion or the sequence changes close to the TSS thereafter. As a consequence we infer that positive selection will play a role in the evolution of these regulatory regions and can often act from the moment of the retrocopy insertion.

Key words: *Drosophila*, retrogene, testis expression, regulatory element, gene duplication, gene relocation.

Introduction

Retrogenes are functional processed copies of genes (Brosius 1999), which originate when an mRNA is reverse transcribed and reinserted into the genome. At the time of insertion retrogenes are devoid of introns, have a poly-A tail, and are flanked by short direct repeats. However, the last two features are quickly lost and not found in most *Drosophila* retrogenes (Betrán et al. 2002). Although the chromosome of origin of the polymorphic retrocopies appears to fit random expectations in *Drosophila* (Schridder et al. 2011), flies exhibit an excess of retrogenes that have originated from genes on the X chromosome and are inserted in autosomes (Betrán et al. 2002; Dai et al. 2006; Bai et al. 2007; Meisel et al. 2009; Vibranovski, Zhang, et al. 2009; Han and Hahn 2012). In addition, most retrogenes exhibit testis-specific expression whereas parental genes are broadly expressed (Betrán et al. 2002; Bai et al. 2007; Vibranovski, Lopes, et al. 2009; Han and Hahn 2012). Similar patterns have also been observed in

placental mammal and marsupial genomes (Emerson et al. 2004; Potrzebowski et al. 2008). The insertion and the expression patterns do not correspond to the retropositional biases across the genome and have, therefore, been attributed to the action of selection (Emerson et al. 2004; Vinckenbosch et al. 2006; Schridder et al. 2011, 2013).

At the time of retroposition, all retroposed copies lack the regulatory regions (enhancer and promoter regions) of their parental genes. Thus, to be preserved in the genome, most retrocopies that will eventually become functional genes will either be inserted downstream of an existing regulatory region or quickly evolve one de novo. It is often assumed that the insertion of a retrocopy between an existing gene and its regulatory region would perturb the gene's regulation and be deleterious. However, in the case of an insertion into a duplicated region where there might be redundancy, the insertion has a higher chance of evolving into a new gene (Long and Langley 1993; Wang et al. 2000). The insertion site may exert particular effects on the transcription of the newly

inserted gene. Initial transcription may occur if the insertion is close to transcribed genes or, in the case of testis-expressed retrogenes, in regions that are enriched with testis-expressed genes, that is, testis neighborhoods (Kalmykova et al. 2005; Vinckenbosch et al. 2006; Bai et al. 2008; Dorus et al. 2008). Either transcriptional facilitation or coregulation has been proposed as mechanisms to explain this effect (Loppin et al. 2005; Vinckenbosch et al. 2006). Transcriptional facilitation could occur if the retrocopy is inserted in an actively transcribed open chromatin region. This mechanism would increase the probability of starting transcription of the new gene using a cryptic promoter. On the other hand, gene coregulation could take place through the sharing of enhancers (Lemos et al. 2008) that can act long or short range and affect more than one gene (Loppin et al. 2005; Vinckenbosch et al. 2006). However, in this case, the evolution of a core promoter region in the retrocopy might be needed for the proper start of transcription.

In this study, we experimentally characterize the regulatory regions of two relatively young retrogenes, *Dntf-2r* (CG10174) and *Pros28.1A* (CG17268), and their evolutionary origin. We show that a short (i.e., 14 bp) *cis*-regulatory region similar to the $\beta 2$ -*tubulin* gene upstream element 1 ($\beta 2$ -UE1) of the $\beta 2$ -*tubulin* gene (CG9359; Michiels et al. 1989) is needed for the testis-specific expression of *Dntf-2r* in *Drosophila melanogaster*. We also determine that the motif was present before *Dntf-2r* insertion, it is found upstream of other testis-specific retrogenes and either orientation (+ or –) can drive testis-specific expression. A novel regulatory region that includes 23-bp upstream of the transcription start site (TSS) is needed to drive testis-specific expression of *Pros28.1A* in *D. melanogaster*. Only *Pros28.1B*, another testis-specific gene of the same gene family, seems to share part of this region. *Pros28.1B* is an older DNA-mediated duplication of *Pros28.1* and it is, at this point, unclear how its regulatory region was acquired. These case studies illustrate that the evolutionary origin of male germline-specific *cis*-regulatory regions depends on relocating a coding region to a new context where regulatory regions for testis expression either already exist or can evolve close to the TSS.

Results

Dntf-2r Is Expressed from a $\beta 2$ -UE1-Like Element

Dntf-2r originated less than 12 Ma (Betran and Long 2003) and is present in four species of *Drosophila*: *D. melanogaster*, *D. simulans*, *D. sechellia*, and *D. mauritiana* (Betran and Long 2003). It shows testis- or male-biased expression in all species where it is present (Betran and Long 2003; Chintapalli et al. 2007; supplementary results, Supplementary Material online). To identify the regulatory region that drives the testis-specific expression of *Dntf-2r* in *D. melanogaster*, clones carrying the region upstream of the previously described TSS (Betran and Long 2003), its 5'-UTR and *Dntf-2r* coding region fused to EGFP (Enhanced Green Fluorescent Protein) as a reporter gene were transformed in *Drosophila* (see Materials and Methods). The longest construct contained 151-bp upstream of the TSS including 25 bp of the 5'-UTR of the gene *bicoid*

stability factor (*bsf*) located upstream of *Dntf-2r* in a head-to-head orientation (fig. 1A). The transformed flies expressed the green fluorescent fusion protein (fig. 1A and supplementary fig. S4A, Supplementary Material online) in a pattern that mimics the *Dntf-2r* RNA in situ profile (fig. 1B) suggesting that this construct harbors the complete testis-specific regulatory region of *Dntf-2r*. Subsequently, we generated constructs carrying –101, –77, –50, and –4 bp upstream of the TSS. Of those, only the –101 and –77 bp constructs show testis-specific green fluorescence (fig. 1A and supplementary fig. S4A–C, Supplementary Material online). The others did not show fluorescence (i.e., the observed fluorescence was comparable to the autofluorescence of the white mutant control, w^{1118} ; supplementary fig. S4D and E, Supplementary Material online). Independent insertion lines show similar results for each construct (supplementary fig. S4A–E, Supplementary Material online) indicating that the constructs themselves, and not the region where the constructs are inserted, are responsible for the observations. Green fluorescence is also observed in the male gonads of the second instar larvae of the –151 bp lines (supplementary fig. S4F, Supplementary Material online). No consistent fluorescence is observed in other tissues such as the gut and ovaries for this line (supplementary fig. S4G, Supplementary Material online) when compared with the control (w^{1118}), thus confirming the testis-specific expression of the construct. We conclude that a 27-bp region (i.e., the region between –77 and –50 bp shaded in yellow in fig. 2A) is needed to drive the expression of the *Dntf-2r*-EGFP fusion gene in testis.

We observed differences in the intensity of the green fluorescence between –151 and –101 and –77 bp, which we could not properly quantify. Specifically, flies containing the –151 bp construct appear to fluoresce stronger in testis than the flies carrying the shorter fluorescing constructs –101 and –77 bp, but in the same pattern. Interestingly, we found two potential BEAF-32 insulator binding sites at –135 to –129 and –115 to –109 bp. These two BEAF-32 insulator sequences appear to be used at least during embryogenesis (FlyBase; Negre et al. 2010) and, consistent with the previously described cases, might prevent coexpression of head-to-head genes (Yang et al. 2012). These insulator sequences are absent in the shorter constructs and it is unclear if or how the lack of these sequences can actually affect the levels of expression of the genes and our constructs in testis. Therefore, we concluded that a 27-bp region (i.e., the region between –77 and –50 bp shaded in yellow in fig. 2A) is needed to drive the expression of the *Dntf-2r*-EGFP fusion gene in testis although other upstream regions may also influence the expression level.

The 27-bp region contains a $\beta 2$ -UE1-like sequence and the upstream region of *Dntf-2r* also harbors a sequence (fig. 2A) that is identical to the 7-bp motif with quantitative effects described in the testis-specific $\beta 2$ *tubulin* gene (GGATATT; Michiels et al. 1989). Those motifs are found in *D. simulans*, *D. sechellia*, and *D. mauritiana* but have positional differences compared with *D. melanogaster* (supplementary fig. S3, Supplementary Material online). We infer from the sequence

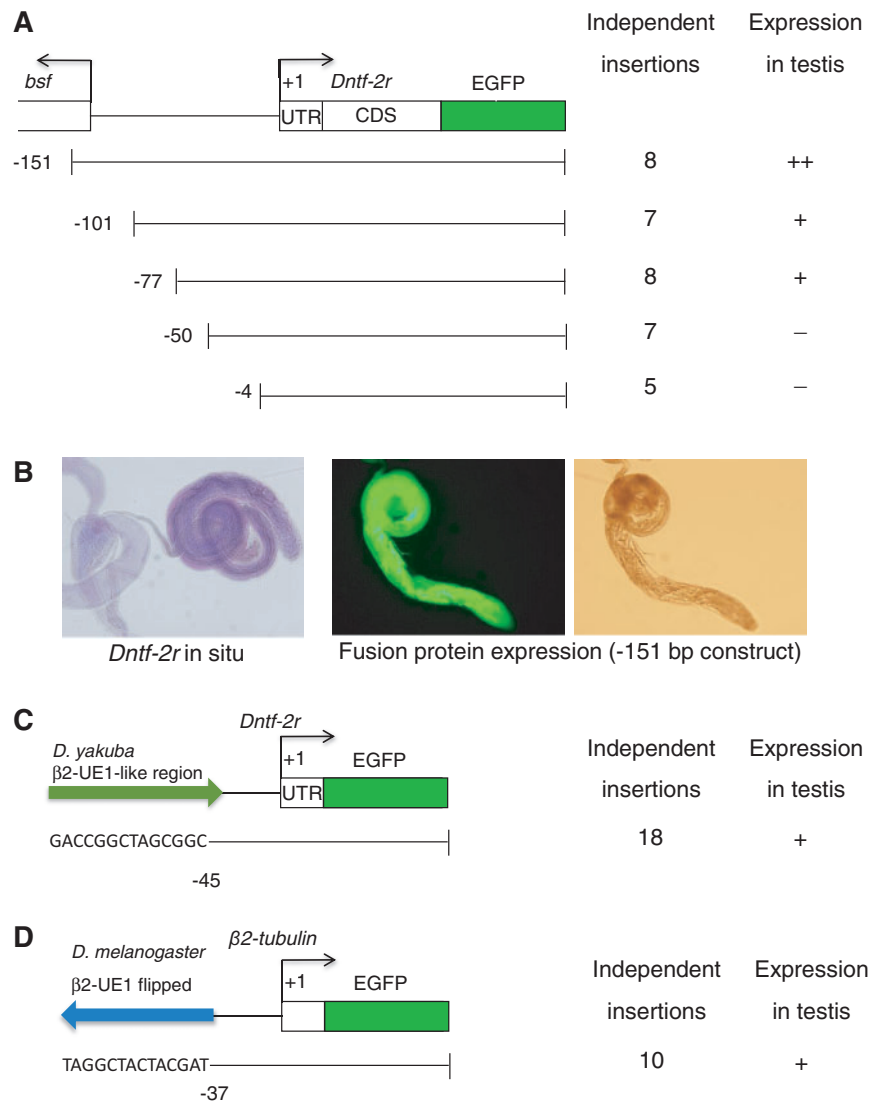


Fig. 1. (A) *Dntf-2r*-EGFP fusion constructs used for transformation in *Drosophila melanogaster*. The drawn line depicts the regions included in every construct. It is drawn proportional to the region and its length varies from -151 to -4 bp from the TSS. The white box depicts 5'-UTR and coding region (CDS) of *Dntf-2r* (not drawn to scale). +1 denotes the start of transcription (i.e., TSS) of the gene fusion. Green box depicts EGFP coding region (not drawn to scale). Fluorescence results due to the testis-specific expression of the fusion protein are summarized. See [supplementary figure S4, Supplementary Material](#) online, for fluorescent pictures of all independent insertions. All constructs except the last two show strong fluorescence only in testis revealing that a testis-specific element is located between -78 and -51 bp from the TSS. (B) In situ results revealing the transcription in testis of *Dntf-2r* are shown and coincide with the observed expression of the fusion protein. (C) The construct including *D. yakuba* β 2-UE1-like region ([fig. 2](#)) is outlined and the fluorescence results summarized. See [supplementary figure S5, Supplementary Material](#) online, for fluorescent pictures of all independent insertions. (D) The construct including β 2-UE1 flipped and the β 2-tubulin region ([fig. 2](#)) is outlined and the fluorescence results summarized. See [supplementary figure S8, Supplementary Material](#) online, for fluorescent pictures of all independent insertions.

conservation that these elements are most likely functional in the other species as well.

The β 2-UE1-Like Element Was Present before the Insertion of *Dntf-2r*

To address the evolutionary origin of the β 2-UE1-like element of *Dntf-2r*, we compared the region before and after the insertion of the retrogene and attempted to reconstruct the ancestral state of this genomic location using maximum parsimony. The multiple alignment of the orthologous

sequences revealed that the motifs upstream of *Dntf-2r* are highly conserved even in species that do not have the retrogene ([fig. 2B](#)). We inferred with confidence that the motif with quantitative effects (GGATATT) was present at the time of insertion and that the β 2-UE1-like region was at least partially present ([fig. 2B](#)).

Indirect evidence supports that this region was functional before the insertion of *Dntf-2r* and drove testis expression. First, we find that the β 2-UE1-like and the motif with quantitative effects are partly and completely conserved, respectively, in *D. yakuba*, *D. teissieri* and *D. erecta*, where *Dntf-2r* is

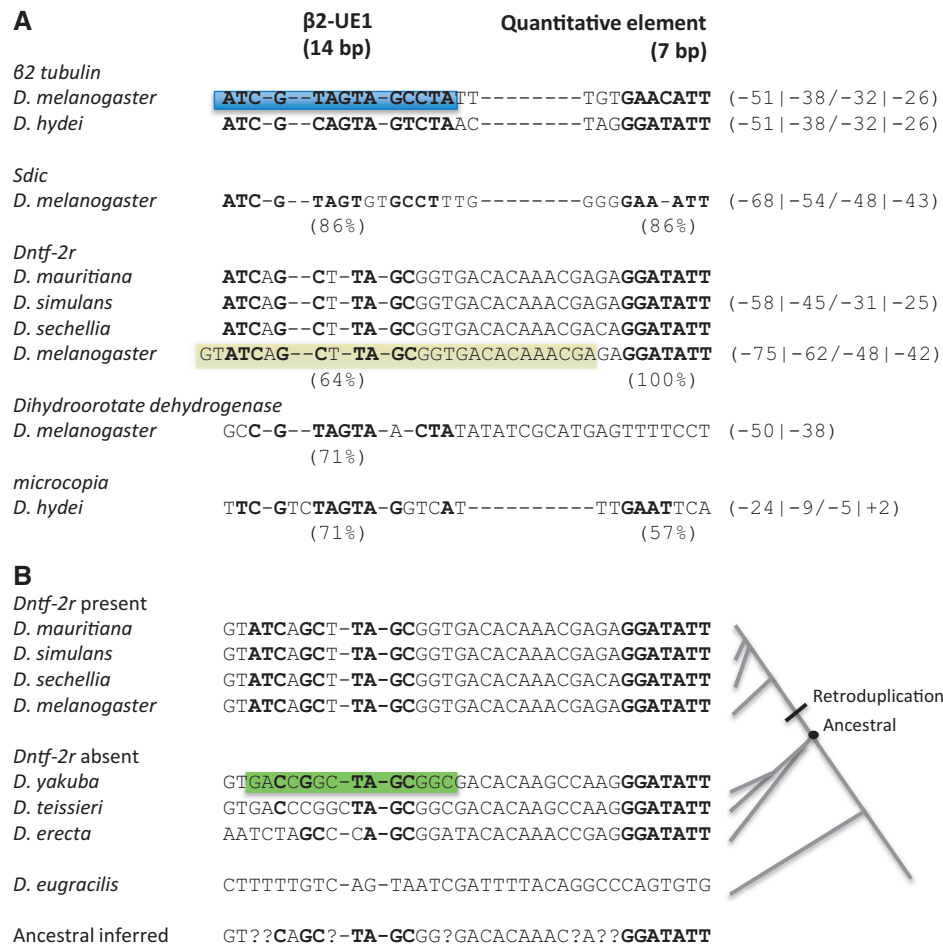


Fig. 2. (A) Percent identity of the upstream region of several testis-specific genes to the known 14-bp β 2-UE1 motif and the 7-bp motif with quantitative effect of the β 2 *tubulin* gene. Identical bases are shown in bold. Negative numbers show the location of both elements as distance to the TSS. For example, β 2-UE1 motif of the β 2 *tubulin* gene in *Drosophila melanogaster* spans the base pairs from -51 to -38 and the 7-bp motif with quantitative effects in the same gene and species spans the base pairs from -32 to -26. This information is consolidated from our study and work of others (Michiels et al. 1989; Lankeau et al. 1994; Yang et al. 1995; Nurminsky et al. 1998). Yellow highlighted region shows the sequence required for testis-specific expression of *Dntf-2r* in *D. melanogaster*. (B) *Dntf-2r* upstream region in the species where the gene is present and the same region in species where the gene is absent. Similar sequences were observed in the syntenic region in distantly related species suggesting that a testis-specific regulatory region existed before *Dntf-2r* insertion. The nucleotide state at the ancestral node was inferred when the same nucleotide was observed in species where *Dntf-2r* is present and in species where *Dntf-2r* is absent and not inferred (?) otherwise. We highlight in green the region used from *D. yakuba* to make a construct to show the current regulatory potential of the 14 bp in this species.

absent (fig. 2). In addition, the BEAF-32 binding sites are also conserved in those species (Data not shown). This supports a head-to-head arrangement of *bsf* with another gene even before the insertion of *Dntf-2r*. Recently, a noncoding RNA gene (*let-7-C*; CR43344) has been annotated in *D. melanogaster* downstream of *Dntf-2r*. This noncoding RNA shows a male-biased pattern of expression similar to *Dntf-2r* according to modENCODE data (Marygold et al. 2013) and its location is conserved in *D. yakuba*, *D. teissieri* and *D. erecta* (Data not shown), indicating that the regulatory regions of *Dntf-2r* likely predated its origin.

To verify that this region is functional in species where *Dntf-2r* does not exist (i.e., species most likely carrying the ancestral configuration of this genomic region), we fused and transformed a construct containing the 14-bp orthologous region (See green highlight in fig. 2B) from *D. yakuba* and the rest of the region from *D. melanogaster*

in front of EGFP and checked whether it drives testis-specific expression (See Materials and Methods). As proposed above, we observed testis-specific fluorescence in most of the independent transformants containing the construct (supplementary fig. S5, Supplementary Material online) and concluded that the region is capable of driving testis-specific expression at a discernible level (fig. 1C) when compared with constructs lacking the *D. yakuba* β 2-UE1-like region (supplementary fig. S4D and E, Supplementary Material online). All the evidence indicates that the region contained a testis-specific regulatory element before the insertion of *Dntf-2r*.

Pros28.1A Is Expressed from a Novel Testis-Specific Cis-Regulatory Element

Pros28.1A is estimated to be approximately 39 My old (Sorourian and Betran 2010). Previous work (Sorourian and Betran 2010) revealed that transgenic flies carrying 246 bp

upstream of the TSS, the 5'-UTR of *Pros28.1A*, and EGFP as a reporter gene show a pattern of expression that recapitulates the known testis expression pattern of the gene. Here, in an effort to elucidate in more detail the testis-specific regulatory region of this gene, five additional constructs were made. The constructs contained -189, -80, -46, -23, and 0 bp upstream of the *Pros28.1A* TSS (fig. 3A). Only the flies transformed with the construct lacking any upstream region lacked EGFP expression in the testis (fig. 3 and supplementary fig. S6, Supplementary Material online). So, we used site-directed mutagenesis to completely change the 23-bp region immediately upstream of *Pros28.1A* TSS of *D. melanogaster* to make another construct (see supplementary materials and methods, Supplementary Material online, and fig. 3A). This construct carrying a different 23-bp region lacked EGFP expression in the testis (fig. 3 and supplementary fig. S6, Supplementary Material online). The transformants showing the testis-specific expression pattern in adult testis showed expression in larval male gonads (supplementary fig. S6A–D and G, Supplementary Material online), but not in other tissues (supplementary fig. S6H, Supplementary Material online). In addition, we observe a lower level of expression as the constructs are shortened (fig. 3A and supplementary fig. S6A–D, Supplementary Material online). However, we have not quantified this effect. The decrease in expression level could potentially be explained by the presence of motif/s with quantitative effects upstream of the testis-specific element as proposed for the $\beta 2$ tubulin gene (Michiels et al. 1989) or the loss of unknown insulator elements in shorter constructs as we have inferred for *Dntf-2r* in this work (see above). We concluded that information contained within the 23 bp upstream of *Pros28.1A* TSS (i.e., 5'-GAAAAATTCATATTGTTCAAGT-3'; fig. 3B) is needed for testis-specific expression of *Pros28.1A* in *D. melanogaster*.

The 23 bp upstream of *Pros28.1A* TSS included in the shortest fluorescent construct (fig. 3B; bold highlight) does not appear to be entirely conserved among species that carry this retrogene but there are at least 5 bp (i.e., GAAAA) conserved. In addition, we could not find any of the promoter elements previously described in this region, including the $\beta 2$ -tubulin testis-specific promoter (Michiels et al. 1989; Ohler et al. 2002; Fitzgerald et al. 2006). Thus, we infer that the 23-bp region contains a novel testis-specific *cis*-regulatory element.

Interestingly, there are 13 base pairs of this 23-bp region that show the highest similarity between *Pros28.1A* and *Pros28.1B* upstream region (i.e., 5'-CATATTGTTCAA-3' *Pros28.1A* vs. 5'-CATATTGTTTAA-3' *Pros28.1B*). The expression of both genes (*Pros28.1A* and *Pros28.1B*) is testis specific and sensitive to mutations in meiotic arrest genes (see below and Mikhaylova et al. 2006), and this shared sequence might potentially be responsible for the similar regulation of these two genes. However, as with *Pros28.1A*, the upstream region of *Pros28.1B* does not appear to be more conserved among species than other adjacent noncoding regions (data not shown). This may indicate that the 23-bp regulatory region of *Pros28.1B* has changed in other lineages or recently originated in the common ancestor of *D. melanogaster*, *D. simulans*, *D. sechellia*, and *D. mauritiana*. It remains possible that

the few base pairs that are conserved between all species that have *Pros28.1A* (fig. 3B) are part of the motif/s needed for testis-specific expression in *D. melanogaster* and that the *D. melanogaster* regulatory region evolved from within a region that was preprimed for testes expression.

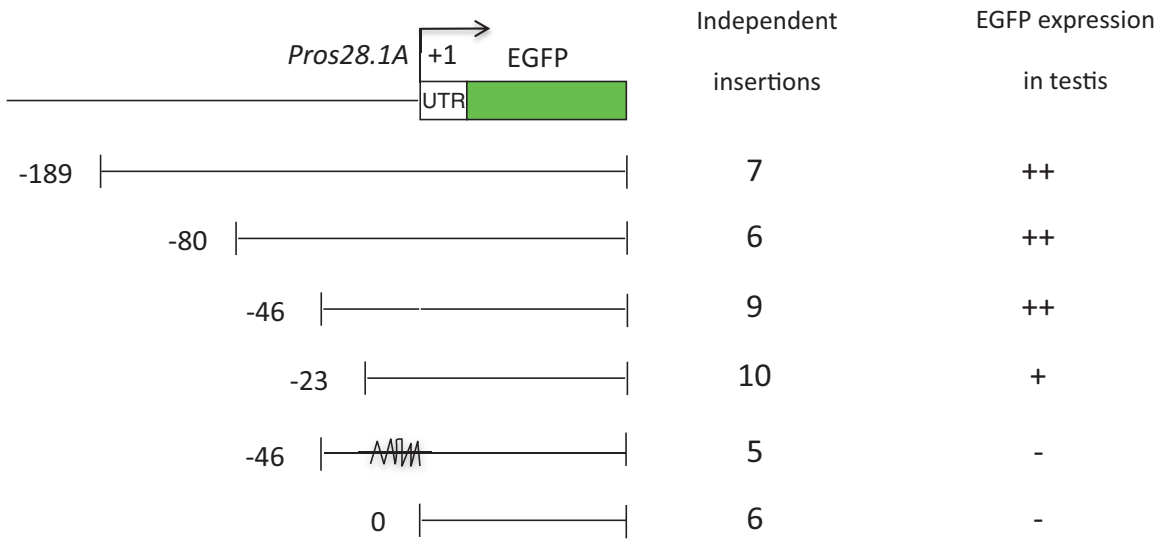
The $\beta 2$ -UE1 Motif Occurs in the Upstream Region of Other Testis-Specific Retrogenes

We studied whether the 14-bp $\beta 2$ -UE1 motif of the $\beta 2$ tubulin gene (5'-ATCGTAGTAGCCTA-3') and the 13-bp conserved portion between *Pros28.1A* and *Pros28.1B* (5'-CATATTGTTT[C/T]AA-3'), also occur in upstream regions of other genes. A preliminary BLAST analysis of both motifs found only the 14-bp $\beta 2$ -UE1 motif in several testis-specific genes. We generated a position-specific scoring matrix (PSSM; supplementary fig. S7 and table S1, Supplementary Material online) from the BLAST hits alignment, which we gave to GLAM2SCAN as an input matrix to scan the upstream region of all *D. melanogaster* genes (see Materials and Methods for more details). We found that 120 upstream regions of the 1,245 testis-specific genes (10%) and 476 upstream regions of the 13,253 nontestis-specific genes (i.e., 4%) contained a copy of the $\beta 2$ -UE1 motif indicating that testis-specific genes are significantly enriched for this motif ($P = 2.91 \times 10^{-19}$; Fisher's exact test). Remarkably, 60 of the 476 nontestis-specific genes containing a copy of the $\beta 2$ -UE1 motif upstream of their TSS are actually testis-biased (i.e., genes that are nontestis-specific under our definition [see Materials and Methods], but whose level of transcription in testis was 10-fold higher than the average in the whole fly). Out of the 120 testis-specific genes, 13 are known testis-specific retrogenes (Betran et al. 2002; Dai et al. 2006; Bai et al. 2007).

Interestingly, 54 $\beta 2$ -UE1-like motifs follow the same strand orientations as the testis-specific gene, yet there are 61 in the opposite strand. This observation prompted us to generate a construct that carries the upstream region and 5'-UTR of the $\beta 2$ -tubulin gene but where the 14-bp $\beta 2$ -UE1 motif has been flipped (i.e., inverted; fig. 1D). Ten independent transformants were examined and showed fluorescence specifically in testis (supplementary fig. S8, Supplementary Material online) demonstrating that the $\beta 2$ -UE1 motif works in both orientations and could potentially work bidirectionally.

Transcription of *Dntf-2r* and *Pros28.1A* Is Sensitive to Mutations in Meiotic Arrest Genes

It has been suggested that the $\beta 2$ -UE1-like promoter of the testis-biased gene *Sdic* is bound by the regulatory protein Modulo and it has been shown that transcription of *Sdic* is sensitive to mutations in meiotic arrest genes (e.g., *rye*) and *modulo* (*mod*; Mikhaylova et al. 2006). The $\beta 2$ tubulin gene transcription is also affected by mutations in meiotic arrest genes (Mikhaylova et al. 2006). We analyzed the transcription of *Dntf-2r* in wild-type (Oregon-R), *rye* (a meiotic arrest gene), and *mod* mutant strains by quantitative real-time RT-PCR (QRT-PCR) (See Materials and Methods). Significant differences ($F_{[2,6]} = 5.6551$; $P = 0.0416$) were observed among the

A *D. melanogaster* constructs

B

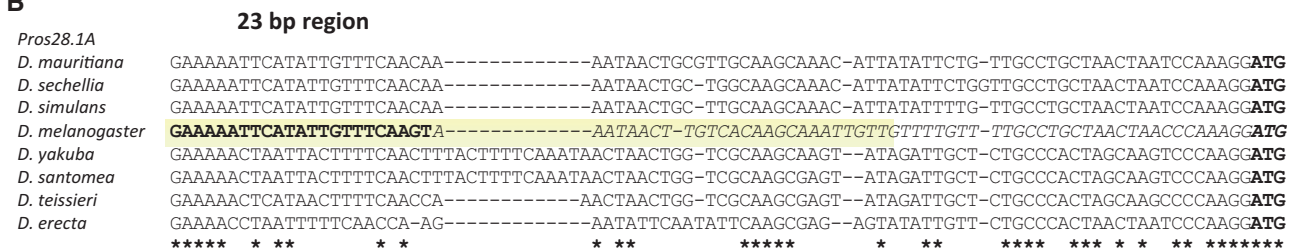


FIG. 3. (A) Constructs carrying the *Pros28.1A* upstream region and 5'-UTR in front of EGFP used for transformation in *Drosophila melanogaster* and a summary of their level of fluorescence are shown. Line depicts the region included in the construct and its length varies from 189 to 0 bp (not drawn to scale). White box depicts 5'-UTR of *Pros28.1A* (not drawn to scale). +1 denotes the start of transcription of the gene. Green box depicts EGFP coding region (not drawn to scale). Zigzag line denotes the 23-bp region that was mutagenized. Lack of EGFP expression in the shortest construct and the mutagenic construct of a 23-bp region is observed revealing that the regulatory regions needed for testis expression include the regions between -23 and 0 bp. (B) Alignment of the upstream region of *Pros28.1A* of *D. melanogaster* with the same region in other species that contain the gene. Yellow highlighted region (49 bp) shows the sequence that was included in the 23-bp construct in front of EGFP and it also depicts how much of the 5'-UTR was included in all the constructs (A). The 23-bp region required for testis-specific expression is shown in bold and the *D. melanogaster* transcript is shown in italics.

mean difference for the normalized threshold cycle number (ΔC_T ; supplementary fig. S9, Supplementary Material online) between strains. Post-hoc Tukey tests revealed a significant difference between Oregon-R and *rye* mutant strain ($P=0.0464$) and the same trend, although not significant ($P=0.0867$), was observed after a posteriori testing between Oregon-R and *mod* mutant strains. Fold-down changes in the expression were calculated using the expression $2^{-\Delta\Delta C_T}$ (Schmittgen and Livak 2008). *Dntf-2r* fold changes are similar in both mutant comparisons with Oregon-R: 1.8109 for *mod* and 2.0280 for *rye*. Although many genes can be affected in meiotic arrest mutants because spermatogenesis does not occur (White-Cooper 2010), our observation is consistent with what has been observed for genes that harbor the β 2-UE1-like promoter.

We also analyzed the transcription of *Pros28.1A* in a wild-type strain (Oregon-R), and *rye* and *mod* mutant strains by QRT-PCR (see Materials and Methods). Significant differences

($F_{[2,6]} = 11.0127$; $P=0.0098$) were observed among the mean difference for the normalized threshold cycle number (ΔC_T ; supplementary fig. S9, Supplementary Material online) between strains. Post-hoc Tukey tests revealed a significant difference between Oregon-R and the *rye* mutant strain ($P=0.0181$), and between the *mod* and *rye* mutants ($P=0.0138$). No differences were observed between Oregon-R and the *mod* mutant strain ($P=0.9674$). Fold-down regulatory changes in the expression were similar in both comparisons with the *rye* mutant strain: 5.2659 times lower expression than Oregon-R and 5.8411 times lower than *mod* mutant strain. We conclude that *Pros28.1A* is a testis-biased gene whose expression is sensitive to mutations in meiotic arrest genes, but not to mutations in the protein modulo. This is similar to the sensitivity observed for *Pros28.1B* transcription (a different *Pros28.1* duplicate that also has spermatogenesis-specific expression; Yuan et al. 1996). *Pros28.1B* transcription was affected in the *rye*

mutant strain, but only slightly affected in *mod* mutants (Mikhaylova et al. 2006). Again, although many genes can be affected in meiotic arrest mutants because spermatogenesis does not occur (White-Cooper 2010), our observation is consistent with a similar regulatory mechanism used for these two *Pros28.1* duplicates as the effects that we tested are consistent between these genes.

Discussion

Nature and Origin of Regulatory Regions of Retrogenes

Retrogenes lack regulatory regions at the time of insertion and only retrogenes that either insert downstream of an existing *cis*-regulatory region and promoter or quickly evolve a regulatory region *de novo* will be retained over the course of evolution. In this study, we report one example for each of these evolutionary scenarios.

Dntf-2r is transcribed from a preexisting testis regulatory region upstream from the insertion site. The regulatory region has 64% identity with the β 2-UE1 promoter of the *β 2-tubulin* gene in *D. melanogaster* and also contains the 7-bp motif with quantitative effects of the *β 2-tubulin* gene (Michiels et al. 1989). These motifs were most likely present before the insertion of *Dntf-2r* and functional in all the species where *Dntf-2r* is present. We hypothesize that the β 2-UE1 promoter was initially driving testis-specific transcription of a noncoding RNA (*let-7-C*), still present downstream of *Dntf-2r*. Because the regulatory region of *Dntf-2r* does not seem to have duplicated, it is difficult to explain why the insertion of *Dntf-2r* does not affect the regulation of *let-7-C*. Importantly though, *Dntf-2r* insertion would have likely been expressed in testis upon insertion and, therefore, selection would have potentially acted from the very beginning to increase the frequency of this retrogene in the population until all individuals would have carried this gene.

On the other hand, the lack of sizeable conservation (only five base pairs seem to be conserved in all species that contain the gene) of the regulatory sequence of *Pros28.1A* in different species is in agreement with the observation that TSS of *Pros28.1A* varies among species despite the conserved testis-specific expression pattern (Sorourian and Betran 2010). The order of events that eventually led to testis-expression upon insertion of *Pros28.1A* is unclear. Our data support the current existence of a new motif in *D. melanogaster* and, potentially, in some close related species that carry this testis-specific gene but not in others in favor of a constantly evolving testis-specific regulatory region.

Consistent with the results of various transgenic experiments for many testis-specific genes (see White-Cooper 2010 for a review), we not only find that very short *cis*-regulatory regions are required for the expression of *Dntf-2r* and *Pros28.1A* in male germline but also infer the existence of boundary elements and motif with quantitative effects around these regions. In addition, we reveal that relocation is critical for the evolutionary origin of male germline-specific *cis*-regulation of retrogenes because this pattern of expression depends on losing the initially broad/housekeeping

expression of the parental genes (Bai et al. 2007) and gaining short and testis-specific expression from either the site of retrogene insertion or sequence changes close to the TSS thereafter.

The observed effect of mutations in meiotic arrest genes in the expression of *Dntf-2r* and *Pros28.1A* is consistent with the effect found in other genes carrying similar regulatory motifs (Mikhaylova et al. 2006) and with a newly proposed regulatory complex involving *Ntx1* in *Drosophila* that couples nuclear export of intronless genes with the testis meiotic arrest complex and regulates the transcription of those genes specifically during spermatogenesis (Caporilli et al. 2013).

Is There Low Specificity and/or High Turnover of Regulatory Regions of Testis-Specific Genes?

We observe that the known β 2-UE1-like regulatory sequences used by several testis-specific genes in *D. melanogaster* (fig. 2A) actually have low identity to each other even when gaps are allowed in the alignment. We also found that the regulatory region of *Pros28.1A* shows low conservation, most likely revealing a turnover of the regulatory regions in *Drosophila* testis genes. This reveals low specificity of the transcription factors binding those sites or the presence of related, but divergent, regulatory complexes and high turnover of the regulatory regions.

Why should there be low specificity and/or high turnover of regulatory regions of testis-specific genes? It is known that some of the general transcription factors (i.e., TATA-binding protein-associated factors also known as TBP-associated factors or TAFs) have testis-specific isoforms (i.e., tTAFs; Sassone-Corsi 2002; Hiller et al. 2004) that contribute to the regulated deployment of testis-specific genes. Recently, it has been observed that the testis-specific TAFs duplicates and *mod* evolve under positive selection (Haerty et al. 2007; Li et al. 2009), and that a new testis-specific copy of a transcriptional cofactor has originated through duplication modifying *Drosophila* testis expression (i.e., a duplicate of the ortholog of human *Rcd-1*; Quezada-Diaz et al. 2010; Chen et al. 2012). Although the high turnover of regulatory regions (i.e., same regulation despite the lack of sequence conservation) compared with protein-coding regions might be a general feature of regulatory regions (Ludwig et al. 2000; Hare et al. 2008), rapid evolution under positive selection of the testis-specific TAFs and *mod* and new regulatory gene acquisitions reveal that there are particular selective pressures in this tissue that likely drive an even faster evolutionary turnover of regulatory regions. Many genes have been observed to evolve under positive selection in testis probably as a response to male–male competition, sexual antagonism, segregation distortion, and/or parasite-related conflicts (Zhang et al. 2004; Haerty et al. 2007; Presgraves 2007; Presgraves and Stephan 2007). These selective pressures might also act on regulatory regions because the gain or loss of testis expression (i.e., gain or loss of a *cis*-regulatory region) of newly duplicated genes (Kaessmann 2010) or of antisense RNAs to regulate parasites in testis (Lankenau et al. 1994) will be under strong selection. We predict that more data will accumulate supporting this idea

and the evolutionary pattern observed for *Drosophila* male germline gene regulation in this work should hold in other lineages.

Materials and Methods

Drosophila Stocks and Fly Handling

Several wild-type strains were used to PCR amplify the regulatory region of *Dntf-2r* in different species: *D. simulans* from Florida (provided by J. Coyne), *D. sechellia* (provided by J. Coyne), *D. mauritiana* (163.1; Lemeunier and Ashburner 1976), *D. teisseri* (118.2; Lemeunier and Ashburner 1976), *D. yakuba* (115; Lemeunier and Ashburner 1976), *D. erecta* (154.1; Lemeunier and Ashburner 1976), and *D. eugracilis* (14026-0451-07; UC San Diego *Drosophila* Stock Center).

Quantitative Real-Time RT-PCR

Tissues were homogenized and total RNA was prepared from whole males of three strains (*Taf12L*^{KG00946} [rye], and *mod*⁰⁷⁵⁷⁰ mutant strains and Oregon R) for comparison. The QIAGEN RNeasy mini kit (QIAGEN, Valencia, CA) and protocol were used for RNA extractions. QRT-PCR was conducted for the retrogenes using oligoprimers 5'-AGCCACGAA GAGGGATCCTC3' and 5'-TTGTCCAGCAGTACTACGCC-3' for *Dntf-2r* and 5'-AACTGGGTACCTGCAGAAGG-3' and 5'-TGCCTGCTAACTAACCCAAAC-3' for *Pros18.1A*. *Gapdh2* was used as a normalizer in the QRT-PCR. The oligoprimers used were 5'-TCAGCCATCACAGTTCGATTC-3' and 5'-CAAAC GAACATGGGAGCATC-3'. QRT-PCRs were performed using the ABI 7300 Real Time PCR system and the SYBR Green PCR Core reagents from Applied Biosystems (Foster City, CA). Three replicates of RT+ and RT- were performed for every retrogene, normalizing gene, and RNA extraction. The primers described above produced similar amplification plots in the logarithmic scale that correspond to similar efficiency (Schmittgen and Livak 2008). QRT-PCR products were run in gels to control for any spurious amplification. Threshold cycle numbers (C_T values) were obtained with the default ABI software parameters. C_T values obtained for the retrogenes were normalized by subtracting the C_T value of the normalizing gene (ΔC_T). The mean difference for the normalized threshold cycle number ($\Delta \Delta C_T$) in different tissues was tested using analysis of variance. Post-hoc Tukey tests were also performed. Changes in the expression were calculated using the expression $2^{-\Delta \Delta C_T}$ (Schmittgen and Livak 2008).

5'-Rapid Amplification of cDNA Ends

RNA ligase-mediated rapid amplification of cDNA ends (RLM-RACE) was performed to identify the TSS for *Dntf-2r* in *D. simulans* using the First choice RLM-RACE kit from Ambion (Applied Biosystems/Ambion, Austin, TX). One microgram of RNA extracted from adult males was used for 5'-RACE in *D. simulans* (fig. 2). The gene-specific outer primer 5'-CCGTTGGGCTTCAGCAAAAAGAT-3' and inner primer 5'-CATCGATTTTAGTCTTCCAAGGACG-3' were used along with the 5'-RACE outer and inner primers provided by the manufacturer, respectively.

DNA Samples and Sequencing

To obtain divergence data for the regulatory region of *Dntf-2r*, DNA was extracted from a single fly using the Puregene DNA purification system from QIAGEN (Valencia, CA). The region was amplified by PCR from genomic DNA using oligoprimers 5'-TGCAGGGCGCATTGTTTCAG-3' and 5'-CATACGCCTGCC AATACGAGT-3' to amplify *Dntf-2r* in strains of *D. simulans*, *D. sechellia*, and *D. mauritiana*. PCR products were sequenced directly after purification (QIAGEN PCR purification kit; QIAGEN, Valencia, CA) on an ABI automated DNA sequencer using fluorescent DyeDeoxy terminator reagents (Applied Biosystems, Foster City, CA).

Preparation of the Constructs, Transformation, and Imaging

Dntf-2r, *Pros28.1A*, and Flipped $\beta 2$ -UE1 Constructs

The complete *Dntf-2r* coding region and variable lengths of the upstream regions were amplified from genomic DNA and cloned into the plasmid pEGFP1 (U55761; Clontech, Mountain View, CA) to put the *Dntf-2r* in frame with the EGFP gene and generate a fluorescent fusion protein. The regions of these clones containing different lengths of the putative regulatory region, 5'-UTR of *Dntf-2r*, *Dntf-2r*-EGFP fused coding regions and the SV40 polyadenylation site were further cloned into the P element *Drosophila* transformation vector—pCaSpeR 4 (X81645) and used for fly transformation.

The 14-bp $\beta 2$ -UE1-like sequence of *D. melanogaster* was replaced by the orthologous 14 bp of *D. yakuba* to understand whether this sequence drives testis-specific expression. The region of *Dntf-2r* was amplified using the forward primer 5'-CACCGACCGCTAGCGGCGACACAAACGAGAGGATA TTTG-3' that added the *D. yakuba* sequence (italics) to the *D. melanogaster* region and the reverse primer 5' TTTAGTTC AAGTATATACGGGGTA-3'. The PCR product was cloned with topoisomerase-catalyzed reaction using pENTRYTM Directional TOPO cloning kit (pENTR/D-TOPO Cloning Kit, Invitrogen catalog # K240020; Invitrogen, Carlsbad, CA). The forward primer added the necessary recombination site to the PCR product for directional cloning into the entry clone using the Gateway system. Colonies were screened by colony PCR using primers flanking the gene followed by sequencing of the positive clones. Miniprep for a good clone was performed, making this clone our master entry clone for recombination into the UAS-GFP destination vector. To perform the recombination into the destination vector, we used the Gateway LR Clonase Enzyme Mix (Invitrogen, Carlsbad, CA). Transformation was performed using Library Efficiency DH5alpha competent cells and the colonies were then sequenced by PCR to find positive clones. One good clone was chosen after sequencing and sent for transformation.

Another construct was made that included the described regulatory regions of the $\beta 2$ -tubulin gene and its 5'-UTR (Michiels et al. 1989) but in which the $\beta 2$ -UE1 element was flipped. This region was introduced upstream of EGFP in the pCaSpeR 4 transformation vector and used for fly transformation.

PCR primers were designed to amplify an increasingly shorter upstream region of *Pros28.1A* from a previous *Pros28.1A*–EGFP pCaSpeR 4 transformation vector (Sorourian and Betran 2010). The products containing different lengths of the *Pros28.1A* putative regulatory region, 5′-UTR of *Pros28.1A* in front of EGFP and the SV40 polyadenylation site were cloned in the pCaSpeR 4 transformation vector and used for fly transformation.

These clones were sent for transformation into *w¹¹¹⁸* flies to Genetic Services Inc. (Cambridge, MA). Several independent transformants were obtained for every construct (see Results). More details about the primers designed and cloning procedures are given as [supplementary materials and methods, Supplementary Material](#) online.

Imaging

Tissues from transformant flies were dissected in Ringer solution and images were taken using an Olympus BX51TRF fluorescent microscope setting the UV exposure time manually at a level established by examining the control strain (*w¹¹¹⁸*, provided by Genetic Services Inc., Cambridge, MA). Controlling the time allowed us to fine tune for any autofluorescence in the original white mutant stock and enabled us to compare the level of green fluorescence across preparations and lines.

In Situ hybridization

In situ hybridization of whole testis of young males of the Besançon strain was performed to detect the transcript of *Dntf-2r* following the protocol described by Morris et al. (2009) with some changes. We used single-stranded DNA probes instead of RNA probes. To denature the probe, it was heated at 100 °C for 10 min. The prehybridization and hybridization steps were carried out at 45 °C. The overnight incubation at 4 °C was done with 0.1% BSA in PBST and the antidigoxygenin antibody. After the color development, the reactions were stopped with PBST washes (four times for 10 min each). After the four washes, all the PBST was removed and 30% glycerol in PBST was added for 30 min, followed by 50% glycerol, and, finally, 70% glycerol was added.

The probe labeling was performed using DIG-High Prime DNA Labeling and Detection Starter Kit I (Roche, Indianapolis, IN). A sense oligoprimers (5′-CTTTTTTTCGGATCGGAAACTCAATCGTACCCCGTATATACTTGAAGTAAATGTCTCTGAATCTGCAGT-3′) allowed us to label the complementary DNA strand using the random priming procedure. This procedure was designed to detect the *Dntf-2r* transcript specifically as the oligo corresponds to a 5′-UTR region that is different from the four transcripts of the parental gene, *Dntf-2-RA*, *Dntf-2-RB*, *Dntf-2-RC*, and *Dntf-2-RE*.

Motif Discovery

We downloaded the expression profile for the whole *D. melanogaster* gene set from Chintapalli et al. (2007), available from FlyAtlas (<http://www.flyatlas.org>, last accessed May 28, 2014). We considered a gene to be transcribed testis-specifically when it was only upregulated in testis and downregulated not in any other tissue (as in Chintapalli et al.

2007). A total of 1,245 genes were selected as testis-specific, including *Dntf-2r*, *Pros28.1A*, and *Pros28.1B*. The region of 300 bp upstream of the TSS of every gene in the *D. melanogaster* genome was downloaded from dmel_r5.24. A sequence containing the β2-UE1 of the β2 tubulin gene (AAATCGTAGTAGCCTATT) and a 13-bp testis-specific putative motif of *Pros28.1A* (see below) was BLASTed against these sequences. GLAM2 (Fritch et al. 2008) was used to build a PSSM from the BLAST hits, which was later given to the program GLAM2SCAN in order to scan the whole genome and find copies of the motifs. We used the top 26,000 motif matches found by GLAM2SCAN in the genome to be sure we included as many potential motif matches as possible in our analyses. Those matches have a score value of 6.21 or higher. Matches located in upstream sequences of testis-specific genes (positive motifs) were used to build another PSSM that is used in the next iteration. This procedure allowed us to refine the PSSM and test for enrichment of the motif in testis-specific genes given an arbitrary score. The process continued for five iterations, that is, until motif enrichment in testis-specific genes compared with nontestis-specific genes was no longer observed.

Supplementary Material

Supplementary materials and methods, results, [table S1](#), and [figures S1–S9](#) are available at *Molecular Biology and Evolution* online (<http://www.mbe.oxfordjournals.org/>).

Acknowledgments

The authors thank J. Coyne, P. Gibert, F. Lemeunier, M. Long, M.-L. Wu, C.-I. Wu, and the UC San Diego *Drosophila* Stock Center for providing wild-type *Drosophila* strains used in this work and Genetic Services, Inc. and the Bloomington *Drosophila* Stock Center at Indiana University for providing the mutant stocks used in this study. They also thank H. White-Cooper for providing them with her in situ hybridization protocol before publication and for all her advice on its implementation in their lab. They want to thank four anonymous reviewers for comments on this work. M.G. wants to thank Arndt von Haeseler for his support when finalizing the manuscript. This work was supported by the National Institute of General Medical Sciences of the National Institutes of Health under award number R01GM071813, an ARRA Supplement and UT Arlington startup funds to E.B. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

- Bai Y, Casola C, Betran E. 2008. Evolutionary origin of regulatory regions of retrogenes in *Drosophila*. *BMC Genomics* 9:241.
- Bai Y, Casola C, Feschotte C, Betran E. 2007. Comparative genomics reveals a constant rate of origination and convergent acquisition of functional retrogenes in *Drosophila*. *Genome Biol.* 8:R11.
- Betran E, Long M. 2003. *Dntf-2r*: a young *Drosophila* retroposed gene with specific male expression under positive Darwinian selection. *Genetics* 164:977–988.
- Betran E, Thornton K, Long M. 2002. Retroposed new genes out of the X in *Drosophila*. *Genome Res.* 12:1854–1859.

- Brosius J. 1999. RNAs from all categories generate retrosequences that may be exapted as novel genes or regulatory elements. *Gene* 238:115–134.
- Caporilli S, Yu Y, Jiang J, White-Cooper H. 2013. The RNA export factor, NXT1, is required for tissue specific transcriptional regulation. *PLoS Genet.* 9:e1003526.
- Chen S, Ni X, Krinsky BH, Zhang YE, Vibranovski MD, White KP, Long M. 2012. Reshaping of global gene expression networks and sex-biased gene expression by integration of a young gene. *EMBO J.* 31: 2798–2809.
- Chintapalli VR, Wang J, Dow JA. 2007. Using FlyAtlas to identify better *Drosophila melanogaster* models of human disease. *Nat Genet.* 39: 715–720.
- Dai H, Yoshimatsu TF, Long M. 2006. Retrogene movement within- and between-chromosomes in the evolution of *Drosophila* genomes. *Gene* 385:96–102.
- Dorus S, Freeman ZN, Parker ER, Heath BD, Karr TL. 2008. Recent origins of sperm genes in *Drosophila*. *Mol Biol Evol.* 25:2157–2166.
- Emerson JJ, Kaessmann H, Betran E, Long M. 2004. Extensive gene traffic on the mammalian X chromosome. *Science* 303:537–540.
- Fitzgerald PC, Sturgill D, Shyakhtenko A, Oliver B, Vinson C. 2006. Comparative genomics of *Drosophila* and human core promoters. *Genome Biol.* 7:R53.
- Frith MC, Saunders NF, Kobe B, Bailey TL. 2008. Discovering sequence motifs with arbitrary insertions and deletions. *PLoS Comput Biol.* 4: e1000071.
- Haerty W, Jagadeeshan S, Kulathinal RJ, Wong A, Ravi Ram K, Sirot LK, Levesque L, Artieri CG, Wolfner MF, Civetta A, et al. 2007. Evolution in the fast lane: rapidly evolving sex-related genes in *Drosophila*. *Genetics* 177:1321–1335.
- Han MV, Hahn MW. 2012. Inferring the history of interchromosomal gene transposition in *Drosophila* using n-dimensional parsimony. *Genetics* 190:813–825.
- Hare EE, Peterson BK, Iyer VN, Meier R, Eisen MB. 2008. Sepsid even-skipped enhancers are functionally conserved in *Drosophila* despite lack of sequence conservation. *PLoS Genet.* 4:e1000106.
- Hiller M, Chen X, Pringle MJ, Suchorolski M, Sancak Y, Viswanathan S, Bolival B, Lin TY, Marino S, Fuller MT. 2004. Testis-specific TAF homologs collaborate to control a tissue-specific transcription program. *Development* 131:5297–5308.
- Kaessmann H. 2010. Origins, evolution and phenotypic impact of new genes. *Genome Res.* 20:1313–1326.
- Kalmykova AI, Nurminsky DI, Ryzhov DV, Shevelyov YY. 2005. Regulated chromatin domain comprising cluster of co-expressed genes in *Drosophila melanogaster*. *Nucleic Acids Res.* 33:1435–1444.
- Lankenau S, Corces VG, Lankenau DH. 1994. The *Drosophila* micropia retrotransposon encodes a testis-specific antisense RNA complementary to reverse transcriptase. *Mol Cell Biol.* 14:1764–1775.
- Lemeunier F, Ashburner MA. 1976. Relationships within the melanogaster species subgroup of the genus *Drosophila* (Sophophora). II. Phylogenetic relationships between six species based upon polytene chromosome banding sequences. *Proc R Soc Lond B Biol Sci.* 193:275–294.
- Lemos B, Landry CR, Fontanillas P, Renn SCP, Kulathinal R, Brown KM, Hartl DL. 2008. Evolution of genomic expression. In: Pagel M, Pomiankowski A, editors. *Evolutionary genomics and proteomics*. Sunderland (MA): Sinauer Associates. p. 81–118.
- Li VC, Davis JC, Lenkov K, Bolival B, Fuller MT, Petrov DA. 2009. Molecular evolution of the testis TAFs of *Drosophila*. *Mol Biol Evol.* 26:1103–1116.
- Long M, Langley CH. 1993. Natural selection and the origin of jingwei, a chimeric processed functional gene in *Drosophila*. *Science* 260: 91–95.
- Loppin B, Lepetit D, Dorus S, Couble P, Karr TL. 2005. Origin and neofunctionalization of a *Drosophila* paternal effect gene essential for zygote viability. *Curr Biol.* 15:87–93.
- Ludwig MZ, Bergman C, Patel NH, Kreitman M. 2000. Evidence for stabilizing selection in a eukaryotic enhancer element. *Nature* 403: 564–567.
- Marygold SJ, Leyland PC, Seal RL, Goodman JL, Thurmond J, Strelets VB, Wilson RJ. 2013. FlyBase: improvements to the bibliography. *Nucleic Acids Res.* 41:D751–D757.
- Meisel RP, Han MV, Hahn MW. 2009. A complex suite of forces drives gene traffic from *Drosophila* X chromosomes. *Genome Biol Evol.* 1: 176–188.
- Michiels F, Gasch A, Kaltschmidt B, Renkawitz-Pohl R. 1989. A 14 bp promoter element directs the testis specificity of the *Drosophila* beta 2 tubulin gene. *EMBO J.* 8:1559–1565.
- Mikhaylova LM, Boutanaev AM, Nurminsky DI. 2006. Transcriptional regulation by Modulo integrates meiosis and spermatid differentiation in male germ line. *Proc Natl Acad Sci U S A.* 103: 11975–11980.
- Morris CA, Benson E, White-Cooper H. 2009. Determination of gene expression patterns using in situ hybridization to *Drosophila* testes. *Nat Protoc.* 4:1807–1819.
- Negre N, Brown CD, Shah PK, Kheradpour P, Morrison CA, Henikoff JG, Feng X, Ahmad K, Russell S, White RA, et al. 2010. A comprehensive map of insulator elements for the *Drosophila* genome. *PLoS Genet.* 6: e1000814.
- Nurminsky DI, Nurminskaya MV, De Aguiar D, Hartl DL. 1998. Selective sweep of a newly evolved sperm-specific gene in *Drosophila*. *Nature* 396:572–575.
- Ohler U, Liao GC, Niemann H, Rubin GM. 2002. Computational analysis of core promoters in the *Drosophila* genome. *Genome Biol.* 3:RESEARCH0087.
- Potrzebowski L, Vinckenbosch N, Marques A, Chalmel F, Jegou B, Kaessmann H. 2008. Chromosomal gene movements reflect the recent origin and biology of thalian sex chromosomes. *PLoS Biol.* 6:e80.
- Presgraves DC. 2007. Does genetic conflict drive rapid molecular evolution of nuclear transport genes in *Drosophila*? *Bioessays* 29: 386–391.
- Presgraves DC, Stephan W. 2007. Pervasive adaptive evolution among interactors of the *Drosophila* hybrid inviability gene, Nup96. *Mol Biol Evol.* 24:306–314.
- Quezada-Diaz JE, Mulyil T, Río J, Betran E. 2010. Drcd-1 related: a positively selected spermatogenesis retrogene in *Drosophila*. *Genetica* 138:925–937.
- Sassone-Corsi P. 2002. Unique chromatin remodeling and transcriptional regulation in spermatogenesis. *Science* 296: 2176–2178.
- Schmittgen TD, Livak KJ. 2008. Analyzing real-time PCR data by the comparative C(T) method. *Nat Protoc.* 3:1101–1108.
- Schrider DR, Navarro FC, Galante PA, Parmigiani RB, Camargo AA, Hahn MW, de Souza SJ. 2013. Gene copy-number polymorphism caused by retrotransposition in humans. *PLoS Genet.* 9:e1003242.
- Schrider DR, Stevens K, Cardeno CM, Langley CH, Hahn MW. 2011. Genome-wide analysis of retrogene polymorphisms in *Drosophila melanogaster*. *Genome Res.* 21:2087–2095.
- Sorourian M, Betran E. 2010. Turnover and lineage-specific broadening of the transcription start site in a testis-specific retrogene. *Fly (Austin)* 4:3–11.
- Vibranovski MD, Lopes HF, Karr TL, Long M. 2009. Stage-specific expression profiling of *Drosophila* spermatogenesis suggests that meiotic sex chromosome inactivation drives genomic relocation of testis-expressed genes. *PLoS Genet.* 5:e1000731.
- Vibranovski MD, Zhang Y, Long M. 2009. General gene movement off the X chromosome in the *Drosophila* genus. *Genome Res.* 19: 897–903.
- Vinckenbosch N, Dupanloup I, Kaessmann H. 2006. Evolutionary fate of retroposed gene copies in the human genome. *Proc Natl Acad Sci U S A.* 103:3220–3225.
- Wang W, Zhang J, Alvarez C, Llopart A, Long M. 2000. The origin of the jingwei gene and the complex modular structure of its parental gene, yellow emperor, in *Drosophila melanogaster*. *Mol Biol Evol.* 17: 1294–1301.
- White-Cooper H. 2010. Molecular mechanisms of gene regulation during *Drosophila* spermatogenesis. *Reproduction* 139:11–21.

- Yang J, Porter L, Rawls J. 1995. Expression of the dihydroorotate dehydrogenase gene, *dhod*, during spermatogenesis in *Drosophila melanogaster*. *Mol Gen Genet.* 246:334–341.
- Yang J, Ramos E, Corces VG. 2012. The BEAF-32 insulator coordinates genome organization and function during the evolution of *Drosophila* species. *Genome Res.* 22:2199–2207.
- Yuan X, Miller M, Belote JM. 1996. Duplicated proteasome subunit genes in *Drosophila melanogaster* encoding testes-specific isoforms. *Genetics* 144:147–157.
- Zhang Z, Hambuch TM, Parsch J. 2004. Molecular evolution of sex-biased genes in *Drosophila*. *Mol Biol Evol.* 21: 2130–2139.