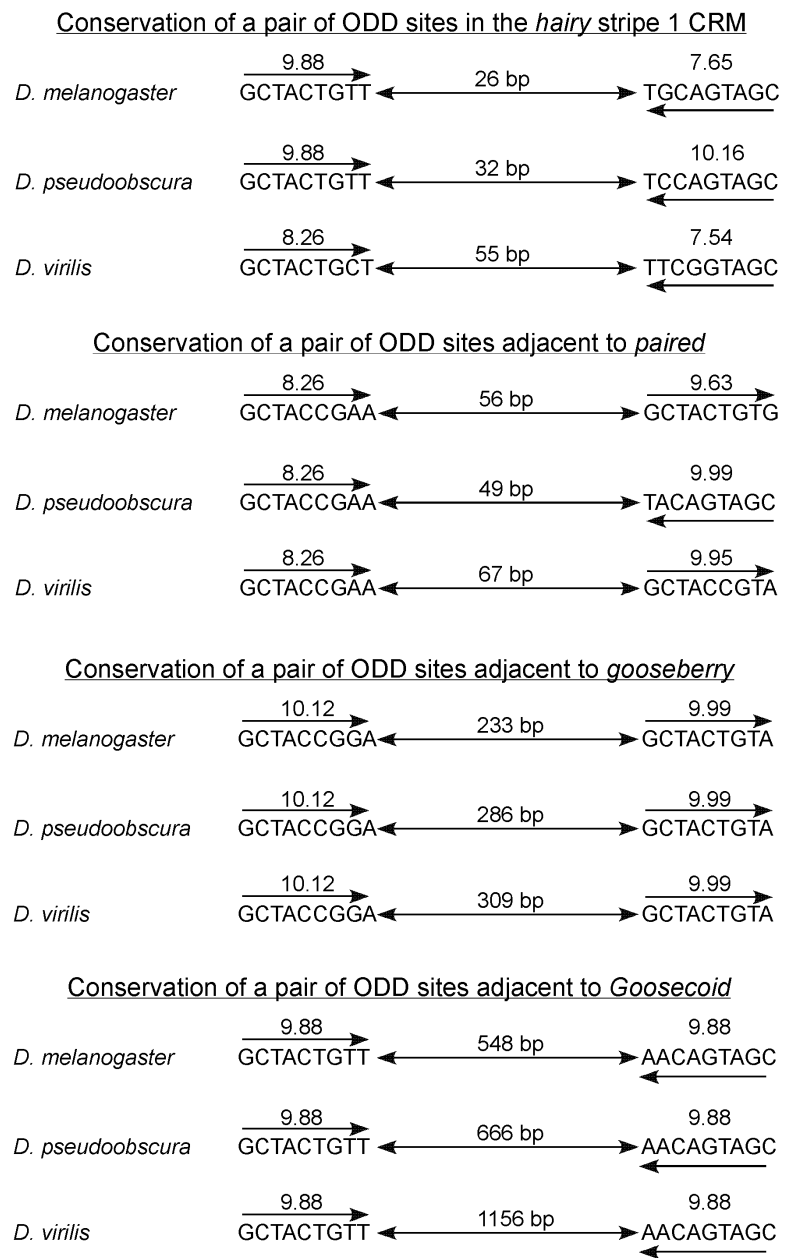


Supplementary Figure 4.

**a) Schematic representation of the conserved Odd binding sites near *hairy*, *paired*, *gooseberry* and *Goosecoid*.** The sequence of the nine base pair binding sites for Odd identified by Target Explorer are shown with the PWM score indicated above the site and the species of *Drosophila* listed to the left of the site. The arrow above or below the sequence indicates the orientation of the binding site. The distance between pairs of sites is indicated above the two-headed arrows between sequences. With the exception of *Gsc*, the distance between neighboring Odd sites is similar in all three species.

The original pair of computational conserved binding sites defined for *Gsc* is conserved in *D. melanogaster* and *D. pseudoobscura* but only a single site is present in *D. virilis*. Subsequent computational analysis of the nearby sequence revealed an additional high scoring Odd site (549 bp distant) that is conserved in these three species as well as *D. yakuba*, *D. ananassae* and *D. mojavensis*. The spacing between the sites is increased in both *D. virilis* and *D. mojavensis* due to a ~600 bp insertion in this region (or a corresponding deletion in the other species). To ascertain the significance of this pair of Odd sites a separate Target Explorer<sup>1</sup> search was performed requiring that two Odd sites to occur within 700 bp, that both score >9.5, and that they be conserved in *D. pseudoobscura*. Only 47 segments in the *D. melanogaster* genome meet these criteria.



**b) Positional conservation of Odd sites in *hairy stripe 1* CRM.**

The sequences for different species used for this analysis were obtained by discontinuous MegaBLAST of whole genome shotgun reads for each species or by BLAT search at the UCSC genome bioinformatics site using the *D. melanogaster* sequence from the genomic segment identified by Target Explorer as the query sequence. The listed order of species reflects their evolutionary distance from *D. melanogaster*. The PWM cut-off score for appearance on this list is  $\geq 7.5$ . PWM scores are the sum of individual weights (scores) for each base in the sequence, which are derived from the frequency of occurrence of that base at that position in the aligned motifs recovered from the B1H selection<sup>1,2</sup>. Relative position refers to the distance in base pairs between a particular Odd binding site (#2) and the first Odd binding site (#1). Only single binding sites are recovered in the corresponding regions of *D. willistoni* and *D. mojavensis*. The first *D. melanogaster* site is on chromosome 3L at position 8630244.

<u>number</u>	<u>species</u>	<u>direction</u>	<u>sequence</u>	<u>relative position</u>	<u>score</u>
1	<i>D. melanogaster</i>	for	GCTACTGTT		9.88
2	<i>D. melanogaster</i>	rev	TGCAGTAGC	26	7.65
1	<i>D. simulans</i>	for	GCTACTGTT		9.88
2	<i>D. simulans</i>	rev	TGCAGTAGC	26	7.65
1	<i>D. yakuba</i>	for	GCTACTGTT		9.88
2	<i>D. yakuba</i>	rev	TGCAGTAGC	26	7.65
1	<i>D. erecta</i>	for	GCTACTGTT		9.88
2	<i>D. erecta</i>	rev	TGCAGTAGC	26	7.65
1	<i>D. ananassae</i>	for	GCTACTGTT		9.88
2	<i>D. ananassae</i>	rev	AGCAGTAGC	13	7.54
1	<i>D. pseudoobscura</i>	for	GCTACTGTT		9.88
2	<i>D. pseudoobscura</i>	rev	TCCAGTAGC	32	10.16
1	<i>D. willistoni</i>	rev	AACAGTAGC		9.88
1	<i>D. mojavensis</i>	for	GCTACTGTT		9.88
1	<i>D. virilis</i>	rev	AGCAGTAGC		7.54
2	<i>D. virilis</i>	for	GCTACCGAA	55	8.26

**c) Positional conservation of Odd sites adjacent to *prd*.**

The sequences for different species used for this analysis were obtained by discontinuous MegaBLAST of whole genome shotgun reads for each species or by BLAT search at the UCSC genome bioinformatics site using the *D. melanogaster* sequence from the genomic segment identified by Target Explorer as the query sequence. The listed order of species reflects their evolutionary distance from *D. melanogaster*. The PWM cut-off score for appearance on this list is  $\geq 7.5$ . Relative position refers to the distance in base pairs between a particular Odd binding site (#n) and the first Odd binding site (#1). The first *D. melanogaster* site is on chromosome 2L at position 12070805.

<u>number</u>	<u>species</u>	<u>direction</u>	<u>sequence</u>	<u>relative position</u>	<u>score</u>
1	<i>D. melanogaster</i>	rev	CACAGAAGC		7.63
2	<i>D. melanogaster</i>	for	GCTACCGAA	84	8.26
3	<i>D. melanogaster</i>	for	GCTACTGTG	149	9.63
1	<i>D. simulans</i>	rev	CACAGAAGC		7.63
2	<i>D. simulans</i>	for	GCTACCGAA	69	8.26
3	<i>D. simulans</i>	for	GCTACTGTG	134	9.63
1	<i>D. yakuba</i>	rev	CACAGAAGC		7.63
2	<i>D. yakuba</i>	for	GCTACCGAA	67	8.26
3	<i>D. yakuba</i>	for	GCTACTGTG	131	9.63
1	<i>D. erecta</i>	rev	CACAGAAGC		7.63
2	<i>D. erecta</i>	for	GCTACCGAA	66	8.26
3	<i>D. erecta</i>	for	GCTACTGTG	130	9.63
1	<i>D. ananassae</i>	for	GCTACCGAA		8.26
2	<i>D. ananassae</i>	rev	TTCAGTAGC	56	8.30
1	<i>D. pseudoobscura</i>	for	GCTACCGAA		8.26
2	<i>D. pseudoobscura</i>	rev	TACAGTAGC	49	9.99
1	<i>D. mojavensis</i>	for	GCTACCGAA		8.26
2	<i>D. mojavensis</i>	for	GCTACCGTA	73	9.95
1	<i>D. virilis</i>	for	GCTACCGAA		8.26
2	<i>D. virilis</i>	for	GCTACCGTA	67	9.95

#### d) Positional conservation of Odd sites adjacent to *gsb*.

The sequences for different species used for this analysis were obtained by discontinuous MegaBLAST of whole genome shotgun reads for each species or by BLAT search at the UCSC genome bioinformatics site using the *D. melanogaster* sequence from the genomic segment identified by Target Explorer as the query sequence. The listed order of species reflects their evolutionary distance from *D. melanogaster*. The PWM cut-off score for appearance on this list is  $\geq 7.5$ . Relative position refers to the distance in base pairs between a particular Odd binding site (#n) and the first Odd binding site (#1). The first *D. melanogaster* site is on chromosome 2R at position 20109481.

<u>number</u>	<u>species</u>	<u>direction</u>	<u>sequence</u>	<u>relative position</u>	<u>score</u>
1	<i>D. melanogaster</i>	rev	TACAGTAGC		9.99
2	<i>D. melanogaster</i>	rev	TCCGGTAGC	233	10.12
1	<i>D. simulans</i>	rev	TACAGTAGC		9.99
2	<i>D. simulans</i>	rev	TCCGGTAGC	236	10.12
1	<i>D. yakuba</i>	rev	TACAGTAGC		9.99
2	<i>D. yakuba</i>	rev	TCCGGTAGC	224	10.12
1	<i>D. erecta</i>	rev	CACAGTAGC		9.63
2	<i>D. erecta</i>	rev	TCCGGTAGC	250	10.12
1	<i>D. ananassae</i>	rev	CACAGTAGC		9.63
2	<i>D. ananassae</i>	rev	TCCGGTAGC	318	10.12
1	<i>D. pseudoobscura</i>	rev	TACAGTAGC		9.99
2	<i>D. pseudoobscura</i>	rev	GCCGGTAGC	199	8.34
3	<i>D. pseudoobscura</i>	rev	GCCGGTAGC	206	8.34
4	<i>D. pseudoobscura</i>	rev	TCCGGTAGC	286	10.12
1	<i>D. willistoni</i>	rev	TACAGTAGC		9.99
2	<i>D. willistoni</i>	rev	TCCGGTAGC	318	10.12
1	<i>D. mojavensis</i>	rev	TACAGTAGC		9.99
2	<i>D. mojavensis</i>	rev	CCCGGTAGC	231	9.76
3	<i>D. mojavensis</i>	rev	TCCGGTAGC	338	10.12
1	<i>D. virilis</i>	rev	TACAGTAGC		9.99
2	<i>D. virilis</i>	rev	GACGGTAGC	226	8.17
3	<i>D. virilis</i>	rev	TCCGGTAGC	309	10.12

1. Sosinsky, A., Bonin, C.P., Mann, R.S. & Honig, B. Target Explorer: An automated tool for the identification of new target genes for a specified set of transcription factors. *Nucleic Acids Res* **31**, 3589-3592 (2003).
2. Hertz, G.Z. & Stormo, G.D. Identifying DNA and protein patterns with statistically significant alignments of multiple sequences. *Bioinformatics* **15**, 563-577 (1999).