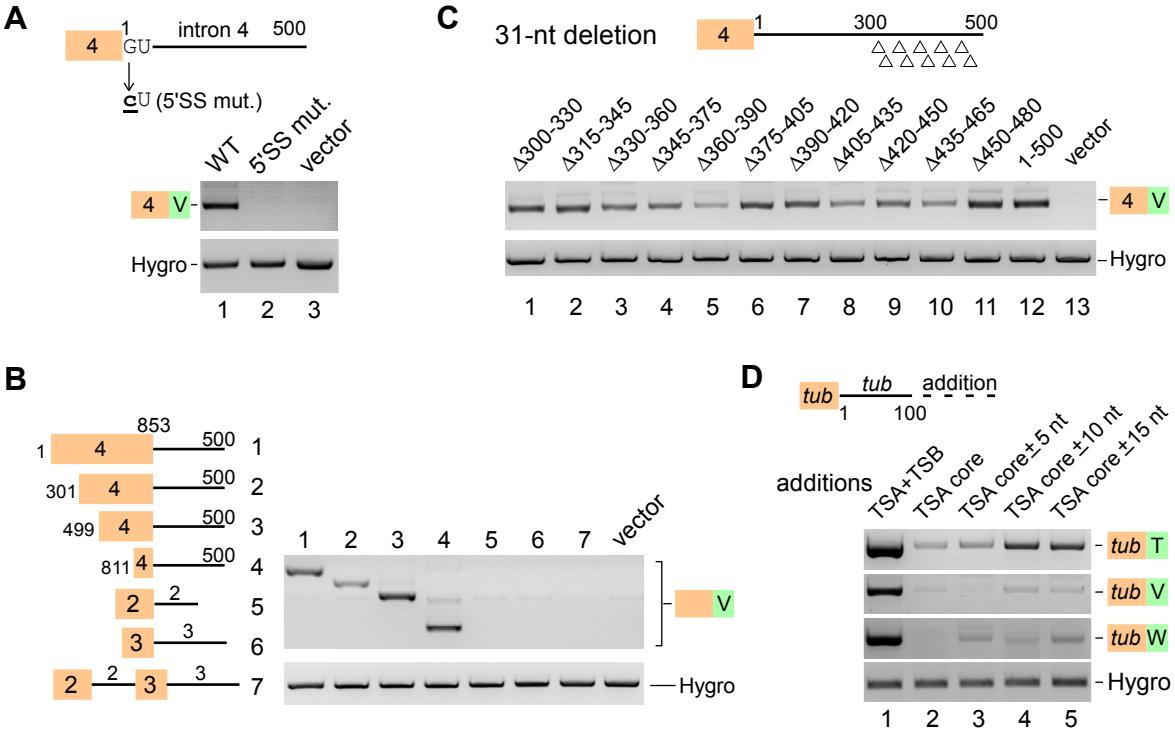

Figure S1. Identification of RNA elements that required for *trans*-splicing of *mod(mdg4).*

(A) *Trans*-spliced product is not detectable when the first nucleotide at the 5'SS of intron 4 is mutated from G to C.

(B) Replacing intron 4 to other 5'-common introns of *mod(mdg4)* abolishes *trans*-splicing activity. However, shorten of the last 5'-exon 4 does not significantly alter *trans*-splicing activity.

(C) Tiled deletions of 31-nt in intron 4 (1-500 nt) reveals several fragments are important to *trans*-splicing activity.

(D) TSA core motif with short flanking sequences is sufficient to promote *trans*-splicing.

Boxes and labels are the same as in Figure 1.

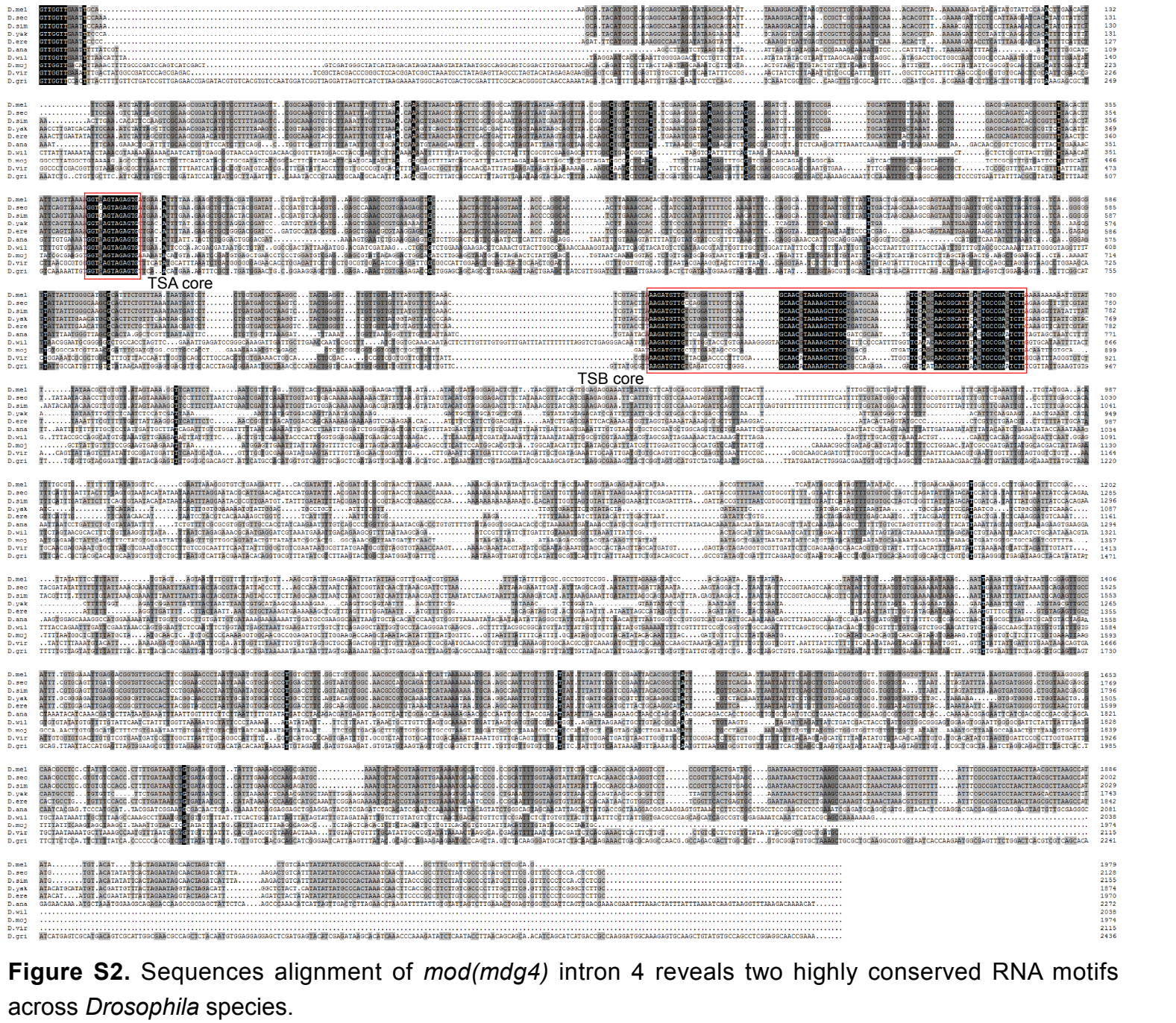

Figure S2. Sequences alignment of *mod(mdg4)* intron 4 reveals two highly conserved RNA motifs across *Drosophila* species.

Intron 4 sequences from available *Drosophila* species were retrieved and aligned by DNAman. Regions of the highly conserved TSA core and TSB core are in red rectangles. Intron 4 from *D. persimilis* cannot be aligned due to missing of a genome fragment. Intron 4 from *D. pseudoobscura* was failed in the alignment, however its TSA core can be found by additional search using MEME software. Enlarged two regions are shown in Figure 2A.

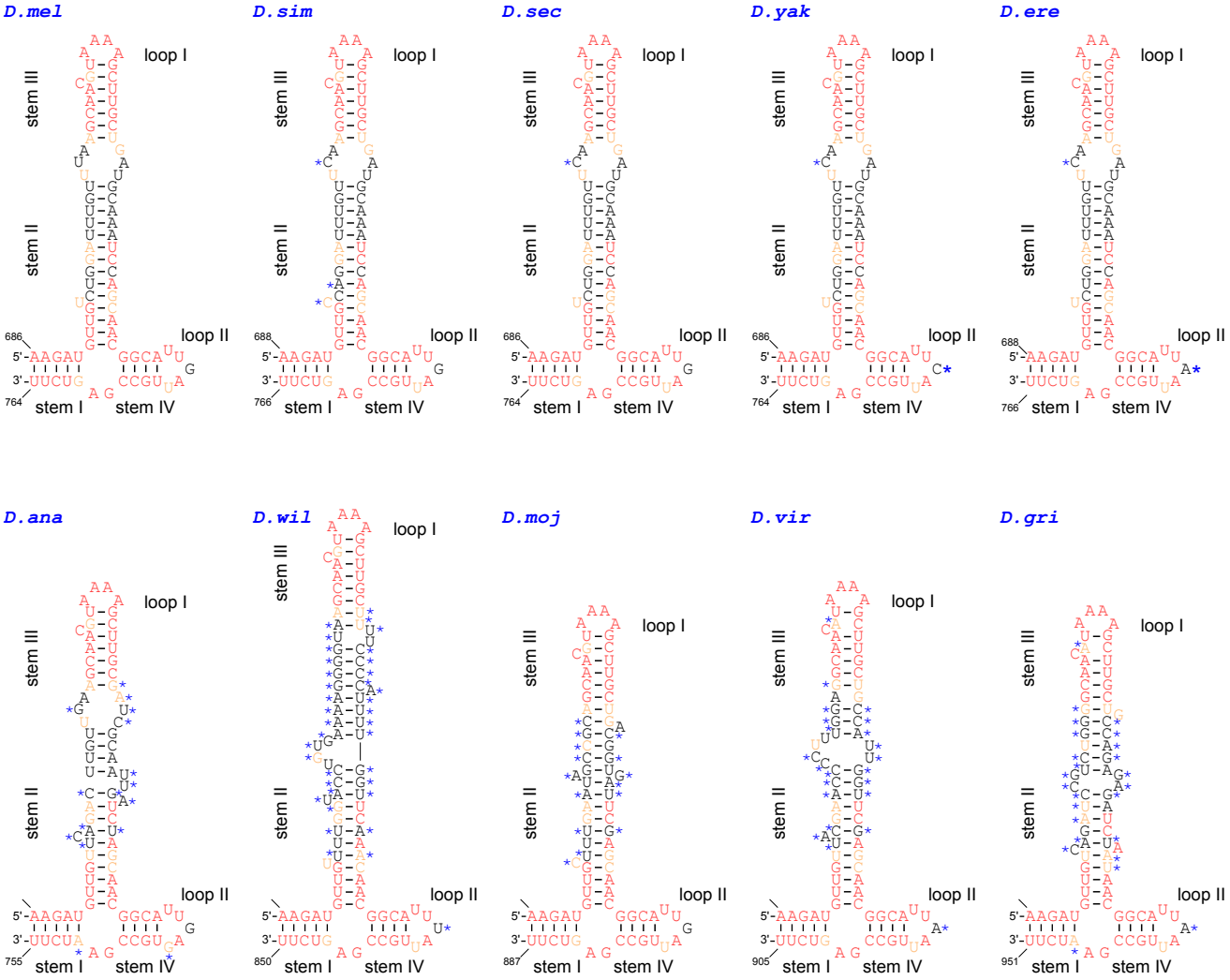

Figure S3. Conserved secondary structures of TSB core RNA across *Drosophila* species. Secondary structures are predicted by online Mfold server. Nucleotide conservation: red (100%); brown (>75%); black (<75%); asterisk (positions with different nucleotide from B core RNA in *D. melanogaster*).

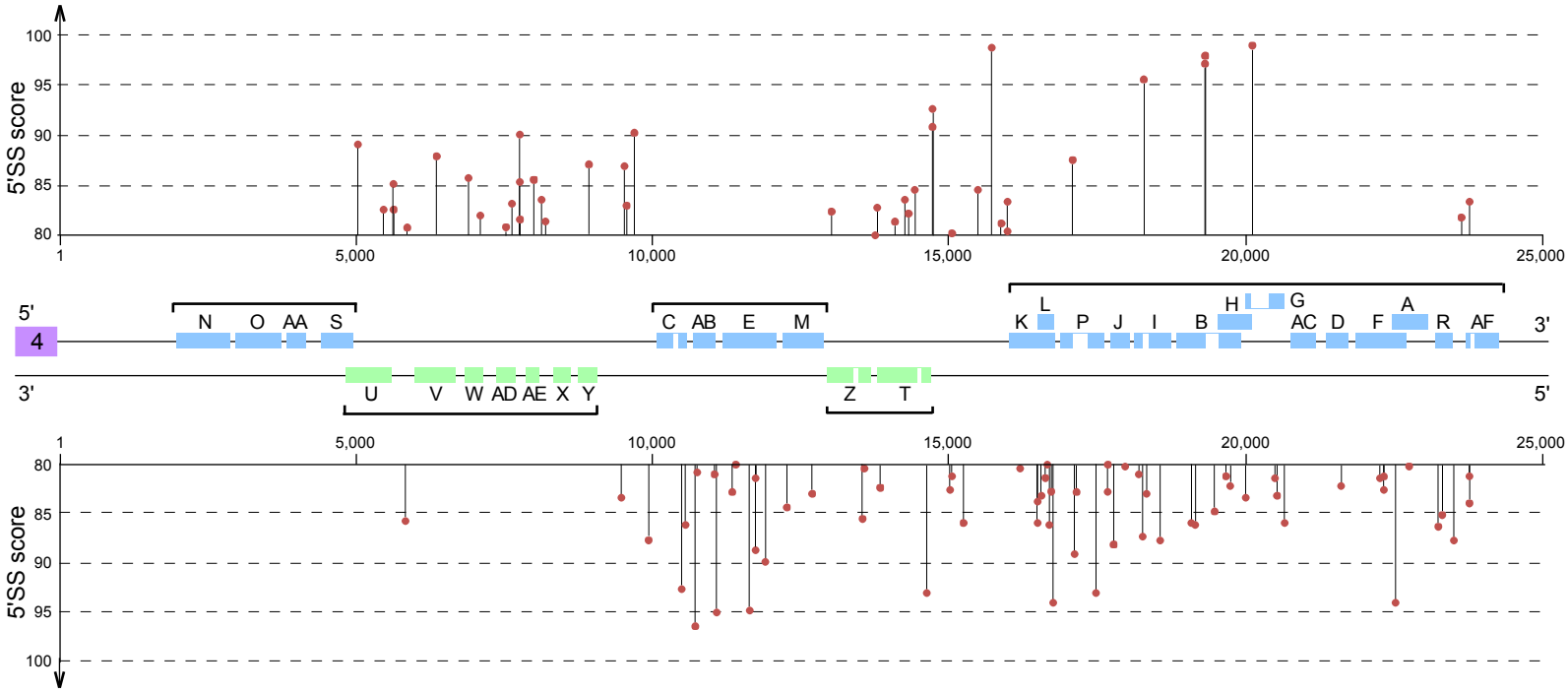

Figure S4. Potential 5'SS in the 3'-transcprits of *Drosophila mod(mdg4)*.

Potential 5'SS are predicted according to high scores (>80) by HSF software. Most of them are located in the UTR regions or internal *cis*-spliced introns (white boxes), but not between the alternatively *trans*-spliced exons.

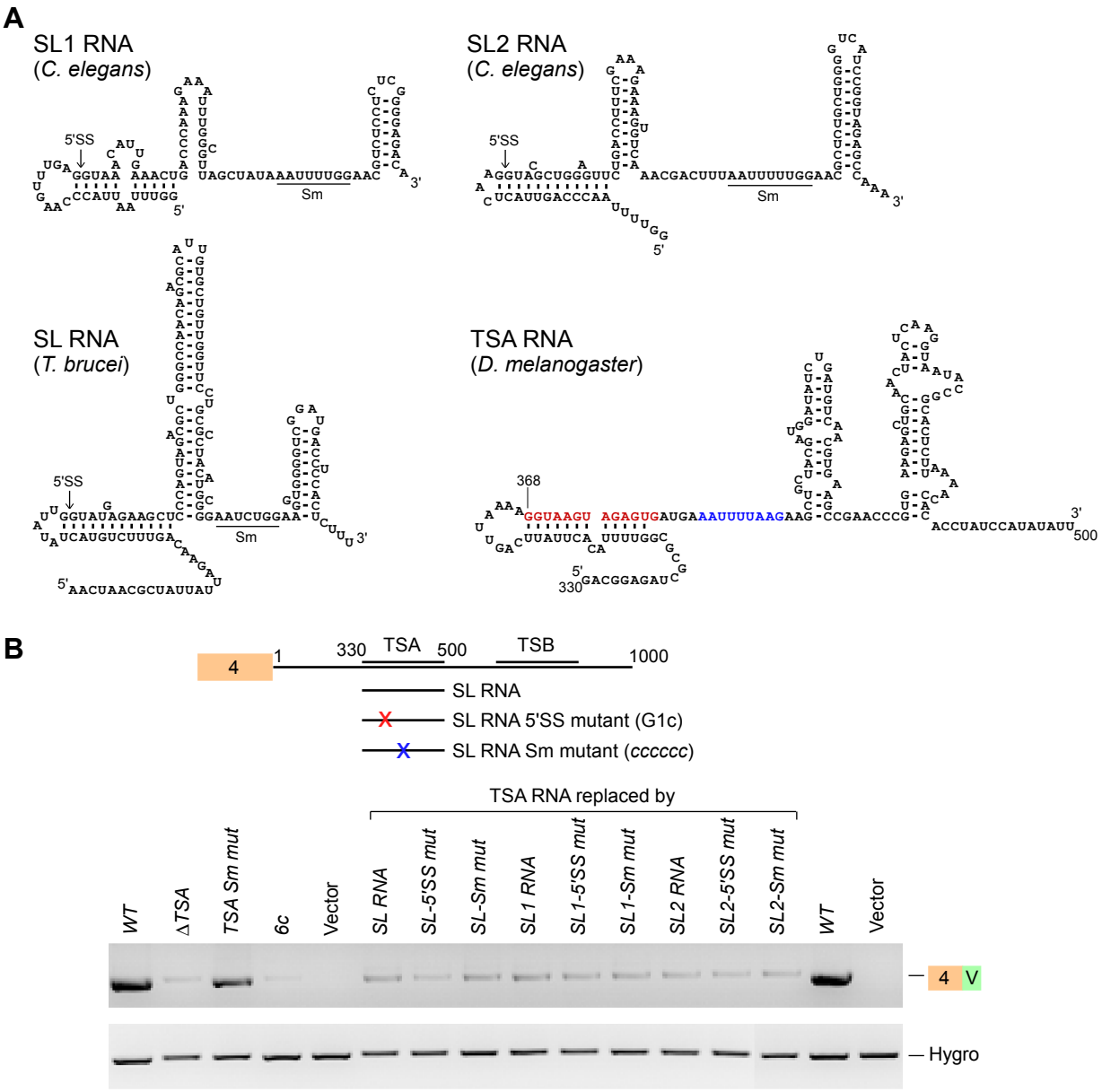

Figure S5. TSA RNA is not an analogue of SL RNA.

(A) Secondary structure of TSA RNA in *D. melanogaster* would be similar to SL RNAs. Three SL RNAs from *T. brucei* and *C. elegans* are adopted from (Bruzik *et al.,* 1988). TSA core motif and potential Sm-binding site in TSA RNA are colored.

(B) Replacing TSA RNA to SL RNAs do not significantly facilitates *trans*-splicing in S2 cells. For mutations in SL RNAs, the first nucleotide of 5'SS are mutated to cytosines, sequences of Sm-binding sites are mutated *cccccc*.

Figure S6. Sequence alignment of the last 5'-intron in *Iola* from *Drosophila* species. Available sequences of intron 8 of *Iola* genes from five *Drosophila* species are retrieved and aligned. The conserved TSA core-like motif is indicated by red rectangle.

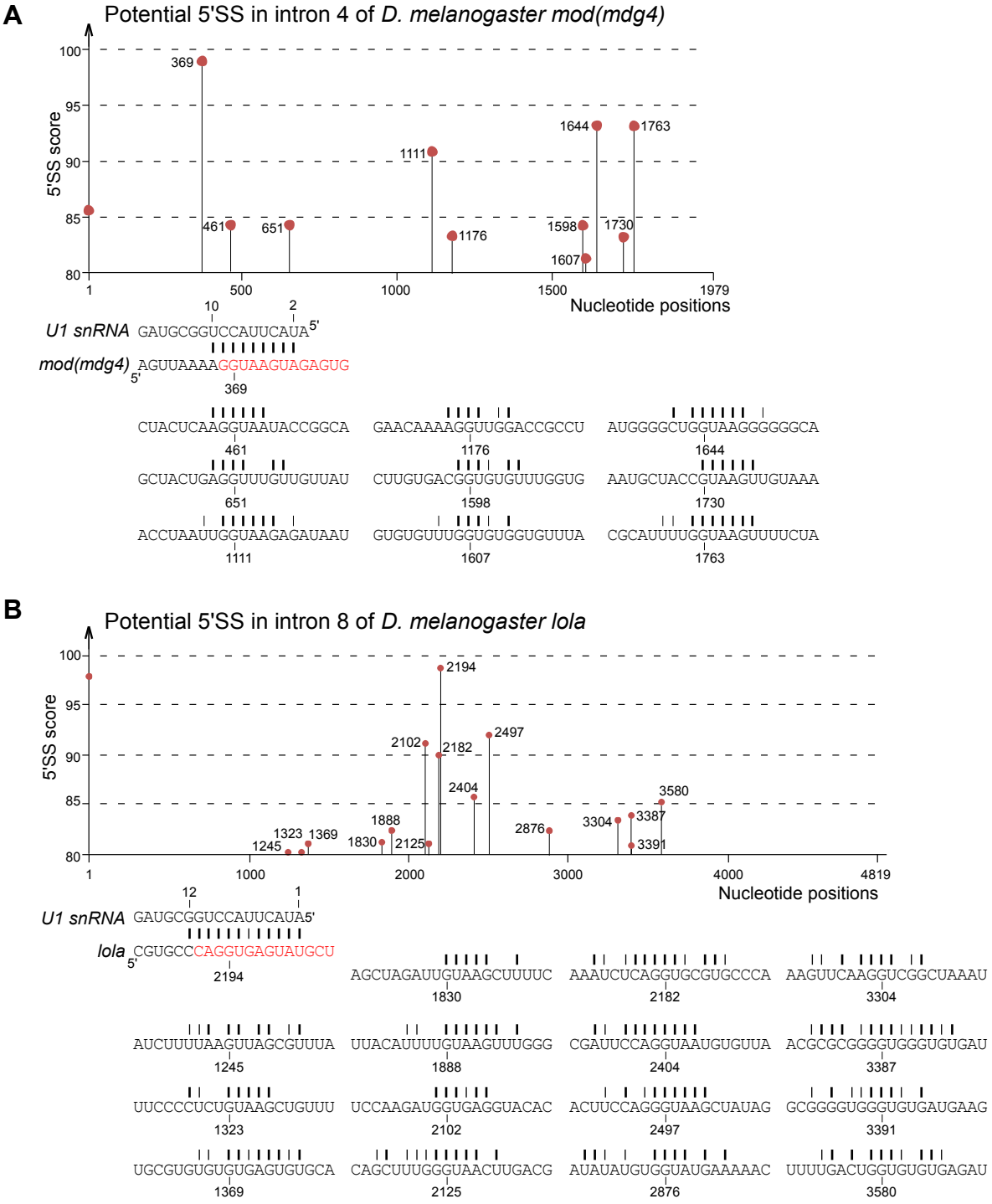

Figure S7. TSA core motif forms the strongest base-pairing with U1 snRNA. Potential 5'SS in the last 5'-introns of *mod(mdg4)* (A) and *lola* (B) are selected according to HSF scores (>80). For each gene, identified TSA core motif (red) forms the longest continuous base-pairs with the 5'-end of U1 snRNA, 9 bp in *mod(mdg4)* and 12 bp in *lola*.

Table S1. Proteins specifically associated with TSA and TSB RNAs.

	D. melanogaster proteins	# of Peptides	MW (KD)
***TSA* RNA associated**	**U1 snRNP**		
	U1-70K	74	52.9
	U1-A	4	24.5
	SmB	7	21.0
	SmD2	4	13.5
	SmD3	2	15.6
	SmE	6	11.1
	SmG	4	8.5
***TSB* RNA associated**	Zinc finger protein on ecdysone puffs	50	78.0
	Protein male-less	39	143.6
	PABP	18	69.9
	Histone H4	12	11.4
	dLarp	6	178.0
	ATP-dependent RNA helicase me31b	5	51.9
	Histone H2A/e	4	13.9
	Symplekin	4	132.0
	CPSF 160	3	164.6
	Testis-specific zinc finger protein topi	2	92.1
	HNRNP 40	2	36.2
	U2 snRNP		
	SF3B1	2	145.7
	SF3A3	1	58.4

Notes: Proteins associated with TSA, TSB and control RNAs are purified through amylose resin and identified by mass spectrometry (see experimental procedures). Specifically bound proteins are listed according to their absence in other two RNA affinity purification samples.

Table S2. TSA core-like motifs in other *Drosophila trans*-spliced genes.

Annotation_id	Genome_loc	Paired_seq
CG9940	X:13608279	UCAGGUGGGUGU
CG9812	2R:19306751	CAGGUGAGU
CG9432	2R:2625000	UGGGUAAGU
CG8529	2R:8338098	CUGGGUAAGU
CG7470	3L:21972476	UAGGUGAGUGU
CG6946	3R:7642917	CAGGUGGGU
CG6803	3R:11002087	CAGGUGGGUG
CG4700	2R:12415185	UCAGGUGAGUA
CG34376	3R:18276869	CAGGUAAGU
CG32556	X:17572059	GGUAAGUGU
CG32490	3R:122224	AGGUAAGUG
CG32045*	3L:9647673	CUGGGUAAGU
CG32045*	3L:9636530	GGUGAGUAU
CG18466*	3R:4872248	CCAGGUGGGUA
CG11326	2L:6690979	UCAGGUAAGUG
CG10851	3R:9488772	CAGGUAAGUG

Notes: *Trans*-splicing events in *Drosophila* are obtained from (McManus et al., 2010) and the TSA core-like motifs in 5'-*trans* introns are retrieved from sequences that contain pseudo-5'SS (score >80 by HSF) and form at least 9 continuous base-pairs with the 5'-end of U1 snRNA. The first two nucleotides of pseudo-5'SS are underlined. All sequences are analyzed on the reference of *dmel_r5.11* from Flybase.

*Experimentally confirmed events in (McManus et al., 2010).

Table S3. Major primers used in this study.

Primer	Sequence (5' to 3')	Notes
FYJ059	AATGCTGGACGGAGATCGCG	sgRNA target sequences for *mod(mdg4)*$^{\Delta TSA}$ (*Δ*340-487 nt)
FYJ062	CACCTATCCATATATTTTTC	
FYJ063	TGGTGATGCTAAGGCTACTG	sgRNA target sequences for *mod(mdg4)*$^{\Delta TSB}$ (*Δ*646-799 nt)
FYJ065	ATAACGCTGTGTTATAGTAA	
FYJ110	ACGAGCAATTCAGCTTGTGC	sgRNA target sequences for *mod(mdg4)*$^{\Delta com}$ (deletion from exon 2-72 nt to Intron 4-1018 nt)
FYJ112	AATTAAAGGGTGTCTGAAGA	
FYJ073	GGATCCATGGTGACATCATTTCAT	For amplification of the 6690-bp WT fragment in rescue construct
FYJ074	ATAGGCCGTGTAATTCGGATGCAA	
G-916	CCCGATGTGACTAGCTCTTTGCTGCAGGCCGTCCTATCCTCTGGTTCCGATAGTTGGCCAGGTTACCCGGCCAAAC	For cloning of U1 snRNA gene into the pMT-hygro vector, these two oligos were used for homologous recombination.
G-918	AGCCATACCACATTTGTAGAGGTTTTACTTGCTTTAAAAAACCTCCCACACCCCTCAACCACAAGCCCTTCACCTAC	