Cell Reports, Volume 37

## Supplemental information

## **Repression precedes independent evolutionary**

## gains of a highly specific gene expression pattern

Jian Pu, Zinan Wang, Haosu Cong, Jacqueline S.R. Chin, Jessa Justen, Cédric Finet, Joanne Y. Yew, and Henry Chung



**Figure S1.** *In situ* hybridization of *bond* in the ejaculatory bulb (EB) in 21 species. *In situ* hybridization of *bond* to the EB of various *Drosophila* and other closely related species showed EB expression of *bond* in *D. melanogaster*, *D. simulans*, *D. erecta*, *D. yakuba*, *D. ananassae*, *D. pseudoobscura*, and *D. subobscura*, *Scaptodrosophila* species, *S. latifasciaeformis*, *S. lebanonensis*, and *S. rufifrons*, but not in other species tested. Arrows indicate *bond* expression in the EB determined by *in situ* hybridization. *Dmel* = *Drosophila melanogaster*, *Dsim* = *D. simulans*, *Dyak*= *D. yakuba*, *Dere* = *D. erecta*, *Dana* = *D. ananassae*, *Dpse* = *D. pseudoobscura*, *Dsub* = *D. subobscura*, *Dneb* = *D. nebulosa*, *Dstu* = *D. sturtevanti*, *Dimm* = *D. immigrans*, *Dnas* = *D. nasuta*, *Dame* = *D. americana*, *Dvir* = *D. virilis*, *Dmer* = *D. mercatorum*, *Dmoj* = *D. mojavensis*, *Cpro* = *Chymomyza* procnemis, *Slat* = *Scaptodrosphila* latifasciaeformis, *Sleb* = *S. lebanonensis*, *Sruf* = *S. rufifrons*, and *Mdom* = *Musca domestica*. The EBs of the *Scaptodrosophila*, *Chymomyza* and *Musca* species are morphologically different from the EBs of the *Drosophila* species. The EB of *M. domestica* is previously referred to as an ejaculatory sac. (Related to Figure 1B).

## D. melanogaster



**Figure S2.** The large intron of *bond* contains sequences that can drive GFP expression in the EB semicircular wall epithelium (*swe*) in *D. melanogaster*. Overlapping fragments from the non-coding region around the *D. melanogaster* bond locus were screened for *cis*-regulatory sequences that were able to drive GFP reporter protein expression in the EB *swe*. Black lines indicate fragments able to drive GFP expression in the EB *swe*. Grey lines indicate fragments not able to drive GFP in the EB *swe*. (Related to Figure 2, 3A)

Dmel : Dsim : Dyak : Dara : Dana : Dpse : Dstu : Dwil : Dstu : Dstu : Drir : Sleb : Mdom :	TTTTCTTAC-C-GTOTOCOCTAL-AGTATORGATGTATCAT TTTTCTTAC-C-GTOTOGGTAL-AGTATORGATGTATCAT TTTTTTAC-C-GTOTOGGTAL-AGTATORGATGTATCAT TTTCTTAC-C-GTOTOGGTAL-AGTATORGATGTATCAT TTTCTTA-C-C-GTOTOGCTAL-TOTTALCCATGCTATCAT AGTGGGGCC-C-GTOTAGCAL-GGATTGGCATTTCCALTTCCAL GTOGGGCC-C-GTOTAGCAL-GGATTGGCATTTCCALTTCCAL GTOGGGCC-C-GTOTAGCAL-GGATTGGCATTGCALTTCCALAT AGTGGGACCC-C-GTOTAGCAL-GGATTGGCALTGCALAT AGTGGGACCC-C-GTOTAGCAL-GGATTGGCALTGCALTCCALT		C TCACT - CGAC CC TTTT TC G TGAT - CGGC CC TTTT TC G TGAT - CGGC CC TTTT TC A TGAT - CGAC CC TTTT TC G TGAT CGC CGAT CC TT GCC TC TC G TGAT CGC CGAT CC TT GCC TC G TGAT CGC CGAT CC CGC CGC G TGAT CGC CGAT CC CGC CGC CGAT CC CGAT CC CGC CGC CGAT CC CGAT CC CC		
	D. melanogaster	Abd-B 1		Abd-B 2	bc3
Dmel : Dsim : Dyak : Dana : Dpse : Dsub : Dsub : Dsub : Dsub : Dsub : Dsub : Sleb : Mdom :	C  SAAAA  AAAAAA  AAA  T  C  T  SAAAA  T  C  T  SAAAA  SAAAAA  AAA  T  C  T  S  SAAAAA  SAAAAA  AAA  T  C  T  T  S  SAAAAA  SAAAAA  SAAAAAA  SAAAAA  SAAAAAA  SAAAAA  SAAAAAA  SAAAAAA  SAAAAAA  SAAAAAAA  SAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA		TRANCATARGC TRANCATARGC ALTERNATION NITTAACATARGC ALTERNATION ALT		A 0T TIGA A 0T TIGA A 0T TIGA A 0T TIGA A 0T TIGA A 0T 0TIGA A
Dmel: Dsim: Dyak: Dere: Dana: Dsub: Dsub: Dsub: Dwil: Dwil: Dwir: Dwir: Dras: Sleb: Sleb: Mdom:	G	AAAAAAATT  AC  AATT    AAAAAATT  AC  AATT    AAAAAATT  AC  AATT    GAAAAATT  AC  AATT    GCAAAATT  C  AATT    GCAAAATT  C  AATT    GCAAAATT  C  AATT    GCAAAATT  C  AATT    GCACAATT  C  AATT    AAATT  CAAACAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	TCARGCTCACTCACTTAC TCARGCTCCCCCCCCCCTTA TCARGCTCCCCCCCCCCCTTA TCCRCTCCCCCCCCCCCCTT TCCRCTCCCCCCCCCC		

**Figure S3. Alignment of the** *bc23* **region.** The 11bp repressor sequences and the two putative Abd-B binding sites in *D. melanogaster* are shown. (Related to Figure 4D, 5C-F)



**Figure S4. Testing 11bp sequences from other** *Drosophila* **species for ability for spatial repression in the EB.** The 11bp sequences from *D. ananassae* and *D. willistoni* at the same location as the *D. melanogaster* 11bp repressor sequences are able to repress expression in the EB *hb* and *hwe*. The 11bp sequences from *D. virilis* did not repress expression in these sites. (Related to Figure 4D)



**Figure S5. Schematic showing positions of Rep region (blue) of EB swe enhancer and the antisense noncoding RNA** *CR44062* **relative to the bond locus.** The Rep region overlaps with the exon of *CR44062*. (Related to Figure 4D)