PNAS RESEARCH ARTICLE | DEVELOPMENTAL BIOLOGY

oskar acts with the transcription factor Creb to regulate long-term memory in crickets

Arpita Kulkarni[a,1], Ben Ewen-Campen[a,2], Kanta Terao[b,3], Yukihisa Matsumoto[c,3], Yaolong Li[b], Takayuki Watanabe[c,d], Jonchee A. Kao[e], Swapnil S. Parhad[f,4,5], Guillem Ylla[a,6], Makoto Mizunami[c], and Cassandra G. Extavour[a,e,g,7]

Edited by Claude Desplan, New York University, New York, NY; received October 29, 2022; accepted March 28, 2023

Novel genes have the potential to drive the evolution of new biological mechanisms, or to integrate into preexisting regulatory circuits and contribute to the regulation of older, conserved biological functions. One such gene, the novel insect-specific gene *oskar,* was first identified based on its role in establishing the *Drosophila melanogaster* germ line. We previously showed that this gene likely arose through an unusual domain transfer event involving bacterial endosymbionts and played a somatic role before evolving its well-known germ line function. Here, we provide empirical support for this hypothesis in the form of evidence for a neural role for *oskar*. We show that *oskar* is expressed in the adult neural stem cells of a hemimetabolous insect, the cricket *Gryllus bimaculatus*. In these stem cells, called neuroblasts, *oskar* is required together with the ancient animal transcription factor *Creb* to regulate long-term (but not short-term) olfactory memory. We provide evidence that *oskar* positively regulates *Creb*, which plays a conserved role in long-term memory across animals, and that *oskar* in turn may be a direct target of Creb. Together with previous reports of a role for *oskar* in nervous system development and function in crickets and flies, our results are consistent with the hypothesis that *oskar*'s original somatic role may have been in the insect nervous system. Moreover, its colocalization and functional cooperation with the conserved pluripotency gene *piwi* in the nervous system may have facilitated *oskar*'s later co-option to the germ line in holometabolous insects.

co-option | neuroblast | mushroom body | Kenyon cells | orphan genes

oskar (*osk*) is an insect-specific gene first discovered in *Drosophila melanogaster,* where it plays a critical role in germ line specification (1). *osk* messenger RNA is localized to the posterior of the developing *D. melanogaster* oocyte (2, 3). Local translation and anchoring of Oskar (Osk) protein leads to the posterior accumulation of the messenger RNA and protein products of several genes with conserved expression and function in animal germ lines, including *vasa* and *piwi* (2, 4, 5). Collectively called germ plasm, these cytoplasmic contents act as necessary and sufficient determinants to specify embryonic germ cells (2, 3). The current model of Osk function in *D. melanogaster* germ plasm assembly is that it serves as a scaffolding protein, facilitating the assembly of the ribonucleoprotein complexes that contain germ plasm components (2, 6, 7).

Interestingly, *osk* and several other genes originally identified as *D. melanogaster* germ line genes, including *vasa*, *pumilio*, *staufen*, *orb*, and *piwi*-related genes including *aubergine* and *argonaute 3*, have since been shown to have a variety of roles in animal nervous systems (8–14). For example, in *D. melanogaster*, *osk* RNA interference (RNAi) in larval dendritic arborization neurons disrupts *nanos* mRNA localization, ultimately leading to a defect in dendrite morphogenesis and an associated defect in motor response to mechanical stimulation (12). Furthermore, *osk* plays a role in the embryonic nervous system, but not in the germ line, in a hemimetabolous insect, the cricket *Gryllus bimaculatus*, where it is important for proper neuroblast divisions and subsequent axonal patterning (15). Our recent analysis of hundreds of previously unidentified *osk* orthologs across insects showed that *osk* is expressed in at least a dozen somatic tissues in species across the insect tree (16). This suggests that a somatic function of *osk* may be ancestral. However, the precise roles of *osk* in any somatic tissue, including the nervous system, remain largely unknown.

Here, we demonstrate a role for *osk* in the adult brain of the cricket *G. bimaculatus*, in a population of neural stem cells in the mushroom body that persist throughout adult life. We show that *osk*, as well as Piwi and Vasa, is enriched in a population of adult neuroblasts in the mushroom body, and that RNAi targeting *osk* or *piwi* in adult crickets impairs long-term, but not short-term, memory formation in an olfactory associative learning assay (17). We also provide evidence that *osk* and *piwi* function in a regulatory feedback loop with the cyclic adenosine monophosphate response element binding

Significance

When new genes evolve, why are they not immediately eliminated from the genome, given that the organism did not previously need them for survival? Here, we examine the hypothesis that novel genes can survive by evolving regulatory interactions with preexisting genes that control essential fitness traits. We provide experimental evidence that the unique insect-specific *oskar* gene plays a role in long-term memory through expression in adult neural stem cells. We show that *oskar* is expressed and required in stem cells of the adult cricket brain for long-term memory. Further, we provide evidence that *oskar's* role in long-term memory involves regulation by the transcription factor Creb, a conserved player in long-term memory that predates the origin of *oskar* in animals.

Preprint servers: https://www.biorxiv.org/content/10.1101/2022.10.24.513429v2.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]Present address: Single Cell Core, Harvard Medical School, Boston, MA 02115.

[2]Present address: Department of Genetics, Harvard Medical School, Boston, MA 02115.

[3]Present address: Institute of Education, Liberal Arts and Sciences Division, Tokyo Medical and Dental University, Tokyo 113-8510, Japan.

[4]Present address: Department of Cell Biology, Harvard Medical School, Boston, MA 02115.

[5]Present address: HHMI, Chevy Chase, MD 20815.

[6]Present address: Faculty of Biochemistry, Biophysics and Biotechnology, Jagiellonian University, Krakow, 30-387 Poland.

[7]To whom correspondence may be addressed. Email: extavour@oeb.harvard.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2218506120/-/DCSupplemental.

Published May 16, 2023.

protein (Creb), a transcription factor with well-described conserved roles in long-term memory across metazoans (18). Our data demonstrate a somatic role for *osk* in the nervous system and shed light on how a novel gene may acquire critical roles by integrating with preexisting gene regulatory systems comprising older, conserved genes.

Results

***osk* Is Expressed in Adult Neuroblasts of the Mushroom Body.** We previously showed that neuroblasts in the cricket embryo express *osk*, *vasa,* and *piwi*, and that *osk* is required for correct neuroblast division and embryonic nervous system morphology (15). Interestingly, in many insects, including crickets, a subset of embryonic neuroblasts persists in the brain throughout adulthood and continuously gives rise to new neurons called Kenyon cells that comprise the mushroom body (19–21). This contrasts with flies like *D. melanogaster*, in which neuroblasts die prior to adulthood (22), and in which adult brains are thus essentially devoid of neurogenesis (23) (although there are reports of potential stem cells in adult *D. melanogaster* brains (23, 24), which may be damage-dependent rather than homeostatic in function (25), and which remain controversial (26)).

Given the role of *osk* in embryonic neuroblasts of crickets, we asked whether *osk* also plays a role in the adult mushroom body neuroblasts (MBNBs). We used in situ hybridization to examine *osk* expression in the adult brain and found expression in a cluster of cells with the large, round nuclei and diffuse chromatin characteristic of stem cells, at the apex of each of the two lobes of the mushroom body, consistent with descriptions of adult neuroblasts in orthopterans (Fig. 1*A*). EdU colocalization (Fig. 1*B*) confirmed the identity of these cells as neuroblasts, the only proliferative cells in the adult brain (27). We also found that MBNBs express high levels of Vasa and Piwi proteins (Fig. 1*E*).

Previous research has shown that MBNBs play an important role in long-term olfactory memory formation in Orthoptera (28). Scotto-Lomassese (28) found that ablation of MBNBs using irradiation led to a dramatic reduction in olfactory, but not visual, learning after 24 and 48 h, suggesting that newborn mushroom body neurons produced by those neuroblasts play a role in forming new olfactory memories. We therefore sought to test whether *osk*, expressed specifically in MBNBs, functions in these cells in the context of long-term memory formation.

We first tested whether *osk* regulates the proliferation or survival of adult MBNBs. Using an established technique for systemic RNAi in the adult cricket brain (29), we injected double-stranded *osk* RNA (dsRNA) into the head capsule and confirmed the efficiency of *osk* knockdown via quantitative PCR (Fig. 2*D* and *SI Appendix*, Table S1) and small RNA profiling of osk^{RNAi} brains (*SI Appendix*, Tables S2–S4 and Fig. S1). osk^{RNAi} adult mushroom bodies showed no gross anatomical defects relative to controls (Fig. 1; and *SI Appendix*, Fig. S1). Moreover, neither the total number of neuroblasts ($P > 0.05$), nor the number of neuroblasts undergoing mitosis as revealed by EdU labeling ($P > 0.05$), was statistically significantly different between osk^{RNAi} adult brains and controls (Fig. 1 *C* and *D*). We stained osk^{RNAi} and control brains with cleaved caspase-3, a marker for apoptosis, and did not observe any evidence of cell death (*SI Appendix*, Fig. S2*A*). We noted that one described role for *piwi* in the *Drosophila* germ line is to prevent DNA damage caused by transposon mobilization (30). However, we observed no detectable increase in γH2A staining, a marker for DNA damage, in osk^{RNAi} brains (*SI Appendix*, Fig. S2*B*). These data suggest that *osk* is not required for the proliferation, survival, or genomic integrity of adult neuroblasts. However, the specific expression of *osk*, Piwi, and Vasa in the MBNBs suggested that some or all of these genes could play a role related to memory or learning.

***osk* RNAi Impairs Long-Term, but Not Short-Term, Memory.** The mushroom body is the anatomical substrate for olfactory memory and learning in insects (27, 31, 32), and ablation of the mushroom body or of the adult MBNBs impairs these processes (28, 33, 34). Based on previous observations that MBNBs play a role in long-term olfactory memory formation in crickets (28), we hypothesized that *osk* might play a role in this process. To test this hypothesis, we assessed the memory of osk^{RNAi} adult male crickets at 1 h ("short-term memory") and 1 d ("long-term memory") post-training using well-established cricket olfactory behavior assays as previously described (17). Briefly, the crickets were injected with double-stranded RNA against the gene of interest (*osk*) or a control gene (*DsRed*) (35), and then subjected to odor preference tests (allowed to freely visit peppermint and vanilla odor sources, quantifying the time spent at each odor source) and conditioning trials (peppermint odor was paired with a water reward, and relative preference for this rewarded odor was compared before and after the conditioning) (17).

In control crickets (injected with dsRNA targeting *DsRed* (35)), four training sessions led to a significant ($P < 0.05$) short-term preference for the rewarded odor (peppermint) at 1 h after training (short-term; Fig. 2*A*). Trained control crickets retained this learned preference ($P < 0.01$) even at 1 d after training (Fig. 2*A*, "*DsRed*"), demonstrating that long-term memory formation is intact in these controls. However, although osk^{RNAi} crickets formed and retained memory for the rewarded odor by 1 h after training (short term; Fig. 2*A osk* dsRNA #1, $P < 0.001$), this memory was lost by 1 d after training (long term; Fig. 2*A osk* dsRNA #1, $P > 0.05$), indicating a specific impairment of long-term memory formation. These results were reproducible in a second experiment using a nonoverlapping fragment of *osk* dsRNA (Fig. 2*A*, *osk* dsRNA #2; $P < 0.05$ for short term, $P > 0.05$ for long term), suggesting that the impact was specific to *osk* knockdown. The efficacy of the knockdown was confirmed via qPCR (Fig. 2*D* and *SI Appendix*, Table S1) and small RNA sequencing (*SI Appendix*, Fig. S1 and Tables S1 and S2), indicating that *osk* is required for cricket long-term memory.

Since both Piwi and Vasa were coexpressed with *osk* in cricket MBNBs (15) (Fig. 1*E*), we also assessed the role of these two genes in olfactory memory. We found that *piwi* (Fig. 2*B*, $P < 0.001$) but not *vasa* (*SI Appendix*, Fig. S3*A*, $P > 0.05$; *SI Appendix*, Fig. S3*B*) was also required for cricket long-term memory. qPCR analyses showed that *osk* RNAi led to a notable decrease in *piwi* transcript levels (Fig. 2*D*), suggesting that *osk* positively regulates *piwi* in the cricket brain. However, *osk* transcript levels remained unaffected in $piwi^{RNAi}$ animals (Fig. 2*E*). Consistent with the phenotype of the single gene knockdowns, $osk^{RNAi}/piwi^{RNAi}$ double knockdown adults also showed a long-term memory impairment phenotype (Fig. 2*C*, $P < 0.01$). Thus, *osk* and *piwi* do not globally disrupt olfaction, learning, or short-term memory formation, but are required for consolidation of long-term memory in this species.

***osk* and *piwi* Positively Regulate the Nuclear Transcription Factor CrebA.** To understand how a novel gene like *osk* might have gained a role in an ancient animal function like long-term memory consolidation, we investigated the hypothesis that it might interact with conserved regulators of animal memory. To test this hypothesis, we asked whether we could detect a functional or regulatory interaction between *osk* and a highly conserved transcription factor with well-documented roles in long-term memory formation across animals, Creb (18). We first identified putative *Creb* orthologs in

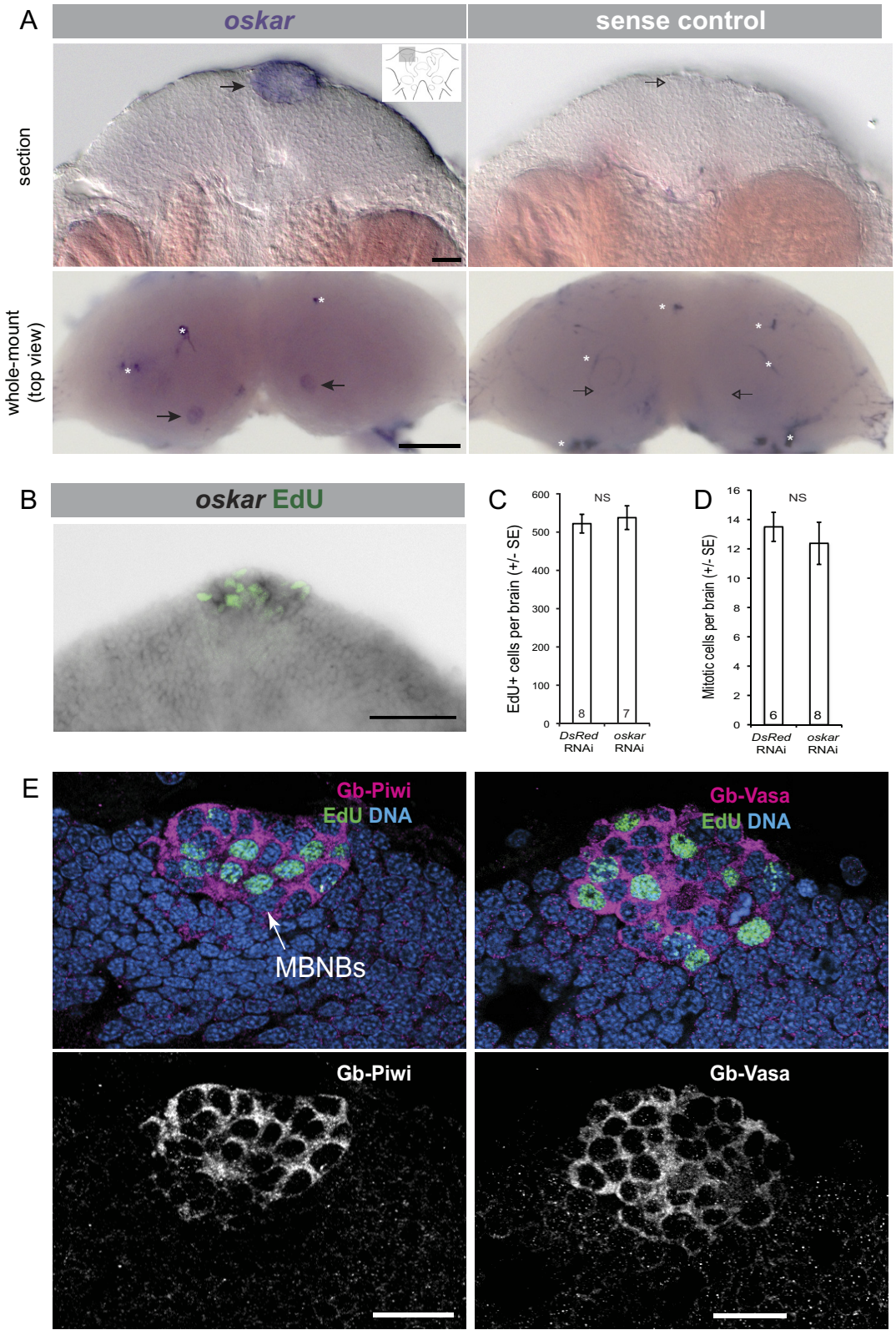

Fig. 1. *oskar*, Piwi, and Vasa are expressed in *G. bimaculatus* adult MBNBs. (*A*) In situ hybridization on adult *G. bimaculatus* brains detects *osk* transcripts in the cells of the mushroom body (arrows). *Inset* in top right corner of top left panel shows the overall structure of the adult brain; shaded box indicates a single mushroom body lobe, corresponding to the region shown in micrographs in top row. Bottom row: dorsal views of both mushroom body lobes, indicating *oskar* expression revealed by in situ hybridization (purple) in neuroblast clusters (arrows). White asterisks indicate nonspecific binding of probe to tracheal remnants in the brain. (*B*) EdU labeling (green) of the adult brain shows that *osk*-expressing cells (gray) are mitotically active, consistent with their identity as neuroblasts. (*C*) Quantification of EdU-positive cells shows no significant difference between osk^{RNAi} and control brains ($P < 0.05$). (*D*) Quantification of total number of mitotically active cells shows no significant difference between osk^{RNAi} and control brains ($P > 0.05$). Numbers within bars indicate sample sizes and NS = no significant difference in (*C*) and (*D*). (*E*) Detection of Vasa & Piwi proteins (magenta) in adult MBNBs. Scale bars, 50 μm in top panels of (*A*) and (*B*) and in (*E*), and 200 μM in bottom panels of (*A*).

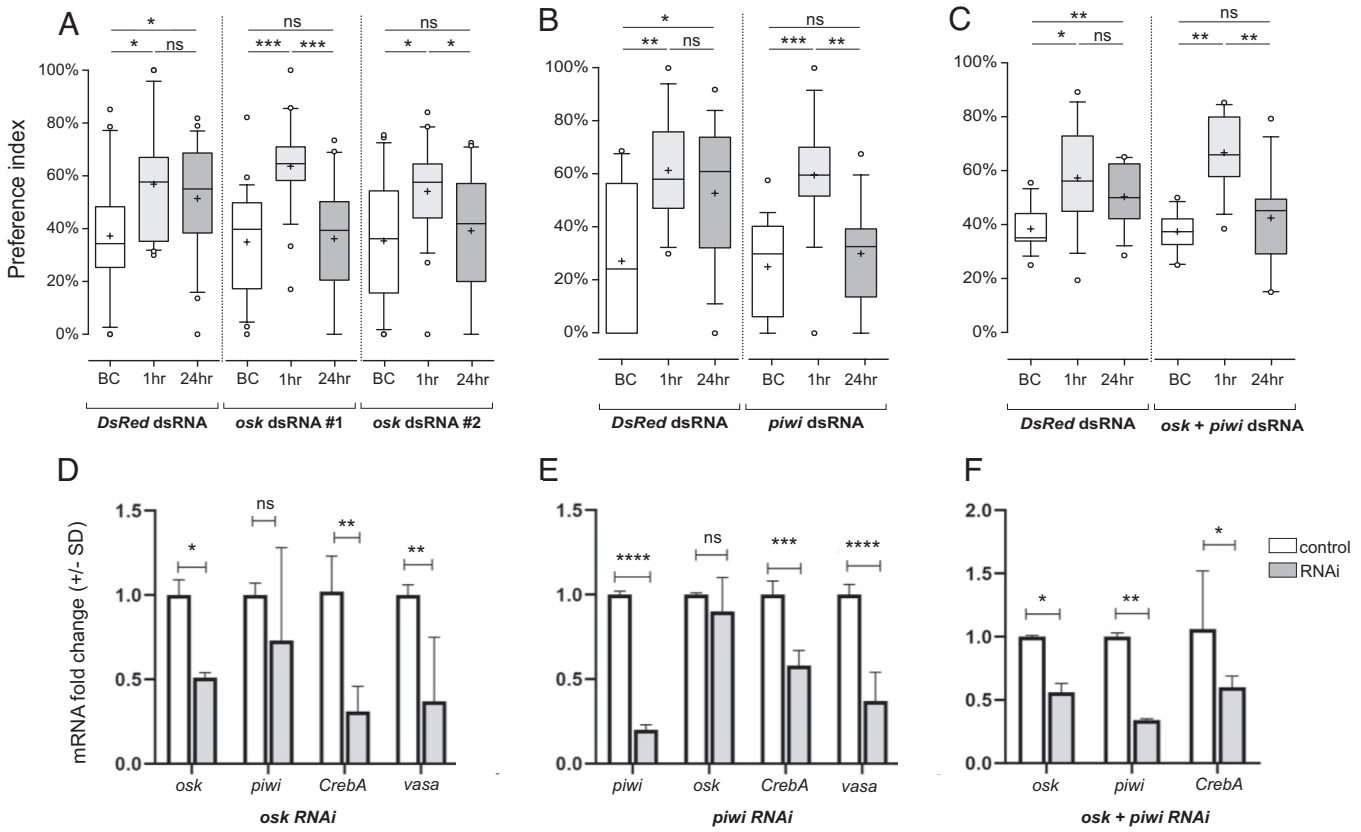

Fig. 2. *oskar*^{RNAi} and *piwi*^{RNAi} impair cricket long-term memory. (*A*) Results of olfactory memory assay in *osk*^{RNAi} animals. (*B*) Results of olfactory memory assay in *piwi*^{RNAi} animals. (*C*) Results of olfactory memory assay in *osk*^{RNAi}/*piwi*^{RNAi} double knockdown animals. For each assay, relative preference between the rewarded odor (peppermint) and control odor (vanilla) was tested before conditioning (BC), 1 h post-training (1 h), and 1 d post-training (24 h) for *DsRed*^{RNAi} controls and for *osk*^{RNAi} (using two different nonoverlapping *osk* fragments #1 and #2), *piwi*^{RNAi}, and *osk*^{RNAi}/*piwi*^{RNAi}. Boxes represent the first and third quartiles surrounding the median (middle line). Whiskers extend to values within 1.5× of interquartile range. Wilcoxon's test was used for comparison of preference before and after conditioning. For multiple comparisons, the Holm method was used to adjust the significance level (*$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, n.s. = not statistically significant). (*D–F*). qPCR results showing the extent of downregulation of different *G. bimaculatus* genes in *osk*^{RNAi}, *piwi*^{RNAi}, and *osk*^{RNAi}/*piwi*^{RNAi} backgrounds. Effectiveness of RNAi per background is also shown in each case. Data are plotted as mRNA fold change (± SD) based on the ΔΔCt method (*$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, ****$P < 0.0001$, n.s. = not statistically significant).

the *G. bimaculatus* genome (36) using a combination of basic local alignment search tool searches and phylogenetic analyses (Fig. 3*A*, and *SI Appendix*, Table S5). These analyses yielded two high-confidence *Creb* orthologs, which we called *CrebA* and *CrebB* based on their closest *D. melanogaster Creb* gene relative (Fig. 3*A*). Analysis of previously generated transcriptomes (37) showed that both genes are expressed in adult cricket brains (*SI Appendix*, Table S6). We performed *CrebA* and *CrebB* RNAi experiments and discovered that *CrebA* (but not *CrebB*; *SI Appendix*, Fig. S3*C*) was required for long-term memory in crickets (Fig. 3*B*; $P < 0.003$ and $P < 0.001$ for dsRNA#1 and dsRNA#2, respectively; *SI Appendix*, Fig. S3*C*). Using qPCR, we then asked whether transcript levels of this memory regulator were altered in *osk* or *piwi* knockdown conditions and found consistent downregulation of *CrebA* transcript levels in both single and double RNAi backgrounds (Fig. 2 *D–F*). In contrast, and consistent with the observation that *vasa*^{RNAi} had no long-term memory impact (*SI Appendix*, Fig. S3*A*), qPCR revealed no reduction of *CrebA* transcripts in *vasa*^{RNAi} conditions (*SI Appendix*, Fig. S3*B*). This suggests that the long-term memory defects observed in *osk*^{RNAi} and *piwi*^{RNAi} conditions (Fig. 2 *D–F*) are due to a downregulation of *CrebA* in these animals.

***osk* and *piwi* Are Regulated by CrebA.** Creb proteins are transcription factors that bind cyclic adenosine monophosphate response element (CRE) binding sites within the regulatory regions of target genes to initiate transcription (18) (Fig. 4*A*). Since target gene transcription and new protein synthesis is crucial for long-term memory formation, and given the similarity in long-term memory phenotypes of *osk*^{RNAi}, *piwi*^{RNAi}, and *CrebA*^{RNAi} animals, we asked whether *osk* or *piwi* might also be *Creb* target genes in this cricket. qPCR revealed that transcript levels of both *osk* and *piwi* are significantly decreased in *CrebA*^{RNAi} conditions (Fig. 4*B*), suggesting that *osk*/*piwi* and CrebA may interact in a positive feedback loop to regulate each other's transcript levels. To evaluate the possibility that *osk* or *piwi* might be direct transcriptional targets of CrebA, we examined the genomic sequences within 10 kb upstream of both loci and found two bioinformatically predicted CRE binding sites within the 6 kb upstream of the transcription start sites for *osk* and *piwi* (Fig. 4*C* and *SI Appendix*, *Supplementary File* 1). These predicted CRE binding sites were found twice as frequently as we would expect to find such sequences in a randomly generated sequence of this length (*Materials and Methods*). Electrophoretic mobility shift assays showed that protein(s) within the adult cricket brain bind specifically to the predicted CRE sites of *osk* (Fig. 4*D* and *SI Appendix*, Table S7). Given the current lack of species-specific CrebA reagents for this cricket species, we cannot rule out the interpretation that a protein(s) other than CrebA present in the adult cricket brain is causing the observed mobility shift by binding the predicted CRE sites of *osk*. However, given our functional data indicating that RNAi against *osk*, *piwi,* and *CrebA* all yield long-term memory defects (Figs. 2 and 3*B*), that *osk* and *piwi* transcript levels are reduced in *CrebA*^{RNAi} brains (Fig. 4*B*), and that *osk* and *piwi* genomic loci contain predicted CRE binding sites (Fig. 4*C*),

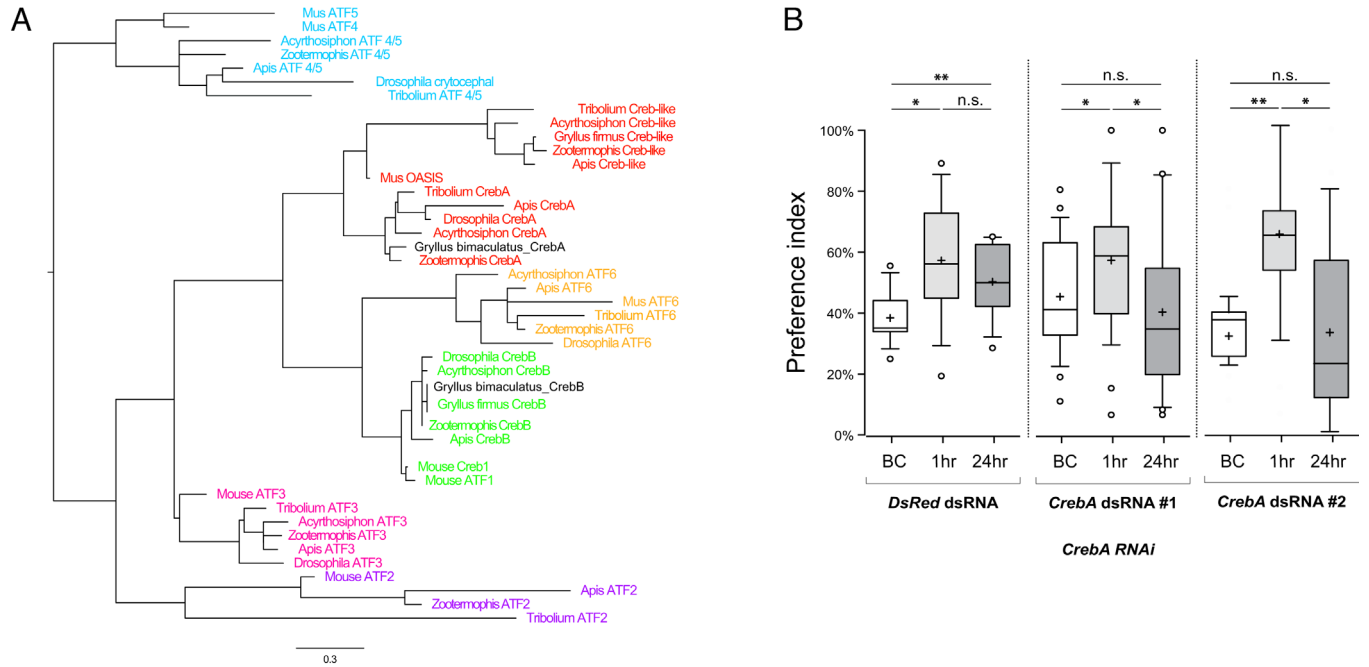

Fig. 3. Cricket *CrebA* is required for cricket long-term memory. (*A*) *Creb/ATF* family member orthologs in mouse and insects (genus name shown) (*SI Appendix*, Table S5) were used to construct a *Creb* phylogenetic tree to infer the evolutionary relationships between mammalian *Creb* proteins and their insect counterparts. *G. bimaculatus CrebA* and *CrebB* are indicated in black in the tree. (*B*) *CrebA*^{RNAi} impairs long-term memory formation in crickets. Relative preference between the rewarded odor (peppermint) and control odor (vanilla) was tested before conditioning (BC), 1 h post-training (1 h), and 1 d post training (24 h) for *DsRed*^{RNAi} controls and *CrebA*^{RNAi} (using two different nonoverlapping *CrebA* fragments #1 and #2 for independent confirmation). Boxes represent the first and third quartiles surrounding the median (middle line). Whiskers extend to values within 1.5× of interquartile range. Wilcoxon's test was used for comparison of preference before and after conditioning. For multiple comparisons, the Holm method was used to adjust the significance level (*P < 0.05, **P < 0.01, n.s. = not statistically significant). n = 9 for *CrebA* and n = 10 for *DsRed*.

the results of our gel shift assay are consistent with the hypothesis that cricket *osk* is a direct transcriptional target of CrebA.

Discussion

We have discovered a role for *oskar* in the adult cricket brain (Fig. 1). We have shown that *osk,* Piwi, and Vasa are coexpressed in MBNBs (Fig. 1 *A* and *B*), a population of neural stem cells required for long-term olfactory memory formation (28), and that knockdown of *osk* and *piwi* disrupts olfactory long-term memory formation. The precise role that the MBNBs play in memory formation remains unknown, as does the molecular role of Osk in these cells. In *D. melanogaster*, where there are no adult neural stem cells in the mushroom body, olfactory long-term memory

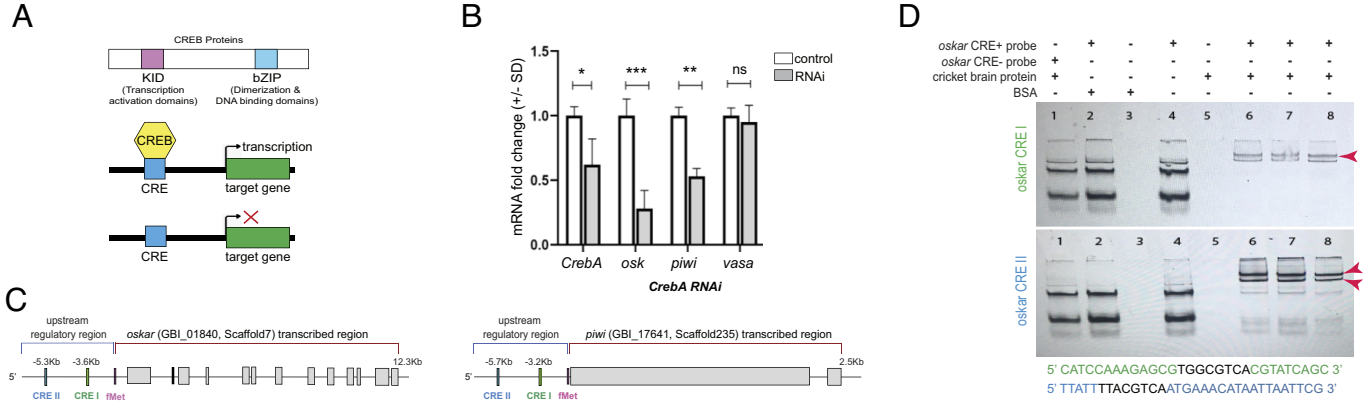

Fig. 4. Cricket CrebA regulates *oskar*. (*A*) Schematic diagram of the transcription factor cAMP response element binding protein (Creb) protein (top *schematic*) displaying only the two domains relevant to this study, the kinase-inducible domain (KID) that can facilitate kinase-inducible transcription activation, and the basic leucine zipper domain (bZIP) that is important for dimerization and DNA binding. Creb protein binding to the CRE, a sequence present in the promoter regions of many cellular genes can increase (middle schematic) or decrease (bottom schematic) transcription of target genes. (*B*) qPCR results showing the relative expression levels of *G. bimaculatus osk, piwi*, and *vasa* in *CrebA*^{RNAi} knockdown conditions. The extent of *CrebA* transcript decrease is also shown to assess the efficiency of RNAi knockdown. Data are plotted as mRNA fold change (± SD) based on the ΔΔCt method (*P < 0.05, **P < 0.01, ***P < 0.001, *P < 0.0001 n.s. = not statistically significant). (*C*) Schematic of *G. bimaculatus osk* and *piwi* genes showing exons (depicted as rectangular grey boxes) along with their presumptive upstream regulatory regions, each containing two predicted CRE sites, which we call CRE I (~3.6 Kb and ~3.2 Kb upstream of the predicted transcription start site for *osk* and *piwi,* respectively, marked by fMet) and CRE II (~5.3 Kb and ~5.7 Kb upstream of predicted fMet for *osk* and *piwi,* respectively). (*D*) Eletrophoretic mobility shift assay to detect possible Creb binding to *osk's* CRE I (top, in green) and CRE II (bottom, in blue). "*oskar* CRE+ probe" indicates predicted CRE site-containing probe; "*oskar* CRE- probe" indicates probe without predicted CRE sites; "cricket brain protein" indicates *G. bimaculatus* brain protein extract; "BSA" indicates 1% bovine serum albumin as a nonspecific protein control. The complete sequences of the EMSA probes used for CRE I and II experiments are indicated in green and blue text underneath the gel image, with black representing the predicted CRE sequence. Shift is marked with red arrowheads.

requires the ~2,500 differentiated Kenyon cell neurons of the mushroom body (38), which respond with high selectivity to a small number of stimuli, allowing the mushroom body to house an explicit representation of a large number of olfactory cues (38, 39). Specific olfactory stimuli are associated with learned behavioral responses via specific sets of neurons connecting the mushroom body to other brain regions in a protein synthesis-dependent fashion, to form long-term memories (9, 40). Thus, one possibility is that adult-born Kenyon cells in *G. bimaculatus* (and other insects that display adult neurogenesis in the mushroom body) are recruited into an existing circuit and allow for a constantly increasing repertoire of olfactory associations. Our results suggest that *osk* could play a role in this process, as *osk* RNAi disrupts long-term memory. We note that of the two mammalian brain regions known to undergo adult neurogenesis, one (the subventricular zone) contributes to the olfactory bulb, and neurogenesis in this region is involved in olfactory memory (41).

Given that adult *D. melanogaster* lacks the MBNBs seen in *G. bimaculatus* (42), a straightforward test for a directly comparable *osk* function in this fruit fly is not possible. However, although *D. melanogaster* mushroom body stem cells are absent in adults, analogous MBNBs remain mitotically active late into pupal development (43). Thus, it will be interesting to test whether *osk* functions in these neuroblasts during larval and/or pupal stages. We note that an insertion of an enhancer trap transposable element over 3 kb upstream of the *osk* transcription start site was recovered in an insertional mutagenesis screen for long-term memory in *D. melanogaster* (8). However, this insertion has not been confirmed as compromising the sequence or function of the *osk* locus, nor has *osk* been tested directly to confirm a potential role in *D. melanogaster* learning or memory.

Although *D. melanogaster* lacks adult MBNBs, it is possible that *osk* could function in fruit fly olfactory long-term memory in a neuroblast-independent manner. A recent study of the mushroom body output neurons has suggested that long-term memory involves the activity-dependent derepression of mRNAs localized to granules containing Pumilio, Staufen, and Orb (oo18 RNA-binding protein) proteins (9). Given that Osk nucleates similar granules containing these proteins in the *Drosophila* oocyte (44, 45), and Osk's ability to nucleate phase-transitioned granules in *D. melanogaster* cells (46), it would be interesting to test whether Osk is involved in the formation and/or activity of these granules in the brain. Given our recent observation of highly similar molecular interactions of conserved molecules in animal germ lines and neural cells (14), future studies could test whether additional genes traditionally known as "germ line" genes other than *vasa* and *piwi*, for example *staufen* (47) and *tudor* (48), also function in *G. bimaculatus* adult neuroblasts, which would suggest that *osk* acts with conserved molecular partners in different cellular contexts. Because mushroom bodies derived from neuroblasts are a conserved arthropod brain structure across and beyond insects (49–52), it seems unlikely that *oskar* played a role in the evolution of insect mushroom body neurogenesis per se.

Both germ cells and neuroblasts are stem cells that give rise to highly specialized daughter cells while remaining proliferative for long periods of time. Thus, the original role of *osk* in both cell types could conceivably be related to stem cell maintenance and/or asymmetric division. Indeed, many different highly conserved "germ line genes" including *vasa*, *nanos*, and *piwi* are found in a variety of multipotent cells in diverse animals (14, 53), raising the possibility that such genes were involved in establishing multipotency rather than specifying germ cell fate per se. In our previous examination of the distribution of *osk* orthologs across insects, we observed that crickets are not the only insects reported to express *osk* in the brain, as *osk* transcripts are detected in transcriptomes from the brains of cockroach, wasp, and beetle species as well (16). This is consistent with the hypothesis that *osk* played an ancient neural role of some kind in insects. We previously showed evidence supporting the hypothesis that evolved changes in the biophysical characteristics of Oskar protein may have driven the evolution of a novel mechanism of germ line specification in the holometabolous insects (16). A broader understanding of the putative ancestral and derived function(s) of *osk* thus requires additional studies of phylogenetically diverse insects, as well as further detailed biochemical analysis in the context of *Drosophila* germ cells and neurons.

Our results provide an example of how novel genes may find stable homes in preexisting genetic regulatory circuits. In the case of *osk*, we hypothesize that by evolving or acquiring binding sites responsive to the conserved transcription factor Creb, *osk* may have gained expression in the brain, opening the door for potential participation in neural roles. An alternative hypothesis to de novo evolution of CRE sites to explain *osk*'s expression in the cricket brain is that *osk* inherited CRE binding sites from an ancestral sequence that contributed to *osk*'s genesis. Our previous work suggested that *osk* arose through a fusion of a eukaryotic LOTUS domain in the 5′ position, coding for *osk*'s N-terminal LOTUS domain, and a prokaryotic SGNH hydrolase-like domain in the 3′ position, coding for *osk*'s C-terminal OSK domain (54). In this scenario, a preexisting LOTUS domain-containing gene could have donated not just its LOTUS domain, but also some upstream 5′ regulatory sequences, including CRE sites and/or neural expression elements, to *osk*. This hypothesis might predict that extant LOTUS domain-containing genes might display one or both of CRE-binding sites, or expression in the brain. To test this hypothesis, we searched the *G. bimaculatus* genome for LOTUS domain-containing genes and identified five such genes (*SI Appendix*, *Supplementary File* 2). These were *osk*, *Tdrd5*, *Tdrd7*, *limkain b1*, and an uncharacterized gene with annotation ID GBI_15344 (*SI Appendix*, *Supplementary File* 2). In transcriptomes previously generated from the adult brain (55), we detected levels of all four non-*oskar* LOTUS domain-containing genes at levels at least as high as those detected for *osk* (*SI Appendix*, Fig. S4 and Table S8). Moreover, in the 10 kb upstream of the first predicted codon of these genes, we detected a putative CRE binding site (*SI Appendix*, Table S9). With the caveat that our transcriptomes do not provide spatial or cell-type resolution for the expression data, both of these findings are consistent with the hypothesis that whatever eukaryotic LOTUS domain-containing sequence was the ancestor of *osk's* LOTUS domain, it also contributed one or both of CRE-responsive or brain-expressed upstream regulatory sequences to *osk*. Future studies will be needed to elucidate the molecular mechanisms of *osk* gene products in the cricket brain, and specifically in learning and memory.

We further speculate that the biophysical properties of Osk protein that make it effective at sequestering RNAs and participating in translational control in the germ line (56–58) may have been advantageous in promoting the rapid translation needed for the synaptic plasticity that underlies learning and memory. These include Osk's ability to form phase-transitioned condensates (46, 59), its regions of high predicted disorder (16, 46, 59), and its ability to achieve and maintain asymmetric subcellular localization, all of which are well known in the germ line and may have provided a selective advantage to Osk in the context of promoting neuronal function.

Materials and Methods

G. bimaculatus husbandry, in situ hybridization, immunostaining, olfactory learning assays, RNAi, and qPCR were performed as previously described. See *SI Appendix, Extended Materials and Methods* for references and detailed protocols.

EdU Assay. Cell proliferation was assayed using the Click-iT EdU Alexa 488 kit (Life Technologies, Cat# C10637). Crickets were injected with 10 to 15 μL EdU either into the abdomen or into the head capsule through the median ocellus (both methods successfully labeled dividing neuroblasts), and brains were dissected to visualize EdU incorporation 4 h postinjection. The brains were dissected and desheathed in ice-cold 1× phosphate buffered saline. Calyces were removed with a microscalpel and incubated in 0.1M citric acid for 15 to 30 min on a poly-lysinated slide (Sigma Aldrich, Cat. No. P8920-100ML). The calyces were then spread into a monolayer by adding a Sigma cote-covered coverslip, and the entire slide was flash-frozen in liquid nitrogen. The coverslip was removed, leaving the mushroom body monolayer on the slide. Slides were air-dried and were then fixed for 15 min in 4% paraformaldehyde. EdU detection was then carried out following manufacturer's instructions. EdU-positive cells were photographed under epifluorescence on a Zeiss AxioImager Z.1 compound microscope using Zen and manually quantified in ImageJ. For any mushroom body where the EdU-positive cluster of cells was damaged or destroyed during preparation, that sample was discarded and not included in the analysis. For tissue double stained to visualize transcripts and EdU incorporation simultaneously, in situ hybridization was conducted before the visualization of incorporated EdU. AxioImager Z.1, LSM 780, or LSM 880 confocal microscopes (Zeiss) were used for microscopy, driven by AxioVision or Zen (Zeiss).

Construction, Sequencing, and Analysis of Small RNA Libraries from *G. bimaculatus* Adult Brains. Unmated adult male crickets within 1 week of their final molt to adulthood were injected with dsRNA as described above (see RNA interference). At 48 h after injection, brains (*SI Appendix*, Table S2) were dissected in ice-cold 1× phosphate buffered saline and transferred into Trizol, following which total RNA was extracted from them following manufacturer's protocols. Next, RNA was size selected for 18 to 30 nt size range after denaturing polyacrylamide gel electrophoresis. A 2S rRNA specific oligo was used for 2S rRNA depletion. The small RNAs were ligated at the 3′ and 5′ ends by the respective adapters and purified by denaturing polyacrylamide gel electrophoresis after each ligation. PCR was performed after reverse transcription. The PCR product was gel purified from an agarose gel to obtain the final library. The libraries were sequenced using Illumina NextSeq500 1×75bp. The resulting data (*SI Appendix*, Table S3) were uploaded onto the National Center for Biotechnology Information Sequence Read Archive (NCBI SRA) database and are publicly available under the BioProject ID PRJNA837371 (60).

Identification of *G. bimaculatus* Creb Genes and Construction of Creb Phylogenetic Tree. Putative orthologs of *Creb/ATF* family members from several animal species were initially identified by basic local alignment search tool (BLAST) searches (*SI Appendix*, Table S5) and then downloaded from NCBI. These sequences were then used to search for putative *G. bimaculatus CrebA* orthologs in the G. *bimaculatus* genome (36). All identified sequences were then aligned with MAFFT (v 7.510) (61). A maximum likelihood tree was created in RAxML using the PROTGAMMAWAG model (62) and plotted with the FigTree package v1.4.4 (http://tree.bio.ed.ac.uk/software/figtree) (Fig. 3*B*).

Bioinformatic Prediction of CRE Sites in *osk* and *piwi* Upstream Regulatory Regions. A position frequency matrix (PFM) for the full CRE octameric palindrome was obtained from the JASPAR database (an open-source database for transcription factor binding sites (63) (*SI Appendix*, *Supplementary File* 1). In addition to CRE, PFMs for the TATA box were also obtained from the same database. We included TATA box proximity among our search criteria for putative CRE sites, since TATA boxes are often a feature of functional promoters, and functional promoter-proximal CRE sites are reported as often occupied by Creb. These raw PFM data (*SI Appendix*, *Supplementary File* 1) were then used as an input in Find Individual Motif Occurrence (FIMO) in the MEME suite [a motif-based sequence analysis tool (64)], and up to 10 Kb of the genome sequence upstream of the predicted transcription start site for each of the *G. bimaculatus oskar*, *piwi*, and *vasa* was scanned for the presence of the CRE and TATA motifs using the annotated *G. bimaculatus* genome (36), and using $P < 0.0001$ as the stringency criteria. For comparison, *G. bimaculatus beta actin*, *alpha tubulin,* and *FGFR* loci were subjected to the same analyses (also with $P < 0.0001$ as stringency criteria) to assess the possibility that any randomly chosen *G. bimaculatus* gene would be predicted to have CRE sites in the 10Kb region upstream of their transcription start site using this method (*SI Appendix*, *Supplementary File* 2). We found that for the latter three genes, there were no CRE predictions in their upstream regions (up to 10 Kb from the transcription start site). Further, we bioinformatically generated one thousand 10 Kb long DNA fragments of random sequence using the "random DNA sequence" tool in the Sequence Manipulation Suite (65) and then tested them for CRE prediction. Our results indicate that a CRE site is expected to occur in a randomly generated sequence at a frequency of ~1.8 CRE sites for every 10 Kb tested (*SI Appendix*, *Supplementary Files* 1 and 2).

Bioinformatic Analysis of LOTUS Domain-Containing Genes in the *G. bimaculatus* Genome. We searched the annotated *G. bimaculatus* genome (36) for genes whose protein product was predicted to contain Pfam motif PF12872, corresponding to the LOTUS domain. This search retrieved five genes (GBI_01840 "*oskar,*" GBI_13502 "*TDRD5/tejas*", GBI_15344 "*uncharacterized,*" GBI_15604 "*limkain b1,*" and GBI_03370 "*TDRD7/tapas*"). We assessed the expression of these genes in brains and gonads using previously published RNA-Seq libraries (55) available at NCBI (PRJNA564136). We analyzed the RNA-Seq data as in ref. 16, including removing adapters and reads shorter than 20 nucleotides with Cutadapt v3.4 (66) and quantifying the gene expression in transcripts per million with RSEM v1.2.29 (67), using STAR v2.7.0e1 (68) as read mapper against the *G. bimaculatus* genome (36) (*SI Appendix*, Table S8). For each of these genes, we retrieved and searched the 10Kb upstream of the first codon annotated for CRE sites as described above in "*Bioinformatic Prediction of CRE Sites in* osk *and* piwi *Upstream Regulatory Region*" (*SI Appendix*, Table S9).

PCR Amplification, Sequence Confirmation, and Cloning of CRE Sites. Based on bioinformatic predictions of putative CRE sites, primers were designed in the upstream regulatory regions of *osk* (*SI Appendix*, Table S7 #1 and #2). Once both CRE sites were sequence confirmed by Sanger sequencing, the ~30bp fragments containing each CRE site were synthetically generated as duplexes (with 3′A overhangs) for use as Electrophoretic Mobility Shift Assay (EMSA) preprobes (*SI Appendix*, Table S7; CRE site in bold). The 3′A overhangs were then used to clone all EMSA preprobes into a pGEM-T easy vector following manufacturer's instructions (Promega, catalog number A1360) using One-Shot chemically competent TOP10 *E. coli* cells (Thermo-Fisher, catalog number C4040-06).

Generation of 5'Cy5-Labeled EMSA Probes and EMSA. Once cloned, pGEM-T easy specific duplex forward primer (5′Cy5-ACGTCGCATGCTCCCGGCCATG, reverse complement 5′Cy5-CATGGCCGGGAGCATGCGACGT) and reverse primer (5′Cy5-GTCGACCTGCAGGCGGCCGCGAATT, reverse complement 5-Cy5-AATTCGCGGCCG CCTGCAGGTCGAC) were designed with 5′Cy5 modifications to amplify inserts and generate fluorescently labeled double-stranded EMSA probes, using a two-step PCR program with the following conditions: (98 °C for 60 s (×1 cycle); 98 °C for 15 s followed by 72 °C for 30 s (×30 cycles); 72 °C for 5 min (×1 cycle) (*SI Appendix*, Table S7). The PCR product was loaded onto a 1% agarose gel, and the desired bands were gel eluted following IBI Scientific's PCR purification and gel elution kit (catalog number IB47030) in 30 μL water. A second round of PCR amplification following the conditions described above was performed using the eluted DNA from previous steps to increase probe yield. All steps starting with the first round of PCR were done in the dark to protect fluorescently labeled probes. Probe concentrations were measured using a Nanodrop and diluted to a final concentration of 40 fmol/probe for use in electrophoretic mobility shift assays (69). Twenty percent native polyacrylamide gele electrophoresis gels were used to study gel shifts. Gels were imaged using an Azure Sapphire Biomolecular Imager (VWR).

Nuclear Protein Extracts from Unmated Adult Male *G. bimaculatus* Brains. Brains were dissected from unmated *G. bimaculatus* males within 1 wk of their final molt that were anesthetized briefly on ice prior to dissection in 1× phosphate buffered saline. Nuclear protein extracts were prepared from dissected brains following manufacturer's instructions (Abcam Nuclear Extraction Kit, catalog number ab113474).

Data, Materials, and Software Availability. Short read next generation sequencing data have been deposited in NCBI SRA (PRJNA837371) (60).

ACKNOWLEDGMENTS. Thanks to Taro Mito (University of Tokushima, Japan) for advice on cricket brain dissection; Venkatesh Murthy (Harvard University) for allowing us to use his laboratory's vibratome; Elena Kramer and Min Ya (Harvard University) for help with polyacrylamide gel electrophoresis setup for electrophoretic mobility

shift assays; William E. Theurkauf (University of Massachusetts Chan Medical School) for guidance on cricket small RNA library preparation and sequencing, and for financial support of S.S.P.; members of the Extavour lab for discussion; and the Faculty of Arts and Sciences Bauer Molecular Biology Core Facility at Harvard University for Illumina Sequencing. This study was supported by NSF award #IOS-0817678, funds from Harvard University and from the Howard Hughes Medical Institute to C.G.E., a NSF Graduate Research Training Fellowship to B.E.-C., and a Grant-in-Aid for Scientific Research from the Ministry of Education, Science, Culture, Sports, and Technology of Japan (no. 21K19245) to M.M.

Author affiliations: [a]Department of Organismic and Evolutionary Biology, Faculty of Arts and Sciences, Harvard University, Cambridge, MA 02138; [b]Graduate School of Life Science, Hokkaido University, Sapporo 060-0810, Japan; [c]Faculty of Science, Hokkaido University, Sapporo 060-0810, Japan; [d]Research Center for Integrative Evolutionary Science, School of Advanced Sciences, Sokendai-Hayama, Kanagawa 240-0193, Japan; [e]Department of Molecular and Cellular Biology, Faculty of Arts and Sciences, Harvard University, Cambridge, MA 02138; [f]University of Massachusetts Chan Medical School, Program in Molecular Medicine, Worcester, MA 01655; and [g]HHMI, Chevy Chase, MD 20815

Author contributions: A.K., B.E.-C., G.Y., M.M., and C.G.E. designed research; A.K., B.E.-C., K.T., Y.M., Y.L., T.W., J.A.K., S.S.P., and G.Y. performed research; A.K., B.E.-C., G.Y., M.M., and C.G.E. analyzed data; and A.K., B.E.-C., and C.G.E. wrote the paper.

1. R. Lehmann, C. Nüsslein-Volhard, Abdominal segmentation, pole cell formation, and embryonic polarity require the localized activity of oskar, a maternal gene in Drosophila. *Cell* **47**, 141–152 (1986).
2. A. Ephrussi, L. K. Dickinson, R. Lehmann, oskar organizes the germ plasm and directs localization of the posterior determinant nanos. *Cell* **66**, 37–50 (1991).
3. J. Kim-Ha, J. L. Smith, P. M. Macdonald, oskar mRNA is localized to the posterior pole of the Drosophila oocyte. *Cell* **66**, 23–35 (1991).
4. B. Hay, L. Y. Jan, Y. N. Jan, Localization of vasa, a component of Drosophila polar granules, in maternal-effect mutants that alter embryonic anteroposterior polarity. *Development* **109**, 425–433 (1990).
5. P. F. Lasko, M. Ashburner, Posterior localization of vasa protein correlates with, but is not sufficient for, pole cell development. *Gene Dev.* **4**, 905–921 (1990).
6. J. R. Jones, P. M. Macdonald, Oskar controls morphology of polar granules and nuclear bodies in Drosophila. *Development* **134**, 233–236 (2007).
7. R. Lehmann, Germ plasm biogenesis - An oskar-centric perspective. *Curr. Top Dev. Biol.* **116**, 679–707 (2016).
8. J. Dubnau *et al.*, The staufen/pumilio pathway is involved in Drosophila long-term memory. *Curr. Biol.* **13**, 286–296 (2003).
9. T.-P. Pai *et al.*, Drosophila ORB protein in two mushroom body output neurons is necessary for long-term memory formation. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 7898–7903 (2013).
10. P. N. Perrat *et al.*, Transposition-driven genomic heterogeneity in the Drosophila brain. *Science* **340**, 91–95 (2013).
11. R. P. Wharton, J. Sonoda, T. Lee, M. Patterson, Y. Murata, The pumilio RNA-binding domain is also a translational regulator. *Mol. Cell* **1**, 863–872 (1998).
12. X. Xu, J. L. Brechbiel, E. R. Gavis, Dynein-dependent transport of nanos RNA in Drosophila sensory neurons requires rumpelstiltskin and the germ plasm organizer oskar. *J. Neurosci.* **33**, 14791–14800 (2013).
13. B. Ye *et al.*, nanos and pumilio are essential for dendrite morphogenesis in Drosophila peripheral neurons. *Curr. Biol.* **14**, 314–321 (2004).
14. A. Kulkarni, D. H. Lopez, C. G. Extavour, Shared cell biological functions may underlie pleiotropy of molecular interactions in the germ lines and nervous systems of animals. *Front. Ecol. Evol.* **8**, 215 (2020).
15. B. Ewen-Campen, J. R. Srouji, E. E. Schwager, C. G. Extavour, oskar predates the evolution of germ plasm in insects. *Curr. Biol.* **22**, 2278–2283 (2012).
16. L. Blondel, S. Besse, E. L. Rivard, G. Ylla, C. G. Extavour, Evolution of a cytoplasmic determinant: Evidence for the biochemical basis of functional evolution of the novel germ line regulator oskar. *Mol. Biol. Evol.* **38**, 5491–5513 (2021).
17. Y. Matsumoto, M. Mizunami, Context-dependent olfactory learning in an insect. *Learn Memory* **11**, 288–293 (2004).
18. A. J. Silva, J. H. Kogan, P. W. Frankland, S. Kida, CREB and memory. *Neuroscience* **21**, 127–148 (1998).
19. S. M. Farris, I. Sinakevitch, Development and evolution of the insect mushroom bodies: Towards the understanding of conserved developmental mechanisms in a higher brain center. *Arthropod Struct. Dev.* **32**, 79–101 (2003).
20. J. Malaterre *et al.*, Development of cricket mushroom bodies. *J. Comp. Neurol.* **452**, 215–227 (2002).
21. R. Urbach, G. M. Technau, Early steps in building the insect brain: Neuroblast formation and segmental patterning in the developing brain of different insect species. *Arthropod Struct. Dev.* **32**, 103–123 (2003).
22. F. Pinto-Teixeira, N. Konstantinides, C. Desplan, Programmed cell death acts at different stages of Drosophila neurodevelopment to shape the central nervous system. *Febs Lett.* **590**, 2435–2453 (2016).
23. A. R. Simões, C. Rhiner, A cold-blooded view on adult neurogenesis. *Front. Neurosci-switz* **11**, 327 (2017).
24. I. Fernández-Hernández, C. Rhiner, E. Moreno, Adult Neurogenesis in Drosophila. *Cell Rep.* **3**, 1857–1865 (2013).
25. A. R. Simões *et al.*, Damage-responsive neuro-glial clusters coordinate the recruitment of dormant neural stem cells in Drosophila. *Dev. Cell* **57**, 1661–1675.e7 (2022).
26. G. Li, A. Hidalgo, Adult neurogenesis in the Drosophila brain: The evidence and the void. *Int. J. Mol. Sci.* **21**, 6653 (2020).
27. N. J. Strausfeld, *Arthropod Brains: Evolution, Functional Elegance, and Historical Significance* (Harvard University Press, 2012), (September 3, 2022).
28. S. Scotto-Lomassese *et al.*, Suppression of adult neurogenesis impairs olfactory learning and memory in an adult insect. *J. Neurosci.* **23**, 9289–9296 (2003).
29. T. Takahashi *et al.*, Systemic RNA interference for the study of learning and memory in an insect. *J. Neurosci. Meth* **179**, 9–15 (2009).
30. C. Klattenhoff *et al.*, Drosophila rasiRNA pathway mutations disrupt embryonic axis specification through activation of an ATR/Chk2 DNA damage response. *Dev. Cell* **12**, 45–55 (2007).
31. J. de Belle, M. Heisenberg, Associative odor learning in Drosophila abolished by chemical ablation of mushroom bodies. *Science* **263**, 692–695 (1994).
32. M. Heisenberg, A. Borst, S. Wagner, D. Byers, Drosophila mushroom body mutants are deficient in olfactory learning: Research papers. *J. Neurogenet.* **2**, 1–30 (1985).
33. M. Mizunami, J. M. Weibrecht, N. J. Strausfeld, Mushroom bodies of the cockroach: Their participation in place memory. *J. Comp. Neurol.* **402**, 520–537 (1998).
34. D. Malun, N. Plath, M. Giurfa, A. D. Moseleit, U. Müller, Hydroxyurea-induced partial mushroom body ablation in the honeybee Apis mellifera: Volumetric analysis and quantitative protein determination. *J. Neurobiol.* **50**, 31–44 (2002).
35. R. E. Campbell *et al.*, A monomeric red fluorescent protein. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 7877–7882 (2002).
36. G. Ylla *et al.*, Insights into the genomic evolution of insects from cricket genomes. *Commun. Biol.* **4**, 733 (2021).
37. C. A. Whittle, A. Kulkarni, C. G. Extavour, Evolutionary dynamics of sex-biased genes expressed in cricket brains and gonads. *J. Evol. Biol.* **34**, 1188–1211 (2021).
38. M. Heisenberg, Mushroom body memoir: From maps to models. *Nat. Rev. Neurosci.* **4**, 266–275 (2003).
39. S. J. C. Caron, V. Ruta, L. F. Abbott, R. Axel, Random convergence of olfactory inputs in the Drosophila mushroom body. *Nature* **497**, 113–117 (2013).
40. C.-C. Chen *et al.*, Visualizing long-term memory formation in two neurons of the Drosophila brain. *Science* **335**, 678–685 (2012).
41. F. Lazarini, P.-M. Lledo, Is adult neurogenesis essential for olfaction? *Trends Neurosci.* **34**, 20–30 (2011).
42. M. Cayre, S. Scotto-Lomassese, J. Malaterre, C. Strambi, A. Strambi, Understanding the regulation and function of adult neurogenesis: Contribution from an insect model, the house cricket. *Chem. Senses* **32**, 385–395 (2007).
43. K. Ito, Y. Hotta, Proliferation pattern of postembryonic neuroblasts in the brain of Drosophila melanogaster. *Dev. Biol.* **149**, 134–148 (1992).
44. W. Breitwieser, F. H. Markussen, H. Horstmann, A. Ephrussi, Oskar protein interaction with Vasa represents an essential step in polar granule assembly. *Gene Dev.* **10**, 2179–2188 (1996).
45. J. S. Chang, L. Tan, P. Schedl, The Drosophila CPEB homolog, Orb, is required for oskar protein expression in oocytes. *Dev. Biol.* **215**, 91–106 (1999).
46. K. E. Kistler *et al.*, Phase transitioned nuclear Oskar promotes cell division of Drosophila primordial germ cells. *Elife* **7**, e37949 (2018).
47. D. S. Johnston, D. Beuchle, C. Nüsslein-Volhard, staufen, a gene required to localize maternal RNAs in the Drosophila egg. *Cell* **66**, 51–63 (1991).
48. R. E. Boswell, A. P. Mahowald, tudor, a gene required for assembly of the germ plasm in Drosophila melanogaster. *Cell* **43**, 97–104 (1985).
49. F. J. Maza, J. Sztarker, M. E. Cozzarin, M. G. Lepore, A. Delorenzi, A crabs' high-order brain center resolved as a mushroom body-like structure. *J. Comp. Neurol.* **529**, 501–523 (2021).
50. N. J. Strausfeld, G. H. Wolff, M. E. Sayre, Mushroom body evolution demonstrates homology and divergence across Pancrustacea. *Elife* **9**, e52411 (2020).
51. N. Strausfeld, M. E. Sayre, Shore crabs reveal novel evolutionary attributes of the mushroom body. *Elife* **10**, e65167 (2021).
52. C. Doeffinger, V. Hartenstein, A. Stollewerk, Compartmentalization of the precheliceral neuroectoderm in the spider Cupiennius salei: Development of the arcuate body, optic ganglia, and mushroom body. *J. Comp. Neurol.* **518**, 2612–2632 (2010).
53. C. E. Juliano, S. Z. Swartz, G. M. Wessel, A conserved germline multipotency program. *Development* **137**, 4113–4126 (2010).
54. L. Blondel, T. E. M. Jones, C. G. Extavour, Bacterial contribution to genesis of the novel germ line determinant oskar. *Elife* **9**, e45539 (2020).
55. C. A. Whittle, A. Kulkarni, N. Chung, C. G. Extavour, Adaptation of codon and amino acid use for translational functions in highly expressed cricket genes. *Bmc Genomics* **22**, 234 (2021).
56. P. M. Macdonald, M. Kanke, A. Kenny, Community effects in regulation of translation. *Elife* **5**, e10965 (2016).
57. C. Rongo, E. R. Gavis, R. Lehmann, Localization of oskar RNA regulates oskar translation and requires Oskar protein. *Dev. Camb Engl.* **121**, 2737–46 (1995).
58. A. Kulkarni, C. G. Extavour, Convergent evolution of germ granule nucleators: A hypothesis. *Stem Cell Res.* **24**, 188–194 (2017).
59. M. Bose, M. Lampe, J. Mahamid, A. Ephrussi, Liquid-to-solid phase transition of oskar ribonucleoprotein granules is essential for their function in Drosophila embryonic development. *Cell* **185**, 1308–1324.e23 (2022).
60. A. Kulkarni, S.S. Parhad, C.G. Extavour, *Small RNA*-seq of unmated adult cricket Gryllus bimaculatus brains. *National Center for Biotechnology Information* Short Read Archive (NCBI SRA). https://www.ncbi.nlm.nih.gov/bioproject/?term=Prjna837371. Deposited 12 May 2022.
61. K. Katoh, K. Misawa, K. Kuma, T. Miyata, MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**, 3059–3066 (2002).
62. A. Stamatakis, RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
63. J. A. Castro-Mondragon *et al.*, JASPAR 2022: The 9th release of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* **50**, gkab1113- (2021).
64. T. L. Bailey, J. Johnson, C. E. Grant, W. S. Noble, The MEME Suite. *Nucleic Acids Res.* **43**, W39–W49 (2015).
65. P. Stothard, The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *Biotechniques* **28**, 1102–1104 (2000).

66. M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads. *Embnet J.* **17**, 10–12 (2011).
67. B. Li, C. N. Dewey, RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *Bmc Bioinformatics* **12**, 323 (2011).
68. A. Dobin *et al.*, STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
69. Y.-W. Hsieh, A. Alqadah, C.-F. Chuang, An optimized protocol for electrophoretic mobility shift assay using infrared fluorescent dye-labeled oligonucleotides. *J. Vis. Exp.*, 54863 (2016), 10.3791/54863.

PNAS

Supporting Information for

oskar acts with the transcription factor Creb to regulate long-term memory in crickets

Arpita Kulkarni[1,2], Ben Ewen-Campen[1,3], Kanta Terao[4,5,], Yukihisa Matsumoto[4], Yaolong Li[6], Takayuki Watanabe[4,7], Jonchee A. Kao[8], Swapnil S. Parhad[9,10, 12], Guillem Ylla[1,11], Makoto Mizunami[4] and Cassandra G. Extavour[1,8,12*]

* Cassandra G. Extavour

Email: extavour@oeb.harvard.edu

This PDF file includes:

Other supporting materials for this manuscript include the following:

Extended Materials & Methods

G. bimaculatus *husbandry*

For behavior experiments, *G. bimaculatus* crickets were maintained in the Mizunami laboratory at 27ºC on a 12:12 light cycle, with a diet of insect food pellets, as previously described(1). For gene expression analysis, quantitative PCR, and cell proliferation experiments, crickets were maintained in the Extavour laboratory at 28ºC and 35% relative humidity on a 12:12 light cycle, with a diet of grain and cat food, as previously described(2).

In situ hybridization

For in situ hybridization, brains were dissected and de-sheathed in ice-cold 1x Phosphate Buffered Saline (1X PBS) as previously described(3, 4). Brains were fixed one hour in 4% paraformaldehyde in 1X PBS, followed by an additional overnight fixation in the same solution at 4ºC, or for an additional 3-4 hours at room temperature. *osk* transcripts were detected using a 788 bp probe, following standard protocols(2) with the following modifications to reduce background: 20-minute Proteinase K (Thermo Fisher Scientific, Cat# EO0491) treatment followed by a 30-minute fixation in 0.8% glutaraldehyde in 1X PBS and 4% paraformaldehyde in 1X PBS. The *osk* probe was used at 1.0 ng/µl concentration and hybridized at 69-70ºC. Brains were sectioned after in situ development was completed, by embedding in 4% low-melt agarose in distilled water, and sectioning at 50-90µM using a Leica VT1000S vibratome.

Immunostaining

For immunostaining, primary antibodies used were as follows: rabbit anti-Gb-Vasa and anti-Gb-Piwi(5) 1:300, mouse anti-RNA polymerase II pSer 6 Mab H5 (Covance MMS-129R) 1:100, FITC-conjugated anti-alpha Tubulin (Sigma F2168) 1:100 and rabbit anti-*Drosophila* Vasa (kind gift of Paul Lasko, McGill University) 1:500 following standard procedures as previously described(2). Goat anti-rabbit secondary antibodies conjugated to Alexa 488, Alexa 555 or Alexa 568 (Invitrogen) were used at 1:500 or 1:1000. Counterstains used were Hoechst 33342 (Sigma B2261) at 0.1 to 0.05 mg/ml and FITC-conjugated phalloidin (Sigma P5282) at 1 U/ml. For antibody staining, brains were embedded in 4% low-melt agarose in distilled water, and sectioned at 50-90µM using a Leica VT1000S vibratome, prior to incubation with the primary antibody. For analysis of apoptosis and DNA damage, brains were fixed 4h after EdU injection and sectioned via vibratome, with EdU detection performed first (using Invitrogen's Click-iT protocol), followed by antibody staining and confocal analysis.

RNA extractions, cDNA preparation and cloning of gene fragments for production of dsRNA

Brains from unmated male adults were dissected in ice-cold 1X PBS, then immediately homogenized in TRIzol (Thermo Fisher Scientific, catalog number 15596026). Total RNA was extracted following the manufacturer's instructions, including a 30-minute DNAse treatment. 1µg of total RNA was used as template for cDNA synthesis using SuperScript III (Thermo Fisher Scientific, catalog number 18-080-044) with oligo-dT primers. cDNA was diluted 1:10 prior to PCR with gene specific primers, and 2 µL of template was used per 25 µL PCR reaction. PCR products were run on a 1% agarose gel and desired bands were gel eluted following IBI Scientific's PCR purification and gel elution kit (catalog number IB47030). Then, products were cloned into Zero blunt TOPO PCR cloning kit (Thermo Fisher Scientific, catalog number 450245) using electro-competent *DH5α-E* cells (Thermo Fisher Scientific, catalog number 11319019).

Identification and annotation of Piwi Proteins in G. bimaculatus

We previously used a *G. bimaculatus* transcriptome(6) to identify two RNA fragments corresponding to two Piwi family proteins (*piwi*: JQ434103 and *piwi-2*: KC242806.1(5, 7)). All previous published analyses of *piwi* in *G. bimaculatus* were performed with "*piwi*" (JQ434103), as only this gene showed enriched expression in embryonic germ cells(5). Since the time of our initial studies on *piwi*, we assembled and annotated an updated *G. bimaculatus* genome(8). For the present study, we therefore performed new BLAST searches to clarify the status of *piwi* orthologs in this cricket (Suppl. Fig. S5). We found both previously identified fragmented *piwi* RNA sequences within the new gene annotations(8) with gene IDs GBI_17641 (containing "*piwi*" fragment JQ434103) and GBI_07509 (containing "*piwi-2*" fragment KC242806.1) respectively. We also identified two additional putative novel *piwi*-like genes annotated in the *G. bimaculatus* genome, with gene IDs GBI_09750 and GBI_09796 (8).

Using InterProscan, we confirmed that the amino acid sequences of the proteins encoded by these four putative *piwi* genes contained the typical characteristics of the Argonaut/Piwi proteins(9) namely a Paz domain followed by a C-terminal Piwi domain. Additionally, we inferred the gene tree of the Argonaute/Piwi protein family using the putative *G. bimaculatus* Piwi and Argonaute protein amino acid sequences obtained from the genome, together with sequences of publicly available Piwi and Argonaute proteins from other insects (*Drosophila melanogaster*, *Apis mellifera*, *Bombyx mori*, *Tribolium castaneum*, *Blattella germanica*, *Zootermopsis nevadensis*, *Acyrthosiphon pisum*, and *Locusta migratoria)*. Protein sequence alignments were performed with MUSCLE(10, 11) in Geneious (v3.8.425; www.geneious.com), and the gene tree was inferred with RAxML v8.2.11(12) with 100 bootstrap iterations to obtain the support values of each node. The tree was then visualized with ggtree(13, 14). The resulting tree differentiated four groups of sequences with bootstrap values above 90%, each of which contained different Piwi/Argonaute subfamilies as follows: Argonaute 1 proteins (AGO1; included GBI_02015), Argonaute 2 (AGO2; included GBI_13717), Argonaute 3 (AGO3; included GBI_01357), and Piwi proteins and their paralogs (Aub and Siwi, the Piwi paralogs in *D. melanogaster* and *B. mori* respectively; included GBI_17641) (Suppl. Fig. S5). This indicated that our previous analyses(5) had indeed targeted the true *piwi* ortholog in *G. bimaculatus*. Accordingly, all gene expression and function analyses in the present study were also performed on this true *piwi* ortholog (GBI_17641).

Analysis of the small RNA species present in *piwi*^{RNAi} animals indicated that our *piwi* RNAi experiments specifically targeted the true *piwi* (GBI_17641), and did not impact expression of the other 3 *piwi* subfamily genes (Suppl. Table S4). The quality control of the eight small RNA-seq samples (3 *piwi*^{RNAi}, 2 *osk*^{RNAi}, 2 *DsRed*^{RNAi}, and 1 untreated or wild-type, WT) was performed with FastQC v0.11.8(15), and the adapters were trimmed with Cutadapt v1.8.1(16). Clean reads were mapped to the *G. bimaculatus* genome assembly(8) with Bowtie 2 v2.3.4.1(17) using parameters "-L 18, -N 0". The numbers of sequenced reads, clean reads, and mapped reads are shown in Suppl. Table S1. The mapped reads were retrieved using samtools v1.9(18) for obtaining the read length distributions (Suppl. Fig. S1B). The proportion of miRNAs and piRNAs in each sample was extrapolated as the percentage of reads of 22-23 nts and 28-29 nts respectively (Suppl. Fig. S1B). The FeatureCounts function from the R package Rsubread v2.0.0(19) was used to count the number of reads mapped to all annotated genes and build a table of counts. The counts of o*sk* (GBI_0140) and the four *piwi* genes (GBI_09750, GBI_09796, GBI_07509, and GBI_17641) were retrieved (Suppl. Table S4). The small RNA-seq reads mapping to the target genes were assumed to be reads of the siRNAs produced from the dsRNA(20). The thousands of such reads mapping to our targeted *piwi* (GBI_17641) and the absence of such siRNA reads mapping to other p*iwi* genes, suggest that no off-target effects impaired the expression of other *piwi* genes.

RNA interference (RNAi)

Unmated adult male crickets within one week of their final molt were injected with 2 µL of double-stranded RNA (dsRNA) through a hole pierced in the median ocellus(21) using a 10 µL syringe fitted with a 26S gauge tip (WPI, Tokyo, Japan; Hamilton Inc., Nevada, USA). Behavioral tests were repeated using two non-overlapping fragments of *oskar* (742bp and 503bp), a 646bp fragment of *piw*i, a 541bp fragment of *vasa*, two non-overlapping fragments of *CrebA* (both 387bp fragments), a 384bp fragment of *CrebB*, and a 678bp fragment of *DsRed* as a negative control (7). Double-stranded RNA concentrations used were 10 µM for *oskar*; 3.38 µg/µL for *piwi*; 2.71 µg/µL for *vasa*; 2.97µg/µL for *DsRed*; 6 µg/µL for *oskar/piwi* double knock down*;* 7µg/µL for *CrebA;* 7µg/µL for *CrebB; and* 7µg/µL for *DsRed* (Suppl. Table S2).

Gene expression from mRNA-seq data

To check the expression of *G. bimaculatus* genes in nervous systems of wild-type animals, previously generated mRNA-seq libraries were used(22) and the complete CDS for genes of interest were obtained from the recently published genome(8). Reads were trimmed with Cutadapt v3.4(16), and mapped to the full *G. bimaculatus* CDS using Geneious Read Mapper(23). DESeq2 normalized counts in fragments per kilobase per exon per million mapped fragments (FPKMs) were then obtained for all genes of interest. The expression of *osk* and *piwi* genes (Suppl. Table S4) shows that the depleted *piwi* (GBI_17641) is one of the two *piwi* genes expressed in wild-type brains of males and females. In the same way, we obtained the expression in FPKMs of *vasa* (GBI_17344), *CrebA* (GBI_04244), and *CrebB* (GBI_02305) (Suppl. Table S6).

Olfactory learning behavioral memory assays

Adult male crickets at eight days after the final molt were used in all experiments, because learning and memory capabilities are highly affected by reproductive status and aging (24), and in our experience, young unmated males exhibit more stable memory scores and longer memory retention scores than females or older males. Three days before conditioning, individual crickets were separated into 100-mL beakers and deprived of drinking water to enhance their motivation to search for water. Two days before conditioning (ten days after the imaginal molt), each cricket was injected with dsRNA as described above. Two days after dsRNA injection, each cricket was subjected to an odor preference test, in which the animal was allowed to freely visit peppermint and vanilla odors(25). The time spent at each of the peppermint and vanilla odor sources was measured cumulatively to evaluate relative odor preference(25). Crickets were subjected to 4-trial conditioning, in which an odor was paired with water reward, with an inter-trial interval of five minutes(4, 26). For conditioning, a small filter paper was attached to the needle of a hypodermic syringe. The syringe was filled with water reward (unconditioned stimulus), and the filter paper was soaked with peppermint essence (conditioned stimulus). At one hour and one day after the end of the conditioning, each cricket was subjected to an odor preference test. The relative odor preference of each conditioned and control animal was measured using the preference index (PI) for rewarded odor (peppermint), defined as tP/(tP+tV) *100 (%), where tP is the time spent exploring the peppermint source and tV is the time spent exploring the vanilla (unrewarded) odor. Wilcoxon's test was used to compare odor preferences before and after training. For multiple comparisons, Holm's method was used to adjust the significance level.

Quantitative PCR (qPCR)

Two days after dsRNA injection, brains were dissected from unmated male adults (within a week of final molt to adulthood) in ice-cold 1x PBS, then immediately homogenized in TRIzol (Thermo Fisher Scientific, catalog number 15596026). Total RNA was extracted from a total of six brains per treatment, following the manufacturer's instructions, including a 30-minute DNAse treatment. 1 µg of total RNA was used as template for cDNA synthesis using SuperScript III (Thermo Fisher Scientific, catalog number 18-080-044) with oligo-dT primers. cDNA was diluted 1:10 prior to qPCR, and 6 µL of template was used per 25 µL qPCR reaction. (PerfeCta SYBR Green SuperMix, Low ROX, Quanta Biosciences, catalog number 101414-158). qPCR reactions were conducted in triplicate, and fold change was calculated using the $\Delta\Delta$Ct method(27), with standard deviation propagated following standard methods. *Beta-tubulin* was used as a reference gene(2). Primers amplifying a 130bp fragment of *Gb-CrebA*, a 140bp fragment of *Gb-CrebB*, a 234bp fragment of *Gb-oskar*, a 166 bp fragment of *Beta-tubulin* (2), a 129bp fragment of *Gb-piwi*, a 150bp fragment of *Gb-vasa* and a 120bp fragment of *Gb-FGFR* (Fibroblast Growth Factor Receptor) were used (Suppl. Table S1).

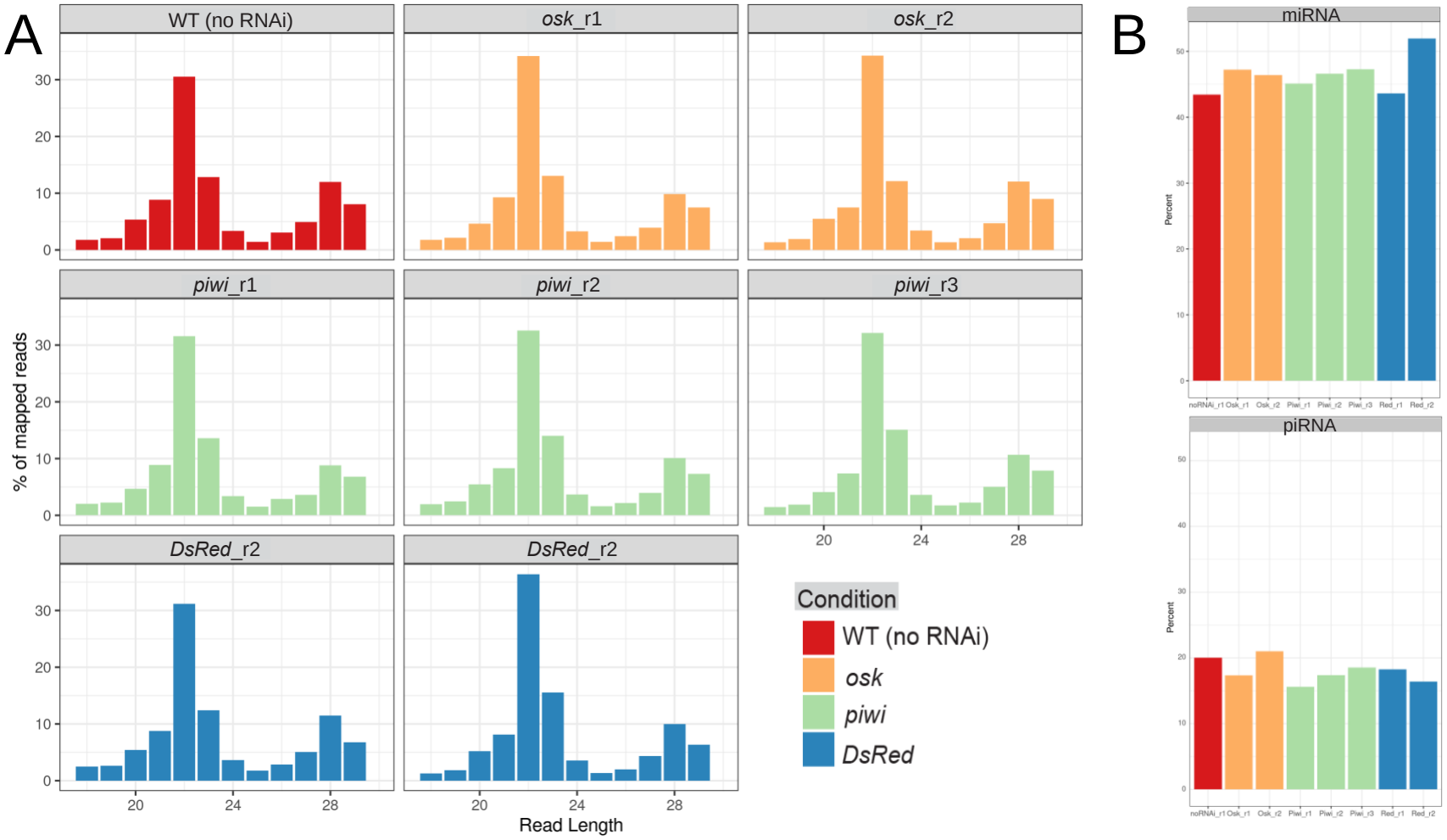

Figure S1. Small RNA library mapping and analysis (A) Read Length Distribution**:** Percentage of mapped reads of each length from 18 to 29 nucleotides in each sequenced small RNA library colored by RNAi treatment. The two peaks at ~22 and ~28 mainly correspond to miRNAs and piRNAs. **(B)** miRNAs vs piRNA: Estimated proportion of reads corresponding to microRNAs and piRNAs in each sample (colored by condition) based on the percentage of reads of 22-23 nucleotides in length and 28-29 nucleotides respectively.

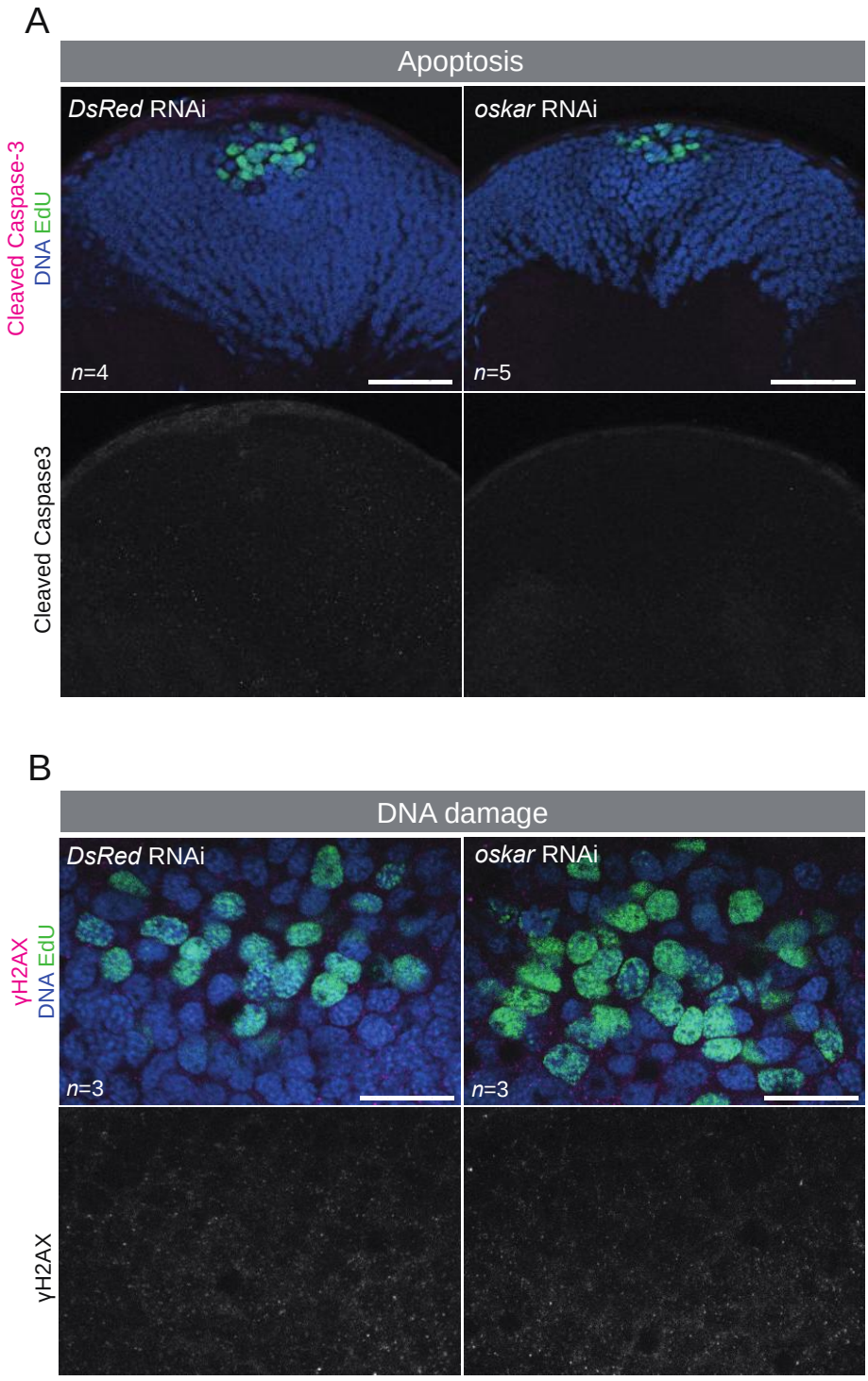

Figure S2. Assessment of DNA damage and apoptosis in *osk*^{RNAi} adult mushroom body neuroblasts. Apoptosis marker Cleaved caspase 3 (**A**) and DNA damage marker gamma H2AX immunostaining (**B**) in adult mushroom bodies, including neuroblasts of control and *osk*^{RNAi} brains.

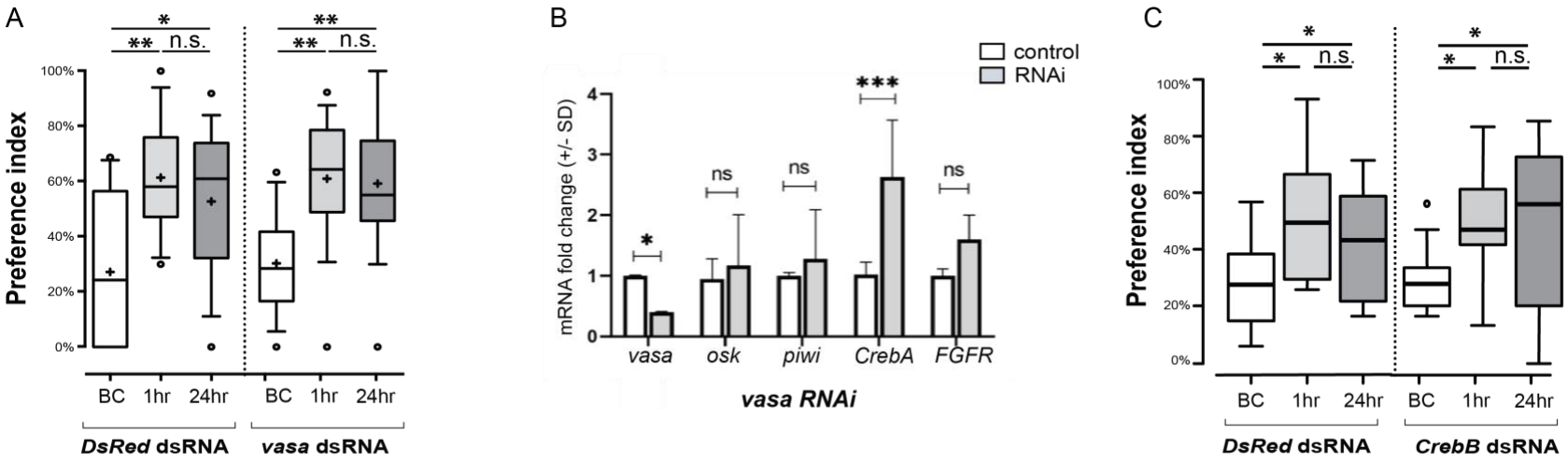

Figure S3. Vasa and CrebB are not required for long-term memory in the cricket *G. bimaculatus*. (**A**) *vasa* RNAi fails to recapitulate the long-term olfactory memory phenotype seen in *osk* and *piwi* RNAi. BC = "Before Conditioning", 1hr = "1 hour post training" and 24hrs = "24 hours after training". (**B**) qPCR on $vasa^{RNAi}$ brains shows significant up-regulation of *CrebA*. (**C**) $CrebB^{RNAi}$ does not recapitulate the long-term memory phenotype shown by $CrebA^{RNAi}$. N=9 for *CrebB,* and N=10 for *DsRed.* (* $p < 0.05$, ** $p < 0.01$, *** $p<0.001$, **** $p < 0.0001$, n.s. = not statistically significant).

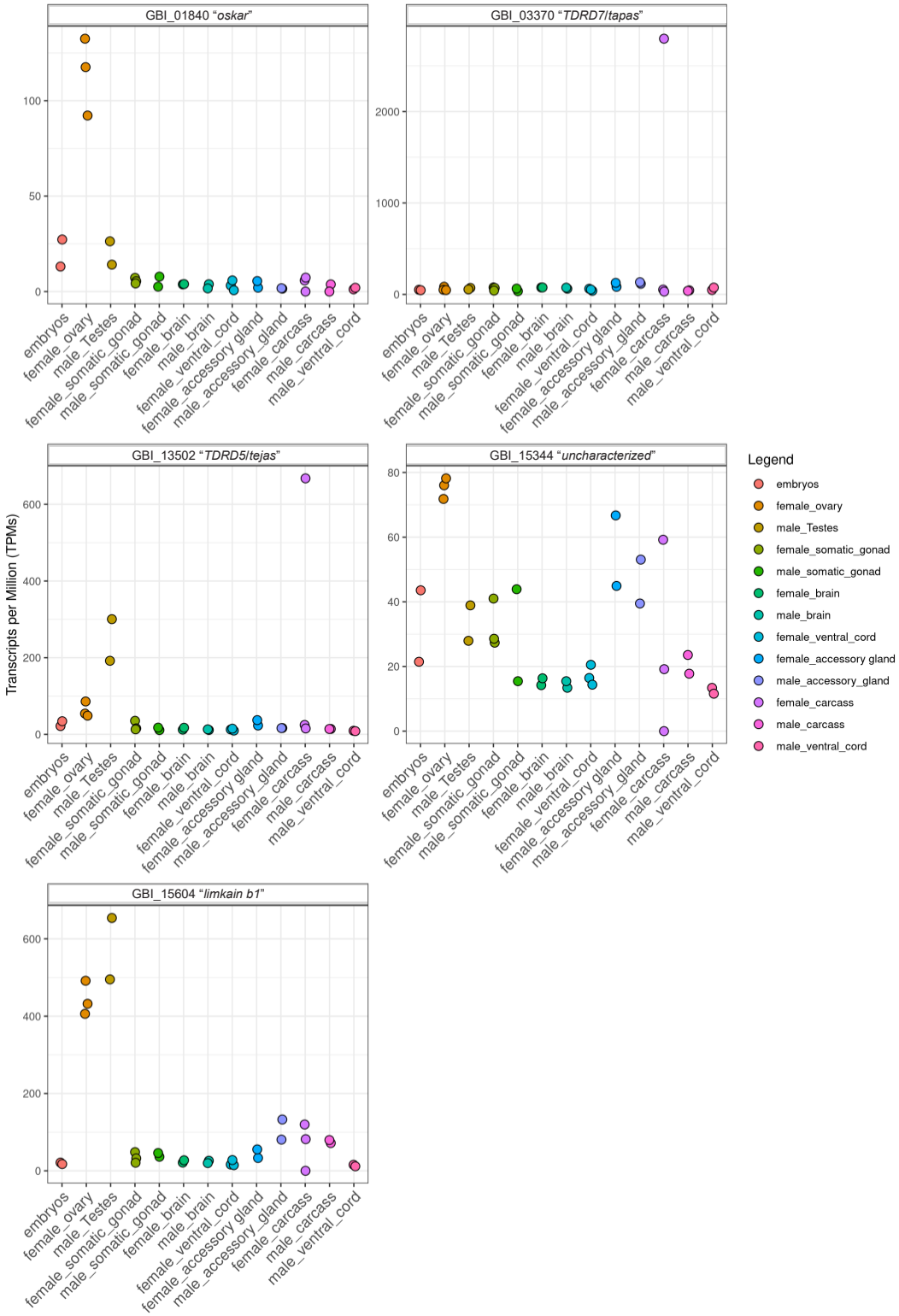

Figure S4. **LOTUS domain-containing gene expression in *G. bimaculatus***. Expression plots (in transcripts per million (TPM)) for LOTUS-domain containing genes in *G. bimaculatus* brain and gonad transcriptomes (corresponds to data shown in Table S8).

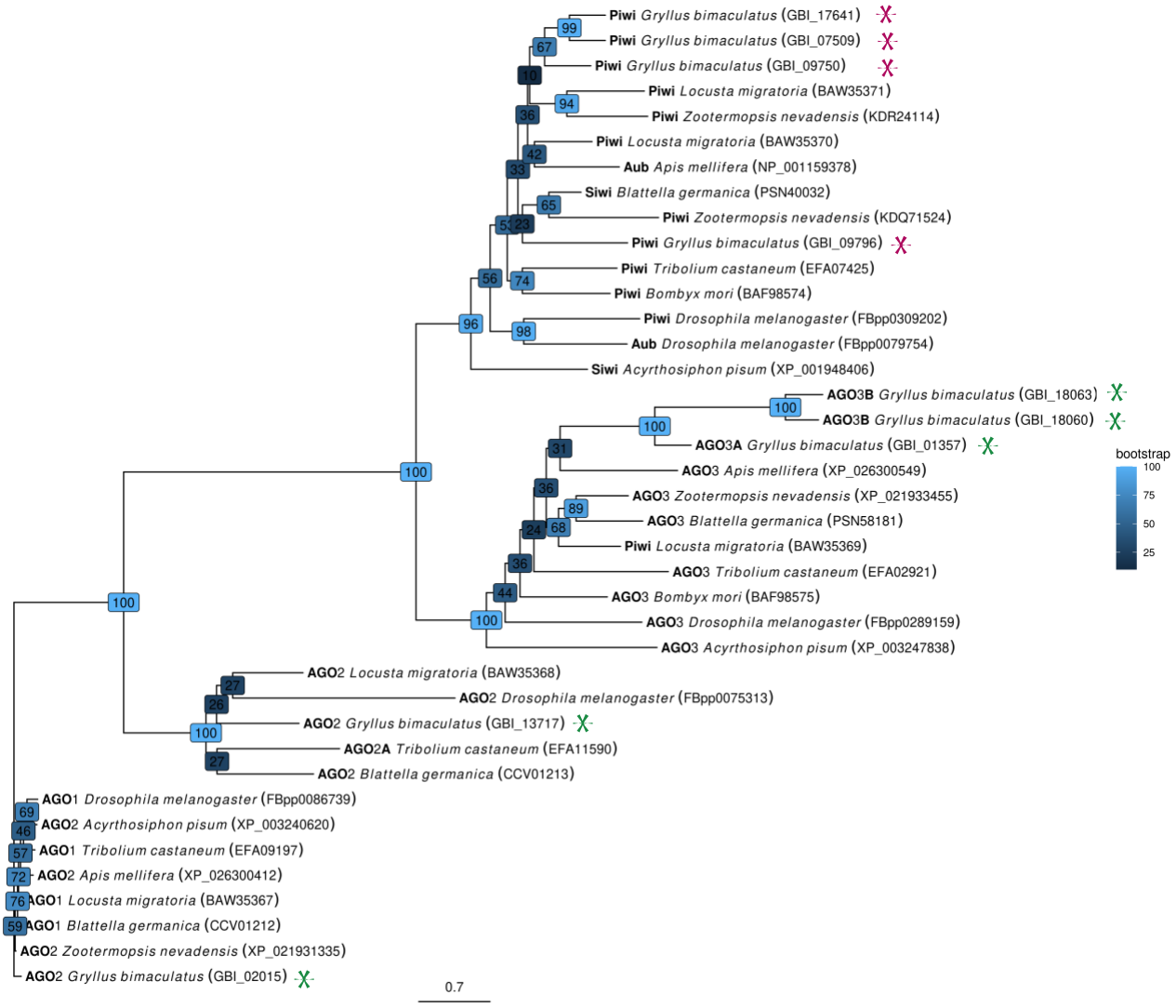

Figure S5. Piwi family genes in the cricket *G. bimaculatus*. *G. bimaculatus piwi* ortholog identification and phylogenetic analysis. Argonaute family gene tree generated with the PIWI, AUB, AGO1, AGO2, and AGO3 protein sequences from *Drosophila melanogaster*, *Apis mellifera*, *Bombyx mori*, *Tribolium castaneum*, *Blattella germanica*, *Zootermopsis nevadensis*, *Acyrthosiphon pisum*, and *Locusta migratoria*. Values at nodes represent bootstrap support, in boxes color-coded from dark (lowest) to light (highest) blue. *G. bimaculatus piwi* and *argonaut* genes indicated by red and green asterisks respectively.

Gene	**qPCR Primers**
CrebA	F:CCGCCTTCACCACCGCAGAC
	R:ATGCTTAGTTTGGGGATGACGACGC
CrebB	F:AGACTCCTGCTAATATTCAGCCTGT
	R:TAGACTGTCATCACTTCCTGCTTCT
piwi	F:TTCGGCCAACTACTTCAAGC
	R:AGAGTTTCCCGATGAACACG
vasa	F:GAACATTGTGAGCCTCATGC
	R:TTGCTGAGCCTGGTGGTAT
oskar	F:TTGTTGACCATTCCCTTCCT
	R:ACTCCACAACACCACTCC
Beta-Tubulin	F:TGGACTCCGTCCGGTCAGGC
	R:TCGCAGCTCTCGGCCTCCTT
FGFR	F:ACCTGTCTTCAGCGAACTAGTG
	R:ACTTGCTTCTTGGCTGGATG

Table S1. Primers used for quantitative PCR of all listed *G. bimaculatus* genes. All sequences are in 5’ to 3’ orientation.

Biological replicate #	RNAi Experimental condition	# adult males injected	µl dsRNA injected	dsRNA concentration
1	Control 1 (uninjected)	16	N/A	N/A
	Control 2 (*DsRed* injected)	13	4	4.4µg/µl
	piwi	17	4	4.6µg/µl
2	Control 2 (*DsRed* injected)	10	3	10 µg/µl
	piwi	17	3-4	8 µg/µl
	oskar	14	4	7.8-14µg/µl
3	Control 2 (*DsRed* injected)	14	3	9.8µg/µl
	piwi	11	3	10µg/µl
	oskar	16	3	11µg/µl

Table S2. A total of 128 unmated, adult male cricket brains (16 brains from Control 1 "uninjected controls", 37 brains from Control 2 "*DsRed* injected", 45 brains from "*piwi*" dsRNA injected, and 30 brains from "*osk*" dsRNA injected) were dissected 48h post dsRNA injection and processed for making small RNA libraries.

File Name	Sample	# Raw reads	# Clean reads	% Clean	# Mapped reads	% Mapped
Bill.smRNA.cricket_dsRed_RNAi.brain.r1.fastq.gz	DsRed_RNAi.brain.r1	9,193,785	4,404,190	47.90%	3,568,586	81.03%
Bill.smRNA.cricket_dsRed_RNAi.brain.r2.fastq.gz	DsRed_RNAi.brain.r2	36,880,382	30,410,536	82.46%	25,931,302	85.27%
Bill.smRNA.cricket_No_RNAi.brain.r1.fastq.gz	No_RNAi.brain.r1	11,311,618	6,373,759	56.35%	5,317,264	83.42%
Bill.smRNA.cricket_osk_RNAi.brain.r1.fastq.gz	osk_RNAi.brain.r1	10,811,667	6,439,713	59.56%	5,086,605	78.99%
Bill.smRNA.cricket_osk_RNAi.brain.r2.fastq.gz	osk_RNAi.brain.r2	29,321,677	24,047,151	82.01%	20,518,912	85.33%
Bill.smRNA.cricket_piwi_RNAi.brain.r1.fastq.gz	piwi_RNAi.brain.r1	7,120,996	3,637,311	51.08%	2,828,184	77.75%
Bill.smRNA.cricket_piwi_RNAi.brain.r2.fastq.gz	piwi_RNAi.brain.r2	18,492,553	13,202,708	71.39%	10,629,195	80.51%
Bill.smRNA.cricket_piwi_RNAi.brain.r3.fastq.gz	piwi_RNAi.brain.r3	21,923,563	17,222,470	78.56%	13,628,698	79.13%

Table S3. Number of raw sequenced small RNA reads, number and percentage of clean reads, and number and percentage of reads mapped to the *G. bimaculatus* genome.

Gene ID	**Gene Name**	***DsRed.r1***	***DsRed.r2***	**WT(no RNAi)**	***osk.r1***	***osk.r2***	***piwi.r1***	***piwi.r2***	***piwi.r3***
GBI_01840	*oskar*	6	65	0	17945	109776	6	2065	1
GBI_09750	*piwi*	1	4	0	0	2	0	4	0
GBI_09796	*piwi*	0	2	0	0	0	1	2	0
GBI_07509	*piwi*	1	0	0	0	0	0	1	0
GBI_17641	*piwi*	156	51	5	6	332	6346	41552	8940

Table S4. Number of small RNA-seq reads mapped to *osk* and *piwi* genes to assess specificity of RNAi knockdowns. These small RNA-seq reads come from the siRNA detected following injections of dsRNA for these respective genes. The targeted sequence in each library is highly enriched by small RNA-seq reads. In the *piwi*^{RNAi} experiments, only the targeted *piwi* (GBI_17641) shows a high number of mapped reads in the libraries generated from animals injected with the dsRNA against *piwi*, indicating that the other three *piwi* orthologs present in the *G. bimaculatus* genome were unlikely to be targeted by our approach.

Mus musculus	*Drosophila melanogaster*	*Tribolium castaneum*	*Apis mellifera*	*Acyrthosiphon pisum*	*Zootermopsis nevadensis*	*Gryllus bimaculatus*
OASIS (NP_036087.2)	CrebA (NP_524087.3)	CREB-A (XP_966968.2) CREB-like (XP_973089.1)	CREB-A (XP_003250132.1) CREB-like (XP_001121941.2)	CREB-A (XP_001948312.1) CREB-like (XP_001949209.1)	CREB-A (KDR23733.1) CREB-like (KDR11962.1)	CREB-like (GAIZ01013153, *Gryllus firmus* TSA), CrebA (GBI_04244, *Gryllus bimaculatus*)
CREB1 (NP_034082.1) CREM (NP_001104320.1) ATF1 (NP_031523.3)	CrebB (NP_001097017.1)	XP_008192794	XP_623392.3	XP_008186705.1	KDR23211.1	GAIZ01012380 and GAIZ01007852 (*Gryllus firmus* TSA), CrebB (GBI_02305, *Gryllus bimaculatus*)
ATF2 (NP_001020264.1)	dATF2 (NP_001033973.1)	XP_974257.1	XP_003249317.1	-	KDR15907.1	-
ATF3 (NP_031524.2)	dATF3 (NP_620473.1)	XP_008192299.1	XP_003251072.1	XP_003243558.1	KDR22659.1	-
ATF4 (NP_001274109.1) ATF5 (NP_109618.1)	cryptocephal (NP_524897.1)	NP_001280506.1	XP_006562898.1	XP_003247514.1	KDR14457.1	-
ATF6 (NP_001074773.1)	dATF6 (NP_995745.1)	XP_008201619.1	XP_395889.5	XP_008183931.1	KDR16855.1	GAIZ01013605 (*Gryllus firmus* TSA)

Table S5. GenBank IDs of *Creb/ATF* family member orthologs in mouse and other insects including the cricket *G. bimaculatus*. This information was used to construct a Creb phylogenetic tree (Fig. 3B) to infer the evolutionary relationships between mammalian Creb proteins and their insect counterparts.

***G. bimaculatus* gene name**	**Gene ID**	**(gene expression in FKPM per tissue)**							
		Female brain 1	**Female brain 2**	**Male brain 1**	**Male brain 2**	**Female ventral 1**	**Female ventral 2**	**Male ventral 1**	**Male ventral 2**
oskar	GBI_01840	2.76	3.85	3.28	1.24	3.24	0.84	1.38	1.29
vasa	GBI_17344	33.65	31.18	32.84	24.66	55.2	28.47	28.54	22.4
CrebA	GBI_04244	198.35	155.47	170.68	158.7	258.26	225.45	218.68	155.47
CrebB	GBI_02305	46.99	42.01	47.27	36.85	49.45	42.64	27.01	29.12
piwi	GBI_09750	0.34	0.24	0	0	0.24	0.07	0.23	0.4
piwi	GBI_09796	21.4	18.25	36.83	44.19	16.96	22.43	35.34	47.58
piwi	GBI_07509	0.94	0.26	0.95	1.28	3.18	1.14	0.58	1.67
piwi	GBI_17641	64.43	57.71	54.72	59.32	61.65	54.38	55.78	33.71

Table S6. Gene expression levels for *osk*, *piwi*, *vasa*, and *CrebA/B* (in FPKM per tissue) from brain and ventral cord transcriptomes of male and female adult *Gryllus bimaculatus*(22). Gene IDs as per the annotated cricket genome(8).

CRE site #	Gene	Site	Cloning primer	EMSA Probe sequence	EMSA Probe reverse complement sequence
1	*oskar*	CRE-I	F: ACAGCCTGAGGCGCTATCTA	CATCCAAAGAGCG**TGGCGTCA**CGTATCAGC	GCTGATACG**TGACGCCA**CGCTCTTTGGATG
			R: AGCGTCTTCTCTGGCGACTA		
2	*oskar*	CRE-II	F: TCCTAGCGATTTTCGCTGAC	TTATT**TTACGTCA**ATGAAACATAATTAATTCG	CGAATTAATTATGTTTCAT**TGACGTAA**AATAA
			R:TCAACTTCTCCCACATTCCA		

Table S7. Primers used for cloning and generation of EMSA probes for *G. bimaculatus oskar.* All sequences are in 5' to 3' orientation. **Bold face type** indicates predicted CRE site in probe sequences.

GeneID	GBI_01840	GBI_03370	GBI_13502	GBI_15344	GBI_15604
Gene name	**oskar**	**TDRD7 "Tapas"**	**TDRD5 "Tejas"**	**uncharacterized**	**limkain b1 -like**
GB_embryos_sample1	13.1	51.12	21.67	21.46	21.22
GB_embryos_sample2	27.26	46.48	34.29	43.57	17.46
GB_female_accessory gland_sample1	2	83.19	23.1	44.92	33.16
GB_female_accessory gland_sample2	5.49	126.53	37.19	66.72	55.36
GB_female_brain_sample1	3.74	76.1	12.26	14.22	21.25
GB_female_brain_sample2	3.89	76.39	17.28	16.35	27.26
GB_female_carcass_sample1	5.82	52.8	24.59	59.2	119.77
GB_female_carcass_sample2	7.27	31.89	15.42	19.17	81.82
GB_female_carcass_sample3	0	2796.68	667.6	0	0
GB_female_ovary_sample1	132.49	50.54	54.23	71.82	406.07
GB_female_ovary_sample2	92.23	47.76	48.75	78.18	432.07
GB_female_ovary_sample3	117.59	84.3	85.82	76.04	491.6
GB_female_somatic_gonad_sample1	5.41	70.59	15.6	27.36	32.25
GB_female_somatic_gonad_sample2	4.23	42.27	13.09	28.58	20.9
GB_female_somatic_gonad_sample3	7.16	76.76	35.41	41.01	48.4
GB_female_ventral_cord_sample1	3.16	62.18	12.76	16.45	16.23
GB_female_ventral_cord_sample2	0.68	39.68	9.85	14.37	14.63
GB_female_ventral_cord_sample3	5.83	53.59	14.91	20.52	27.84
GB_male_accessory_gland_sample1	1.49	116.51	16.58	53.06	132.66
GB_male_accessory_gland_sample2	1.73	133.82	16.34	39.48	80.7
GB_male_brain_sample1	3.82	62.6	11.38	13.41	26.05
GB_male_brain_sample2	1.63	74.48	13.14	15.45	19.96
GB_male_carcass_sample1	3.74	44.63	14.16	17.77	71.82
GB_male_carcass_sample2	0	38.8	14	23.56	79.59
GB_male_somatic_gonad_sample1	2.55	64	17.6	43.9	46.03
GB_male_somatic_gonad_sample2	7.81	35.91	11.19	15.45	36.34
GB_male_Testes_sample1	26.31	54.91	191.92	27.94	495.09
GB_male_Testes_sample2	14.1	71.36	300.5	38.91	653.98
GB_male_ventral_cord_sample1	1.15	48.55	9.37	13.39	15.5
GB_male_ventral_cord_sample2	1.98	75.09	8.43	11.58	11.73

Table S8. Expression levels of LOTUS domain-containing genes (in transcripts per million (TPM)) in the *G. bimaculatus* genome (data plotted in Supplementary Figure S4).

Sequence name	# putative CRE motifs found	putative CRE motif sequence (5' to 3')	Strand	Start	End	p-value (p<0.0001)	q-value
TDRD5 "tejas"_GBI_13502_upstream10Kb	1	TGACGYMA	plus	3925	3932	5.93E-05	1
TDRD7 "tapas"_GBI_03370_upstream10Kb	2	TGACGYMA	plus	8185	8192	1.47E-05	0.147
		TGACGYMA	minus	8185	8192	1.47E-05	0.147
uncharacterized_GBI_15344_upstream10Kb	3	TGACGYMA	minus	5868	5875	2.67E-05	0.534
		TGACGYMA	plus	5769	5776	7.40E-05	0.573
		TGACGYMA	minus	7355	7362	8.60E-05	0.573
Limkain b1_GBI_15604_upstream10Kb	3	TGACGYMA	plus	6110	6117	1.47E-05	0.137
		TGACGYMA	minus	6110	6117	1.47E-05	0.137
		TGACGYMA	minus	3128	3135	8.60E-05	0.536
oskar_GBI_01840	2	TGACGYMA	minus	8016	8023	2.67E-05	0.883
		TGACGYMA	minus	6419	6426	5.93E-05	0.883
piwi_GBI_17641	2	TGACGYMA	minus	2503	2510	5.93E-05	0.474
		TGACGYMA	minus	5048	5055	5.93E-05	0.474

Sequence name	Matched CRE sequence (5' to 3')	Total # putative TATA motifs in 10Kb	# putative TATA motifs within 1Kb of CRE	putative TATA motif sequence (5' to 3')	Strand
TDRD5 "tejas"_GBI_13502_upstream10Kb	TGACGTAA	19	2	VTATAWAWRVVNNNN	plus/ plus
TDRD7 "tapas"_GBI_03370_upstream10Kb	TGACGTCA	18	4	VTATAWAWRVVNNNN	minus/ minus/ minus/ plus
	TGACGTCA		4	VTATAWAWRVVNNNN	minus/ minus/ minus/ plus
uncharacterized_GBI_15344_upstream10Kb	TGACGCCA	17	3	VTATAWAWRVVNNNN	minus/ minus/ plus
	TGAGGTCA		3	VTATAWAWRVVNNNN	minus/ minus/ plus
	TGACGTCG		3	VTATAWAWRVVNNNN	plus/ minus/ plus
Limkain b1_GBI_15604_upstream10Kb	TGACGTCA	10	1	VTATAWAWRVVNNNN	plus
	TGACGTCA		1	VTATAWAWRVVNNNN	plus
	TGACGTCG		4	VTATAWAWRVVNNNN	minus/ plus/ minus/ minus

oskar_GBI_01840	TGACGCCA	36	4	VTATAWAWRVVNNNN	minus/ plus/ minus/ plus
	TGACGTAA		3	VTATAWAWRVVNNNN	plus/ minus/ plus
piwi_GBI_17641	TGACGTAA	23	2	VTATAWAWRVVNNNN	minus/ plus
	TGACGTAA		1	VTATAWAWRVVNNNN	plus

Sequence name	**Start**	**End**	**p-value (p<0.001)**	**q-value**	**Matched TATA sequence (5' to 3')**
TDRD5 "tejas"_GBI_13502_upstream10Kb	3064/ 4668	3078/ 4682	0.000654/ 0.000843	0.8/ 0.823	TTATTATAAGAATGGC / GTATGAAGACGCCGA
TDRD7 "tapas"_GBI_03370_upstream10Kb	8717/ 8092/ 8068/ 7763	8731/ 8106/ 8082/ 7777	0.000481/ 0.000601/ 0.000638/ 0.00084	0.814/ 0.831/ 0.831/ 0.836	CCATAAATCCCCCTT/ GTATAATGGGGCGGA/ CGATAAAATGGAGTG/ CTATACAAAGAGTCC
	8717/ 8092/ 8068/ 7763	8731/ 8106/ 8082/ 7777	0.000481/ 0.000601/ 0.000638/ 0.00084	0.814/ 0.831/ 0.831/ 0.836	CCATAAATCCCCCTT/ GTATAATGGGGCGGA/ CGATAAAATGGAGTG/ CTATACAAAGAGTCC
uncharacterized_GBI_15344_upstream10Kb	4777/ 6833/ 4691	4791/ 6847/ 4705	0.00021/ 0.000773/ 0.000981	0.725/ 0.916/ 1	GTAAAAAAGGGGACG/ GTACAAAAACTCCGT/ GTATAGAAGTGAAGG
	4777/ 6833/ 4691	4791/ 6847/ 4705	0.00021/ 0.000773/ 0.000981	0.725/ 0.916/ 1	GTAAAAAAGGGGACG/ GTACAAAAACTCCGT/ GTATAGAAGTGAAGG
	7647/ 6833/ 8145	7661/ 6847/ 8159	0.000135/ 0.000773/ 0.000795	0.725/ 0.916/ 0.916/	GTATAAAAAGTAATC/ GTACAAAAACTCCGT/ CTTTAAAAAAAACGT
Limkain b1_GBI_15604_upstream10Kb	5415	5429	0.000681	1	TTATAAAAGAGAATA
	5415	5429	0.000681	1	TTATAAAAGAGAATA
	3559/ 2546/ 2867/ 2865	3573/ 2560/ 2881/ 2879	0.000158/ 0.000594/ 0.00082/ 0.000866	1/1/1/1	GTATAAAAAACAAAT/ CTATAAAATGAATTT/ ATATATAAATATGTG/ ATATAAATATGTGAA
oskar_GBI_01840	8725/ 7245/ 8988/ 7026	8739/ 7259/ 9002/ 7040	0.000499/ 0.000786/ 0.000135/ 0.000293	1/ 1/ 0.879/ 1	GTATTAAAAACAGCT/ CTATTAATGCAAATG/ CTATATAAACAGAA/ TTATAAAGGGGGCAA
	6465/ 6299/ 7026	6479/ 6313/ 7040	0.000408/ 0.000846/ 0.000293	1/1/1/	ACATAAAAGTCCCTC/ ACATAAAAAAGCACT/ TTATAAAGGGGGCAA

piwi_GBI_17641	2338/ 2832	2352/ 2846	0.000563/ 0.000308	0.688/ 0.688	CTTTAAAAATGCACA/ GTATAAAAATGTAAA
	5351	5365	0.000223	0.688	TTATAAAATGGCACC

Table S9. Prediction of putative CRE sites in LOTUS domain-containing *G. bimaculatus* genes. For predictions, p-value was set to less than/equal to 0.0001. The p-value of a motif occurrence is defined as the probability of a random sequence of the same length as the motif matching that position of the sequence with as good or a better score. The score for the match of a position in a sequence to a motif is computed by summing the appropriate entries from each column of the position-dependent scoring matrix that represents the motif. The q-value of a motif occurrence is defined as the false discovery rate if the occurrence is accepted as significant. If there are multiple CRE predictions for one sequence, the table is sorted by increasing p-value for those CRE predictions.

Supplementary File 1: The Frequency Matrices (PFM) from JASPAR database used to predict CRE sites and TATA boxes in the presumptive regulatory regions of *G. bimaculatus* genes.

```
TATA box
MEME version 4
ALPHABET= ACGT
strands: + -
Background letter frequencies
A 0.25 C 0.25 G 0.25 T 0.25

MOTIF POL012.1 TATA-Box
letter-probability matrix: alength= 4 w= 15 nsites= 389 E= 0
 0.156812  0.372751  0.390746  0.079692
 0.041131  0.118252  0.046272  0.794344
 0.904884  0.000000  0.005141  0.089974
 0.007712  0.025707  0.005141  0.961440
 0.910026  0.000000  0.012853  0.077121
 0.688946  0.000000  0.000000  0.311054
 0.925450  0.007712  0.025707  0.015424
 0.570694  0.005141  0.113111  0.311054
 0.398458  0.113111  0.403599  0.084833
 0.143959  0.347044  0.385604  0.123393
 0.213368  0.377892  0.329049  0.079692
 0.210797  0.326478  0.329049  0.133676
 0.210797  0.303342  0.329049  0.156812
 0.174807  0.275064  0.357326  0.192802
 0.197943  0.259640  0.359897  0.182519
URL http://jaspar.genereg.net/matrix/POL012.1


CRE consensus sequence full site
MEME version 4
ALPHABET= ACGT
strands: + -
Background letter frequencies
A 0.25 C 0.25 G 0.25 T 0.25

MOTIF MA0018.2 CREB1
letter-probability matrix: alength= 4 w= 8 nsites= 11 E= 0
 0.000000  0.090909  0.090909  0.818182
 0.000000  0.090909  0.909091  0.000000
 1.000000  0.000000  0.000000  0.000000
 0.000000  0.818182  0.181818  0.000000
 0.090909  0.000000  0.909091  0.000000
 0.000000  0.272727  0.000000  0.727273
 0.181818  0.636364  0.090909  0.090909
 0.727273  0.000000  0.090909  0.181818
```

URL http://jaspar.genereg.net/matrix/MA0018.2

Supplementary File 2: FASTA files containing the simulated one thousand 10-Kb long DNA fragments generated to test the frequency of occurrence of CRE sites in randomly generated sequences, and LOTUS domain-containing genes.

Download Link

SI References

1. Y. Matsumoto, M. Mizunami, Olfactory learning in the cricket Gryllus bimaculatus. *J Exp Biology* 203, 2581–8 (2000).

2. F. Kainz, B. Ewen-Campen, M. Akam, C. G. Extavour, Notch/Delta signalling is not required for segment generation in the basally branching insect Gryllus bimaculatus. *Development* 138, 5015–5026 (2011).

3. A. Hamada, *et al.*, Loss-of-function analyses of the fragile X-related and dopamine receptor genes by RNA interference in the cricket Gryllus bimaculatus. *Dev Dynam* 238, 2025–2033 (2009).

4. T. Takahashi, *et al.*, Systemic RNA interference for the study of learning and memory in an insect. *J Neurosci Meth* 179, 9–15 (2009).

5. B. Ewen-Campen, S. Donoughe, D. N. Clarke, C. G. Extavour, Germ Cell Specification Requires Zygotic Mechanisms Rather Than Germ Plasm in a Basally Branching Insect. *Curr Biol* 23, 835–842 (2013).

6. V. Zeng, *et al.*, Developmental Gene Discovery in a Hemimetabolous Insect: De Novo Assembly and Annotation of a Transcriptome for the Cricket Gryllus bimaculatus. *Plos One* 8, e61479 (2013).

7. B. Ewen-Campen, J. R. Srouji, E. E. Schwager, C. G. Extavour, oskar Predates the Evolution of Germ Plasm in Insects. *Curr Biol* 22, 2278–2283 (2012).

8. G. Ylla, *et al.*, Insights into the genomic evolution of insects from cricket genomes. *Commun Biology* 4, 733 (2021).

9. X. Zhou, Z. Liao, Q. Jia, L. Cheng, F. Li, Identification and characterization of Piwi subfamily in insects. *Biochem Bioph Res Co* 362, 126–131 (2007).

10. R. C. Edgar, MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *Bmc Bioinformatics* 5, 113 (2004).

11. R. C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32, 1792–7 (2004).

12. A. Stamatakis, RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313 (2014).

13. G. Yu, Using ggtree to Visualize Data on Tree-Like Structures. *Curr Protoc Bioinform* 69, e96 (2020).

14. G. Yu, D. K. Smith, H. Zhu, Y. Guan, T. T. Lam, ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol Evol* 8, 28–36 (2017).

15. S. Andrews, Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data (2010) (September 3, 2022).

16. M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads. *Embnet J* 17, 10–12 (2011).

17. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9, 357–359 (2012).

18. H. Li, *et al.*, The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079 (2009).

19. Y. Liao, G. K. Smyth, W. Shi, The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. *Nucleic Acids Res* 47, gkz114- (2019).

20. J. C. Montañés, C. Rojano, G. Ylla, M. D. Piulachs, J. L. Maestro, siRNA enrichment in Argonaute 2-depleted Blattella germanica. *Biochimica Et Biophysica Acta Bba - Gene Regul Mech* 1864, 194704 (2021).

21. Y. Matsumoto, S. Noji, M. Mizunami, Time Course of Protein Synthesis-Dependent Phase of Olfactory Memory in the Cricket Gryllus bimaculatus. *Zool Sci* 20, 409–416 (2003).

22. C. A. Whittle, A. Kulkarni, C. G. Extavour, Evolutionary dynamics of sex-biased genes expressed in cricket brains and gonads. *J Evolution Biol* 34, 1188–1211 (2021).

23. M. Kearse, *et al.*, Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28, 1647–1649 (2012).

24. Y. Matsumoto, M. Mizunami, Context-Dependent Olfactory Learning in an Insect. *Learn Memory* 11, 288–293 (2004).

25. Y. Matsumoto, M. Mizunami, Temporal determinants of long-term retention of olfactory memory in the cricket Gryllus bimaculatus. *The Journal of Experimental Biology* 205, 1429–1437 (2002).

26. Y. Matsumoto, A. Hatano, S. Unoki, M. Mizunami, Stimulation of the cAMP system by the nitric oxide-cGMP system underlying the formation of long-term memory in an insect. *Neurosci Lett* 467, 81–85 (2009).

27. K. J. Livak, T. D. Schmittgen, Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2−ΔΔC T Method. *Methods* 25, 402–408 (2001).