Supplemental information

Competing RNA pairings in complex alternative splicing of a 3′ variable region

Huawei Pan[§], Yang Shi[§], Shuo Chen, Yun Yang, Yuan Yue, Leilei Zhan, Lanzhi Dai, Haiyang Dong, Weiling Hong, Feng Shi, Yongfeng Jin[*]

Institute of Biochemistry, College of Life Sciences, Zhejiang University, Hangzhou, Zhejiang, ZJ3[1]0058, P. R. of China,

* Correspondence should be addressed to Yongfeng Jin: 0086-571-88206479(Tel); 0086-571-88206478(Fax); jinyf@zju.edu.cn (e-mail).

§These authors contributed equally to this work

This PDF file includes:

Supplemental Tables S1 to S4

Supplemental Figs. S1 to S12

Supplemental references

Supplemental Table S1. Species used in this study are listed according to order, family, genus and species. Three letter abbreviations used in sequence alignments are shown in parentheses following species names. *PGRP-LC, CG42235* and *Pip* genomic DNAs obtained and sequenced are indicated in the GeneBank column.

Order	Family	Genus	**Species** (abbreviation)	GenBank	Repeat copy
PGRP-LC					
Diptera	Drosophilidae	*Drosophila*	*D. melanogaster* (Dme)	MWIU0100012	3
Diptera	Drosophilidae	*Drosophila*	*D. simulans* (Dsi)	JPYS01000004	3
Diptera	Drosophilidae	*Drosophila*	*D. sechellia* (Dse)	AAKO0100003	3
Diptera	Drosophilidae	*Drosophila*	*D. yakuba* (Dya)	AAEU02000131	3
Diptera	Drosophilidae	*Drosophila*	*D. erecta* (Der)	AAPQ01006562	3
Diptera	Drosophilidae	*Drosophila*	*D. ananassae* (Dan)	AAPP01019340	3
Diptera	Drosophilidae	*Drosophila*	*D. pseudoobscura* (Dps)	AAFS01000557	3
Diptera	Drosophilidae	*Drosophila*	*D. pseudoobscura* (Dps)	AAFS01000557	3
Diptera	Drosophilidae	*Drosophila*	*D. persimilis* (Dpe)	CH479237	3
Diptera	Drosophilidae	*Drosophila*	*D. willistoni* (Dwi)	CH964168	3
Diptera	Drosophilidae	*Drosophila*	*D. mojavensis* (Dmo)	CH933809	3
Diptera	Drosophilidae	*Drosophila*	*D. virilis* (Dvi)	AANI01017383	3
Diptera	Drosophilidae	*Drosophila*	*D. grimshawi* (Dgr)	CH916366	3
Diptera	Muscidae	*Musca*	*M. domestica* (Mdo)	DQ372063	3
Diptera	Tephritidae	*Ceratitis*	*C. capitata* (Cca)	AOHK0200050	3
Diptera	Cecidomyiidae	*Mayetiola*	*M. destructor* (Mde)	AEGA01027600	2
Diptera	Culicidae	*Culex*	*C pipiens*(Cpi)	AAWU0100981	2
Diptera	Culicidae	*Aedes*	*A aegypti* (Aae)	CH478382	2
Diptera	Culicidae	*Anopheles*	*A gambiae* (Aga)	CM000356	3
Diptera	Culicidae	*Anopheles*	*A darlingi* (Ada)	ADMH0200080	3
Lepidoptera	Bombycidae	*Bombyx*	*B. mori (Bmo)*	DF090323	2
Coleoptera	Tenebrionidae	*Tribolium*	*T. castaneum (Tca)*	KQ971321	2
Hymenoptera	Apidae	*Apis*	*A. mellifera (Ame)*	GL630240	1
CG42235					
Diptera	Drosophilidae	*Drosophila*	*D. melanogaster* (Dme)	AE014297	5
Diptera	Drosophilidae	*Drosophila*	*D. simulans* (Dsi)	CM002913	5
Diptera	Drosophilidae	*Drosophila*	*D. sechellia* (Dse)	CH480828	5
Diptera	Drosophilidae	*Drosophila*	*D. yakuba* (Dya)	CM000160	5
Diptera	Drosophilidae	*Drosophila*	*D. erecta* (Der)	CH954182	5
Diptera	Drosophilidae	*Drosophila*	*D. ficusphila* (Dfi)	KB457506	5
Diptera	Drosophilidae	*Drosophila*	*D. eugracilis* (Deu)	KB465229	5
Diptera	Drosophilidae	*Drosophila*	*D. biarmipes* (Dbia)	KB462579	7
Diptera	Drosophilidae	*Drosophila*	*D. elegans* (Del)	KB458619	5
Diptera	Drosophilidae	*Drosophila*	*D. rhopaloa* (Drh)	KB452482	5
Diptera	Drosophilidae	*Drosophila*	*D. kikkawai* (Dki)	KB459690	5
Diptera	Drosophilidae	*Drosophila*	*D. ananassae* (Dan)	CH902617	6
Diptera	Drosophilidae	*Drosophila*	*D. bipectinata* (Dbip)	KB464355	5

Pip

Diptera	Drosophilidae	*Drosophila*	*D. melanogaster* (Dme)	AE014296	10
Diptera	Drosophilidae	*Drosophila*	*D. yakuba* (Dya)	CM000159	10
Diptera	Drosophilidae	*Drosophila*	*D. erecta* (Der)	CH954178	10
Diptera	Drosophilidae	*Drosophila*	*D. ficusphila* (Dfi)	KB457308	10
Diptera	Drosophilidae	*Drosophila*	*D. eugracilis* (Deu)	KB465074	10
Diptera	Drosophilidae	*Drosophila*	*D. biarmipes* (Dbia)	KB462895	10
Diptera	Drosophilidae	*Drosophila*	*D. elegans* (Del)	KB458607	10
Diptera	Drosophilidae	*Drosophila*	*D. rhopaloa* (Drh)	KB452050	10
Diptera	Drosophilidae	*Drosophila*	*D. kikkawai* (Dki)	KB459887	10
Diptera	Drosophilidae	*Drosophila*	*D. ananassae* (Dan)	CH902618	10
Diptera	Drosophilidae	*Drosophila*	*D. pseudoobscura* (Dps)	CH379070	10
Diptera	Drosophilidae	*Drosophila*	*D. Miranda* (Dmi)	CM001517	10
Diptera	Drosophilidae	*Drosophila*	*D. willistoni* (Dwi)	CH963876	10
Diptera	Drosophilidae	*Drosophila*	*D. mojavensis* (Dmo)	CH933809	8
Diptera	Drosophilidae	*Drosophila*	*D. virilis* (Dvi)	CH940647	9
Diptera	Drosophilidae	*Drosophila*	*D. albomicacs* (Dal)	JH862192	10

Supplemental Table S2. Primers used for RT-PCR and PCR analysis.

Primer	Sequence (5' to 3')	Primer	Sequence (5' to 3')
PGRP-LC			
DmPGRP-5-1	AAGGTACCGCTCCAAACACGAAGATCCGGCG	DmPGRP-5-2	AAGGTACCACCAATTTGTCCTTTTCTGCCC
DmPGRP-3-1	GCGAATTCGATAAAATAACCCCCGATAAAGTAA	DmPGRP-3-2	GCGAATTCGCTAAGCCCCTGGCAAACTAACCTC
DmPGRP-3-10	GCTAAGCATGTTAGTCTTTTTGT	DmPGRP-3-17	AGAGGATGTTGGATGCGCAA
DmPGRP-3-15	GAGGAAACCTAAAGCCATTTA	DmPGRP-3-13	CAATATAAACTATGACTCGCAG
DmPGRP-RT1	TAGTYCCAGCCACGCCCCTCGTA	DmPGRP-RT3	AGCAGYTGRCATGCCTCTAA
DmPGRP-3-a-s	CAGCAGACGAACTCGGAGCACA	BGH	TAGAAGGCACAGTCGAGGC
MT-F	CATCTCAGTGCAACTAAAGG		
DmPGRP-GZ-5-c1	TACCTGCAGCTCGTGGGCGGCGATGGCCGTGTG	DmPGRP-GZ-3-c1	CGTCAATTTCTTAAGACGAACGC
DmPGRP-GZ-5-c3	GTTCGTCTTAAGAAATTGACGCTAAGTGTTACTCGTCTTTTGCT	DmPGRP-5-M1-R	CAATGGATCCTAATAGGTACAGTTTTGTTTC
DmPGRP-3-M1-R	ATTAGGATCCATTGTAGACTTATTTGTCTGTAATCGTCGTCATCT	DmPGRP-5-M2-R	CATTCTAGACTTAGCCAGTGAATCGACCTCGTGCGT
DmPGRP-3-M2-R	GCTAAGTCTAGAATGGAGCCTTTTAGGCCACACTTTTGATTTAT	DmPGRP-5-M3-R	AGGATCCATTATAGACTTACACGAATATCTTACAAAAGATAC
DmPGRP-3-M3-R	ATTCGTGTAAGTCTATAATGGATCCTTATTGAAAAACTTTAATTTTTCT	DmPGRP-3-R6	AGCGGGAAACAAAACTGTACCTATTGCCTAACCGAGTGGAATATTTTGTCTGTAATC
DmPGRP-5-R6	CTCGGTTAGGCAATAGGTACAGTTTTGTTTCC	DmPGRP-3-R4	AGCGGGAAACAAAACTGTACCTATTGTCTAACTAATTGGAATATTTTGTCTGTAATC
DmPGRP-5-R4	ATTAGTTAGACAATAGGTACAGTTTTGTTTCC	DmPGRP-3-R3	AGCGGGAAACAAAACTGTACCTATTGCTAAATTAATTGGAATATTTTGTCTGTAATC
DmPGRP-5-R3	ATTAATTTAGCAATAGGTACAGTTTTGTTTCC	DmPGRP-3-R2	AGCGGGAAACAAAACTGTACCTATTGTTTAACTGATTGGAATATTTTGTCTGTAATC
DmPGRP-5-R2	ATCAGTTAAACAATAGGTACAGTTTTGTTTCC	DmPGRP-3-FR4	ACAAAACTGTACCTATTGTTGAATTAAGTGAAACATTTTGTCTGTAATCGTCGTCAT
DmPGRP-5-FR4	ATGTTTCACTTAATTCAACAATAGGTACAGTT	DmPGRP-3-FR3	ACAAAACTGTACCTATTGTTGAATTGAGTAGAATACTTTGTCTGTAATCGTCGTCAT
DmPGRP-5-FR3	GTATTCTACTCAATTCAACAATAGGTACAGTT	DmPGRP-3-R7	AGCGGGAAACAAAACTGTACCTATTGTTGAATTAATTGGAATATTTTGTCTGTAATC
DmPGRP-5-R7	ATTAATTCAACAATAGGTACAGTTTTGTTTCC	DmPGRP-3-R8	AGCGGGAAACAAAACTGTACCTATTGTTAAATTGATTGGAATATTTTGTCTGTAATC
DmPGRP-5-R8	ATCAATTTAACAATAGGTACAGTTTTGTTTCC	DmPGRP-3-R9	AGCGGGAAACAAAACTGTACCTATTGCTGAATTGATTGGAATATTTTGTCTGTAATC
DmPGRP-5-R9	ATCAATTCAGCAATAGGTACAGTTTTGTTTCC	DmPGRP-3-R10	AGCGGGAAACAAAACTGTACCTATTGTTAAATTAATTGGAATATTTTGTCTGTAATC
DmPGRP-5-R10	ATTAATTTAACAATAGGTACAGTTTTGTTTCC	DmPGRP-3-R11	AGCGGGAAACAAAACTGTACCTATTG

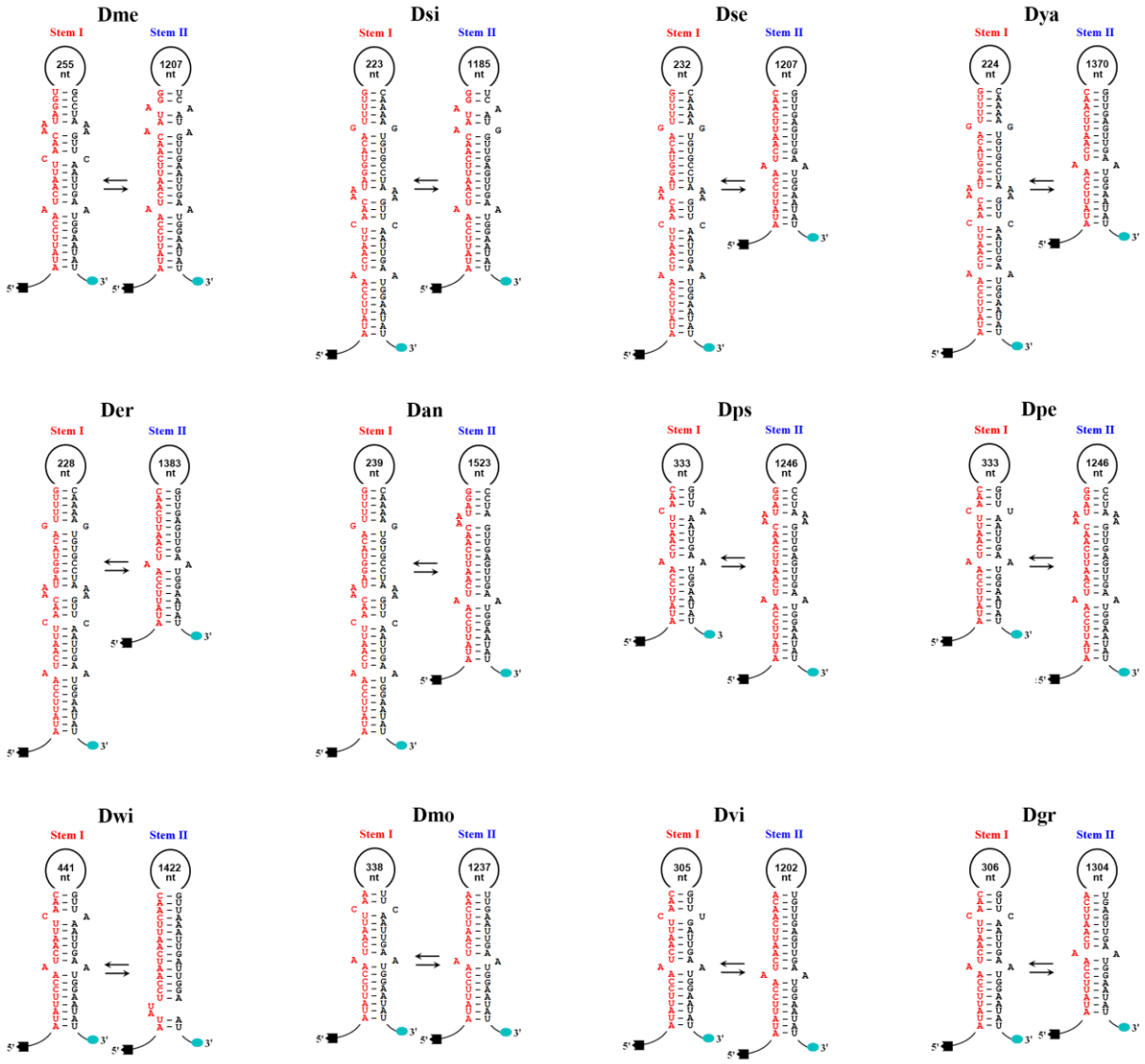
			CTGAATTAATTGGAATATTTTGTCTGTAATC
DmPGRP-5-R11	ATTAATTCAGCAATAGGTACAGTTTTGTTTCC	DmPGRP-5-R12	AGCGGGAAACAAAACTGTACCTATTGCTAAATTGATTGGAATATTTTGTCTGTAATC
DmPGRP-5-R12	ATCAATTTAGCAATAGGTACAGTTTTGTTTCC	DmPGRP-3-R13	AGCGGGAAACAAAACTGTACCTATTGTTTAATTGATTGGAATATTTTGTCTGTAATC
DmPGRP-5-R13	ATCAATTAAACAATAGGTACAGTTTTGTTTCC	DmPGRP-3-R14	AGCGGGAAACAAAACTGTACCTATTGCTAAGTCAATTGGAATATTTTGTCTGTAATC
DmPGRP-5-R14	ATTGACTTAGCAATAGGTACAGTTTTGTTTCC	DmPGRP-3-R15	AGCGGGAAACAAAACTGTACCTATTGCTAACTAAATTGGAATATTTTGTCTGTAATC
DmPGRP-5-R15	ATTTAGTTAGCAATAGGTACAGTTTTGTTTCC		
pip			
Dmpip-5-1K	CGGGGTACCAGCACGATGACATCAACACCCTG	Dmpip3-3-K	GAAACTCAGTTGGTACCACTCGTC
Dmpip-M1-5-1	CAAGTTGAGGAATCACTTACTAATGCATTTGGCCTACAGCATTCAGC	Dmpip-M1-3-1	TTAGTAAGTGATTCCTCAACTTGCTACCGTTGATGGACATTCATCG
Dmpip3-5-B	GCTGGTCAAGGATCCGATAGACC	Dmpip-3-1A	AGGGGGCCCCTTAAAATATTCATTTATTAGTTTATACAGTG
Dmpip-M3-5-1	TTAGTAAGTGATTCCTCAACTTGGTAAATTTGATGATACACTTAAATTGAAC	Dmpip-M3-3-1	CAAGTTGAGGAATCACTTACTAAGCATTTGGGGCAATGGAAAACTATTC
Dmpip-mini3-RT	GCCARATATTGYTTGT	Dmpip-mini3-3	AAGAAYCGKGGWACRTATTT
CG42235			
CG42235F5.1 SpeI	GGGACTAGTCTTCGTGTCTGGAATCTCGCTTCTGGG	CG42235F3.1 XhoI	GGGCTCGAGCATGGAACTACTTTCTACGTCTTC
CG42235PCR	AGCCAACCCATGAAGAAGAAG	CG42235Fan	AACTCATAATCRCAGCCTTC
DmCG42235-M5-1	CATGCCATGGGTACATAGATATTACAATGGGATA	DmCG42235-M3-1	CATGCCATGGACTGTTTAAGTACGTATTATAGACCATT
DmCG42235-M5-2	CATGCCATGGACTGTTTTCATATATGTAAATGTTAAAAG	DmCG42235-M3-2	CATGCCATGGGATAAAACTTAAAACAACCCAAAG
DmCG42235-5-2	CCGCGTAAGTTTGCGATATCC	DmCG42235-3-2	GGATATCGCAAACTTACGCGG

Supplemental Table S3. Primers used for amplifying fragments for *in vitro* transcription for RNA interference.

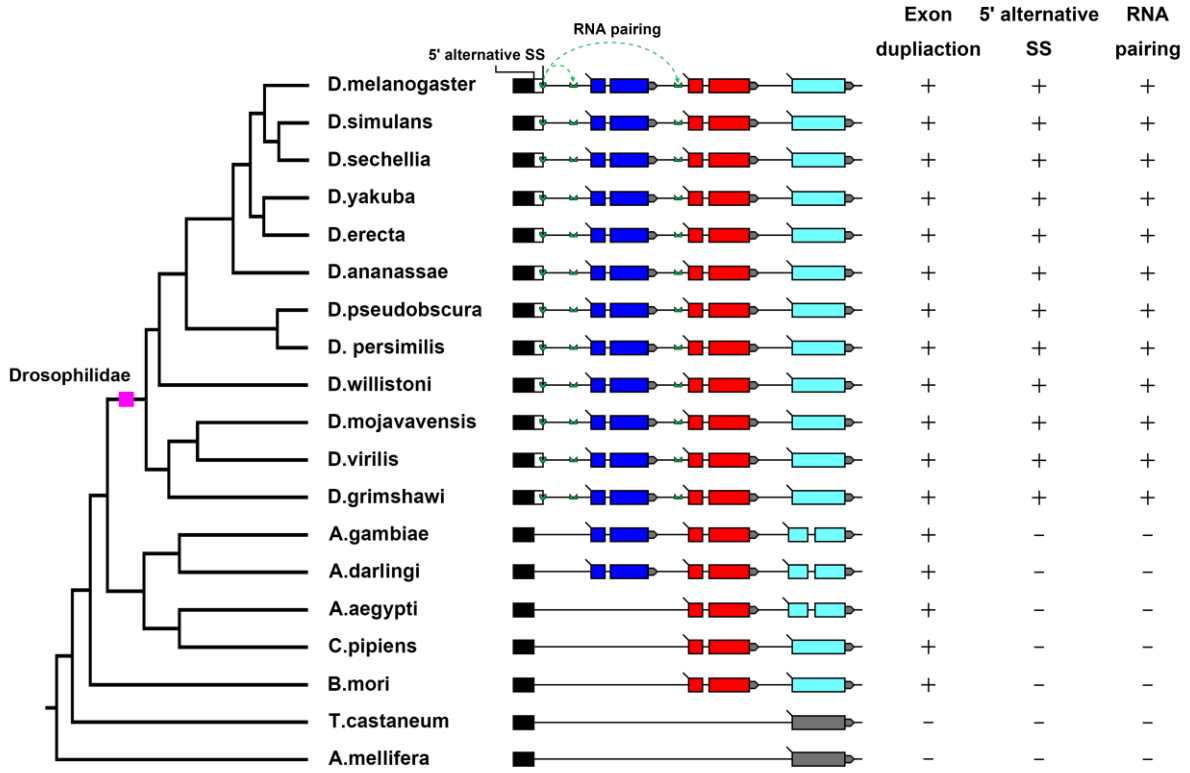
Primer	Sequence (5' to 3')	Primer	Sequence (5' to 3')
B52-i-5	taatacgactcactatagggGACCGCAATAACGAGAGCAT	CG1316-i-3	taatacgactcactatagggATGGTCTTGCCGTTCATCTC
B52-i-3	taatacgactcactatagggCCGGTGTACAAGTCGCAGTA	eIF2α-i-5	taatacgactcactatagggACTTTAACATGGCCCTGACG
U2af38-i-5	taatacgactcactatagggGTGCTCTCGCATCCACAACAAACC	eIF2α-i-3	taatacgactcactatagggCAGGCGGTCTTCTGGTAGAG
U2af38-i -3	taatacgactcactatagggCTCACGCGAGATGGGCTTCAAGTG	Hsc70-4-i-5	taatacgactcactatagggAAGGAACTGGAGGGTGTGTGC
U2af50-i-5	taatacgactcactataggCAAGCGTTGTGCCGGATACACCACA	Hsc70-4-i-3	taatacgactcactatagggTGTCGTTTGACCCGTTTGTA
U2af50-i -3	taatacgactcactataggGTAATCGTGCGGACGCCTAATCTTT	Ote-i-5	taatacgactcactataggTCCGAGGAGCCCGTTGCCGC
Eflα48D-i-5	taatacgactcactatagggTCCACTCCTCATCCACTTCC	Ote-i-3	taatacgactcactataggCCTTGGGCTGCAGTACGCAC
Eflα48D-i-3	taatacgactcactatagggGAGTGCTCTTTTACGGGCTG	pAbp-i-5	taatacgactcactataggGCTGAGGCGGCCGTTCAGGC
Syp-i-5	taatacgactcactataggTGGATCGCCCCAATAAAAAC	pAbp-i-3	taatacgactcactataggCGCTAGCCGCGTTGGGCTGG
Syp-i-3	taatacgactcactataggTGCAGATCGTCGATCAGA	caz-i-5	taatacgactcactataggCTTCAATGACAACAATGGCG
SmB-i-5	taatacgactcactatagggCGGCAAGAACAACAAAATGA	caz-i-3	taatacgactcactataggGAGCGATAACCGTGTGTGTG
SmB-i-3	taatacgactcactatagggGCGGACAAGTTGATAGGCAT	CG6937-i-5	taatacgactcactatagggGTACAAGGATCTGGGCATCGACTTC
Lam-i-5	taatacgactcactataggAACGCAGCACAACACAACTC	CG6937-i-3	taatacgactcactatagggGGTGGCGGCATTCTTGTTGCTCTTA
Lam-i-3	taatacgactcactataggGCAACATGCCTCTCCAGTTT	Hrb87F-i-5	taatacgactcactatagggTGTCAGACCAGGCAGACAAG
pont-i-5	taatacgactcactataggTTCCGCTAGATCTGCTCGAT	Hrb87F-i-3	taatacgactcactatagggAAAACTAAGCGTTCGGGGTT
pont-i-3	taatacgactcactataggTTTTACTGGGAAATCCGTGC	Imp-i-5	taatacgactcactataggAAATTTGCCTCAAAATACTCGC
CG18178-i-5	taatacgactcactatagggCGAATCCGATGAAGATGATGGTGAG	Imp-i-3	taatacgactcactataggACACATGAAGCCTTCTTCGC
CG18178-i-3	taatacgactcactatagggCAACGAACTGGTCTTTGGCAGATTG	La-i-5	taatacgactcactataggCGAGGAGGAGGAGAAGGAGT
sqd-i-5	taatacgactcactataggTTCCCGAAGATCCTTCAATG	La-i-3	taatacgactcactataggGCTTTCTGCTGAGGGAAACA
sqd-i-3	taatacgactcactataggTTGTAGCTGTTGGTTGCTGG	mod-i-5	taatacgactcactataggGTGCATGAACGGACTTTAAACC
CG1316-i-5	taatacgactcactatagggAGCGAGAAAAGTCGAAGTGC	mod-i-3	taatacgactcactataggGGGGACTACTGTCTCCTCTTCC

Supplemental Table S4. Primers used to test the efficiency of RNA interference.

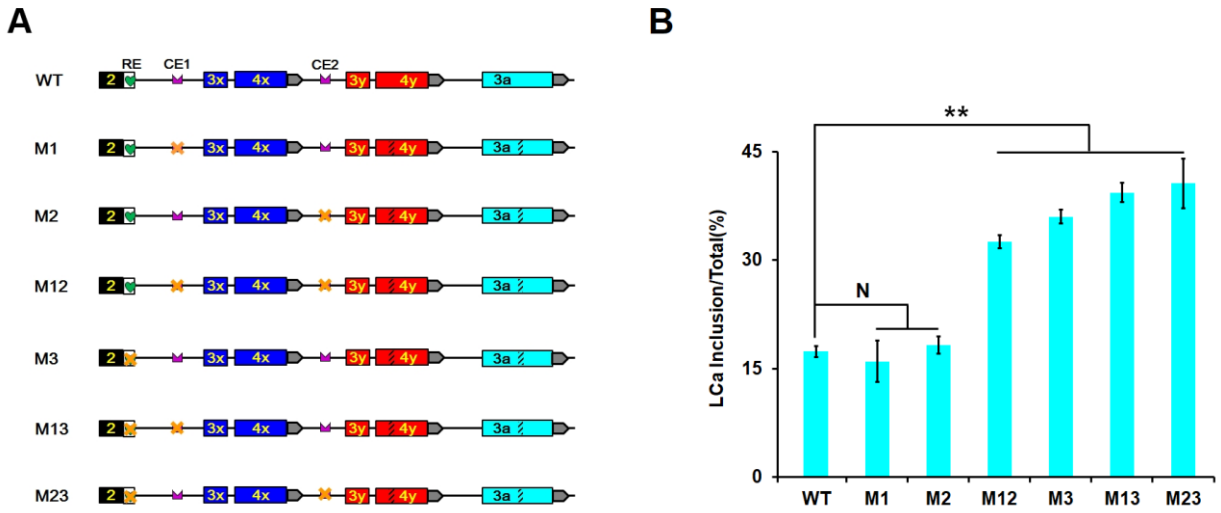
Primer	Sequence (5' to 3')	Primer	Sequence (5' to 3')
B52-t-5	GAAGTTGGATGACACCGAGC	CG1316-t-3	CCTGGTTGACGCAGTTGCTC
B52-t-3	CTGGCGACTTGGACTTGGAG	eIF2α-t-5	CGCCCGAAGATGTCGAGAAG
U2af38-t-5	CCGCGGATGGCTCCCATCTG	eIF2α-t-3	GGATCTTGACAGTGGGCGAG
U2af38-t-3	CGCCCATCTCGTACTGCCGA	Hsc70-4-t-5	TCCTCTGTCCCTGGGTATCG
U2af50-t-5	CCCGACAGGCGCGTCGCCTG	Hsc70-4-t-3	GCGCTCGATGTCCTCCTTGG
U2af50-t-3	GCCATCGAATGCCATGGCCT	Ote-t-5	CTGGATGGCAACAAGGTGGC
Ef1α48D-t-5	AGAAGGAGGCCCAGGAGAT	Ote-t-3	GGACAATCAGGGAGTTGACC
Ef1α48D-t-3	GCAGTTTCGAACTTCCACAG	pAbp-t-5	CGGCAAGAAGGTCTACGTGG
Syp-t-5	CCGCCCGTACAGTTTCACGAC	pAbp-t-3	CAGCTCCTGCTGGCGTTCGG
Syp-t-3	ACCATGTGGATTCGTCCGGC	caz-t-5	GCCTCCAAGGGAGAGGCCAC
SmB-t-5	GCGAATCGTGCTGCAGGACTC	caz-t-3	AGCGTCCGCCGCCACCGCCT
SmB-t-3	GCGACCGATTCCCGGACCAGG	CG6937-t-5	GAAATCTATGTCTCGGATTTAACCA
Lam-t-5	GCGGCTAATCAACGAGAAAG	CG6937-t-3	CTTCTTCTGCACACCACCAG
Lam-t-3	CAGCCGGGAGGAGGATGTGTG	Hrb87F-t-5	GGAGCCACGGTAAAGAAGCT
pont-t-5	TCGGAGGTATTCAGCAATGA	Hrb87F-t-3	TGCTTGGCAATAGCCTTCTT
pont-t-3	TTTGCAAGGCATCAAAGATG	Imp-t-5	GCGAATCATTACGGTCAAGG
CG18178-t-5	CAAGGTGAAGGCTCTGCATC	Imp-t-3	CCTGTAGGTCATTGGGGAAG
CG18178-t-3	GGTACGTTTGTTTACGATGG	La-t-5	AGGGCCTCGTGGAGATTA
sqd-t-5	TCAATGTCAAGACAGATCCC	La-t-3	GGCTTCGCGAATTTCCTC
sqd-t-3	CTCCTCATCGCTGATCTCTG	mod-t-5	GCTGTTCTGCAGGCTAAGCC
CG1316-t-5	CCAAAGGGTTCTACACGCAC	mod-t-3	GGTAAGTTCACTGGTACGTG



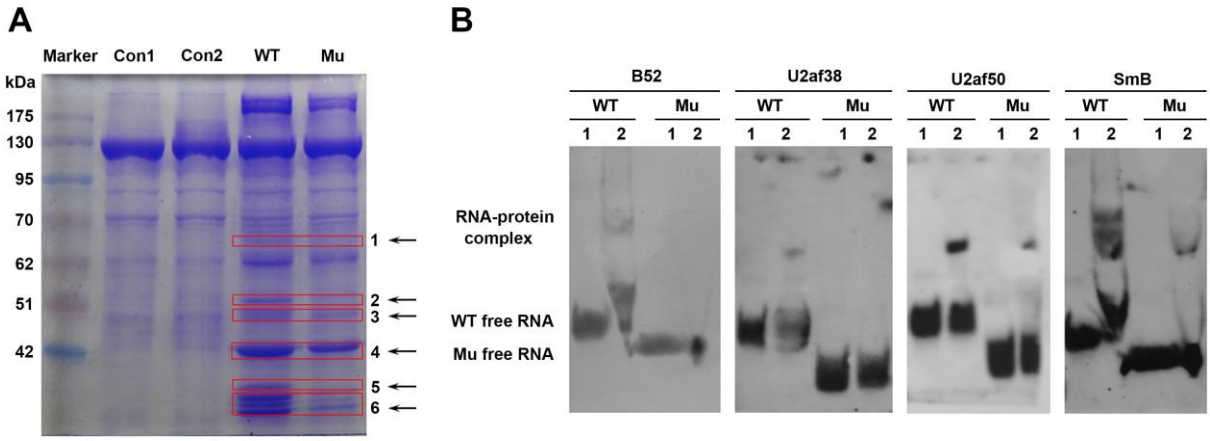
Supplemental Figure S1. The competition between intra- and inter-intron RNA pairing is conserved in *Drosophila PGRP-LC* 3' tandem cluster. Abbreviations used see Supplemental Table S1. The intronic RNA pairings between the selector sequences and the docking site were predicted using Mfold program(Zuker 2003).



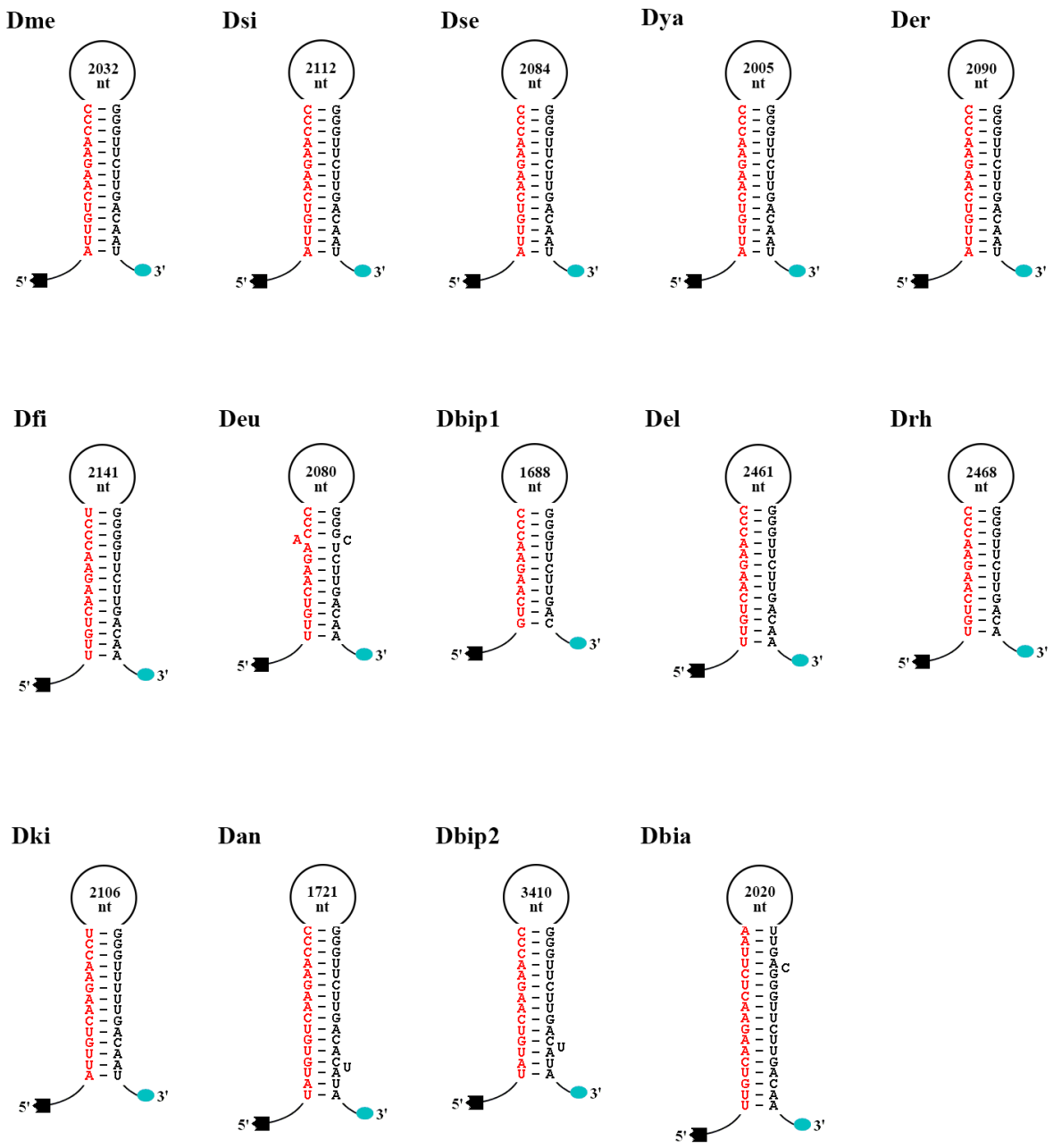
Supplemental Figure S2. Correlations of RNA pairings with isoform selection in insect *PGRP-LC*s. The structure organization of *PGRP-LC* gene was shown from 16 Diptera species, 1 Lepidoptera, Coleoptera and Hymenoptera species, respectively. Evolutionary dynamics of *PGRP-LC* alternatively spliced variants were shown in Diptera Lepidoptera, Coleoptera and Hymenoptera species. Exons and introns are not drawn to scale. This result shows the correlation between the emergence of 5' alternative splice sites and the occurrence of RNA pairing.



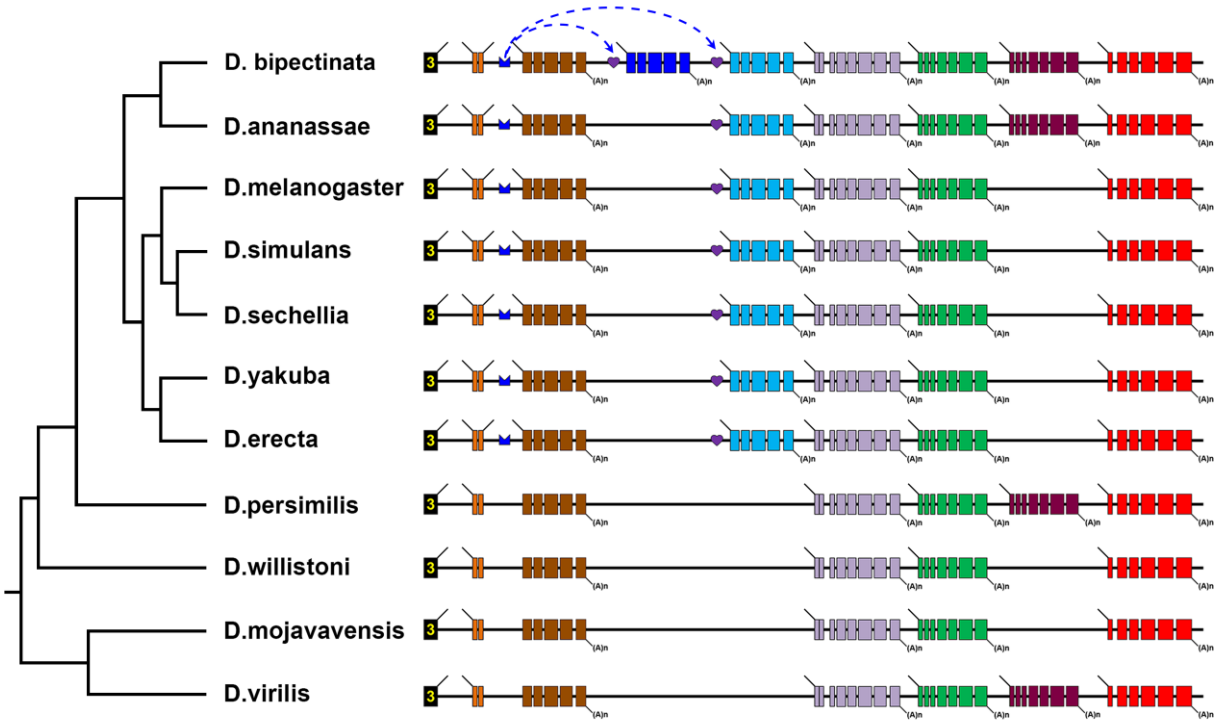
Supplemental Figure S3. The effects of the docking site and RNA pairings on selection of the *LCa* isoform. (**A**) Overview of the minigene constructs of *D. melanogaster* used to test the effects of RNA pairing on splicing. Constitutive exons (in black boxes), alternative exons (in blue boxes), conserved intronic elements CE1 and CE2 (yellow) and intron (line). Symbols used are the same as in Figure 1. The splicing pattern of M1, in which the *LCy* and *LCa* exon sequences have been mutated for RT-PCR detection convenience (marked by slash), was indistinguishable from that of wild-type. (**B**) Effects of RNA pairing modulated the *LCx* inclusion by mutated analysis. Quantitation of the data was shown below. Data were expressed as percentage of mean means ± SD from three independent experiments, (**) P<0.01.



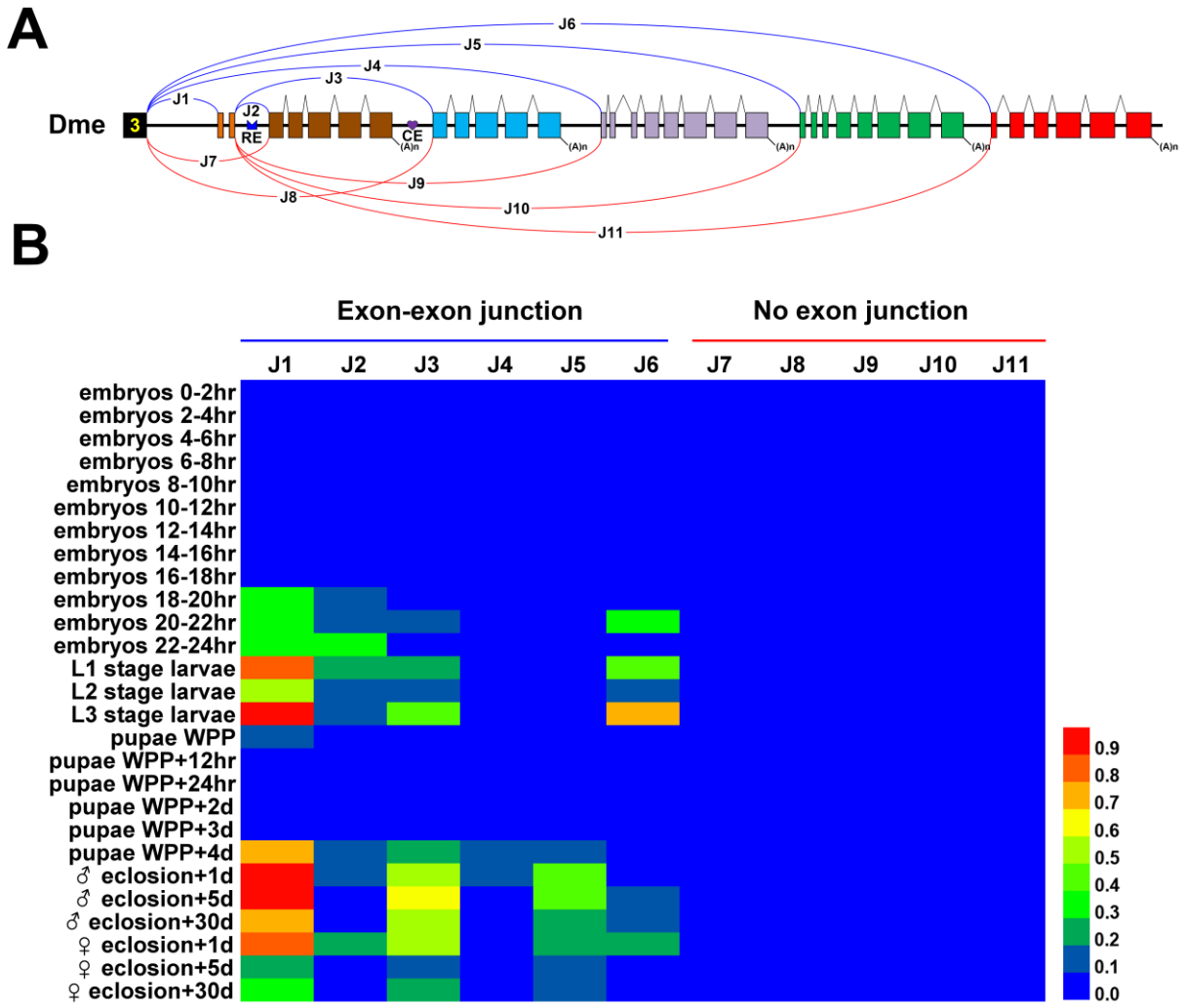
Supplemental Figure S4. The identification of docking site RNA-binding proteins. (*A*) Staining of affinity-purified proteins. Analyzed by mass spectrometry, docking site specifically associated proteins are obtained and screened. **(*B*)** Electrophoretic mobility shift assay of target proteins. Biotin-labeled WT or Mu RNA was incubated with candidate proteins, and complexes were resolved by non-denaturing PAGE. Biotin-labeled WT or Mu RNAs are the same as RNA used for RNA pull-down analysis.



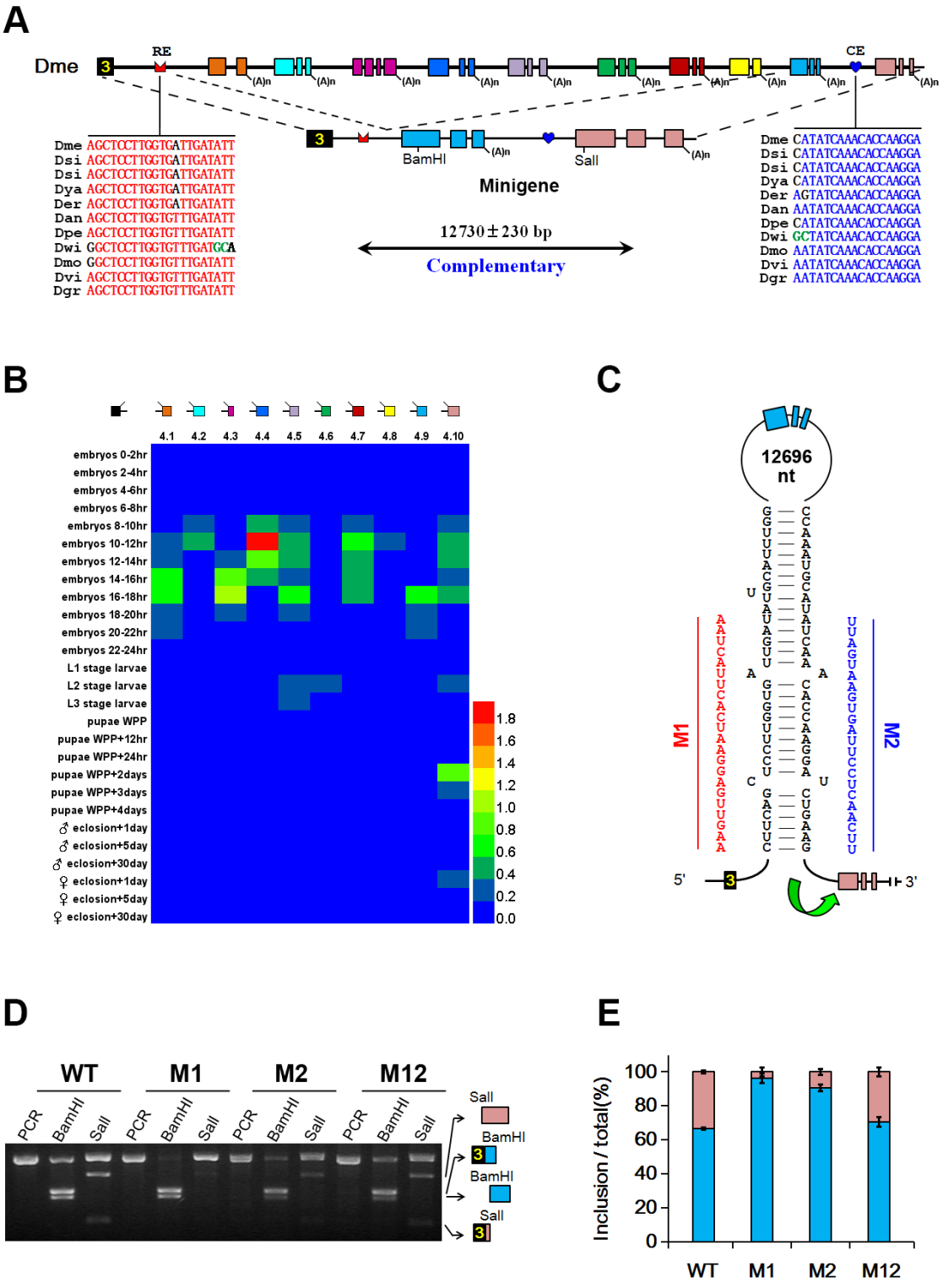
Supplemental Figure S5. The inter-intronic pairing interaction is conserved in *Drosophila CG42235* 3' tandem cluster. Abbreviations used see Supplemental Table S1. The intronic RNA pairings between the selector sequences and the anchor site were predicted using Mfold program[1].



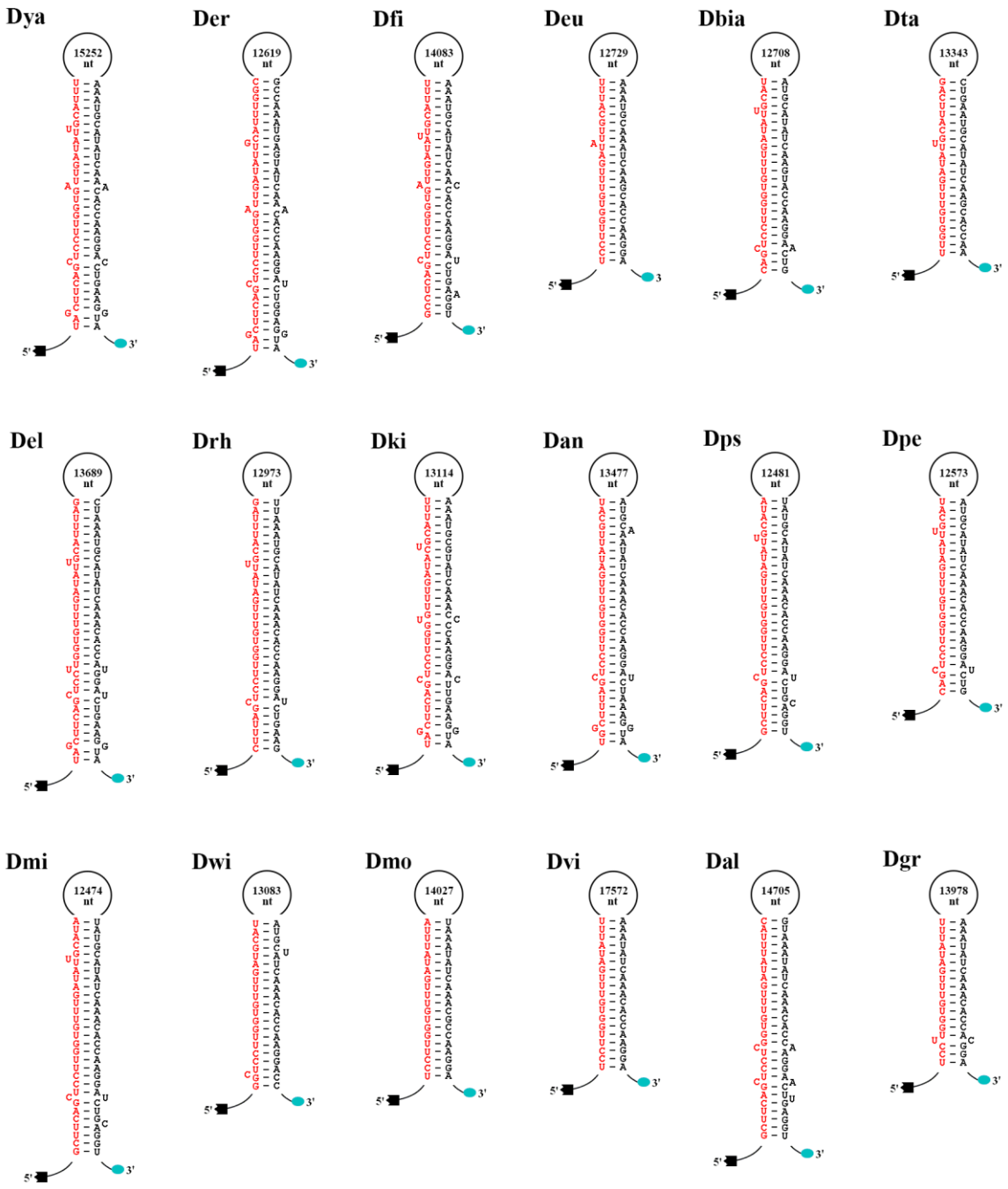
Supplemental Figure S6. Evolutionary dynamics of 3' variable region in *Drosophila* ***CG42235*s.** The structure organization of *CG42235* gene from 11 *Drosophila* species was shown. Exons and introns are not drawn to scale. 4-7 tandemly arrayed domains with 5-7 exons are alternatively spliced to the common exon 3 of the *CG42235* gene in mutually exclusive mode. This result shows that the evolutionary shift from the 3′ constitutive site to an alternative one was concomitant with the creation and expansion of RNA pairing.



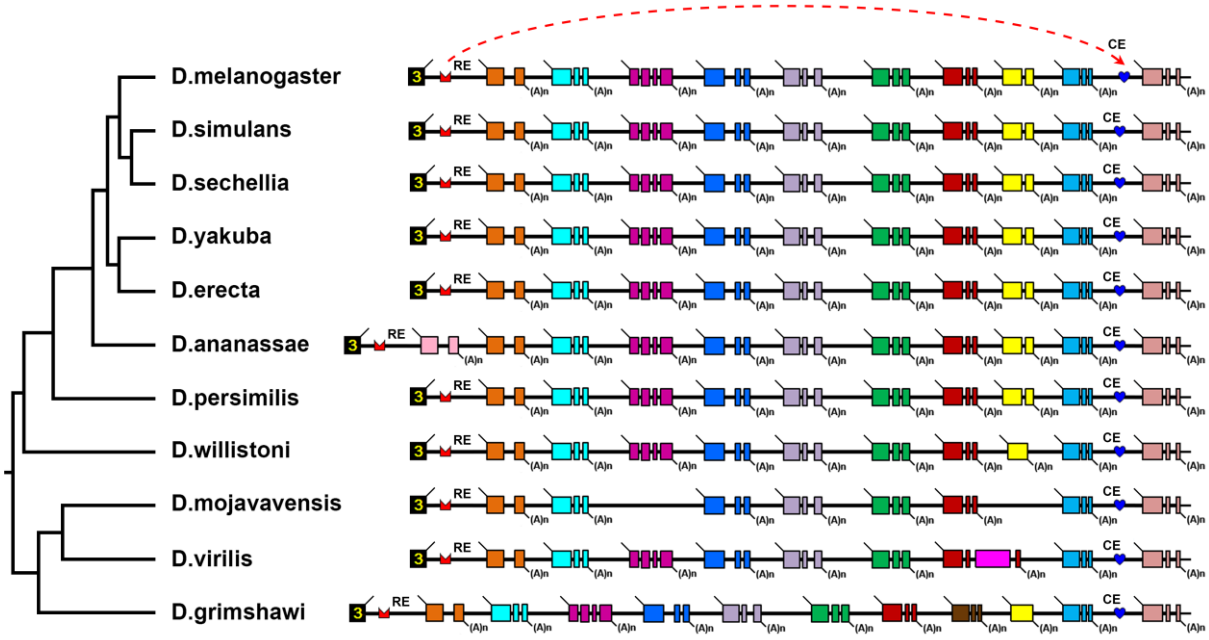
Supplemental Figure S7. The biased expression of 3' variable isoforms of *Drosophila CG42235*. (*A*) Overview of 3' variable region of the *Drosophila CG42235*. Constitutive exon (in black box), alternative exons (in colour box), and introns (line). This gene contained five tandemly arrayed cassettes with 5-8 exons at 3' variable region, which are alternatively spliced to the common exon 3. **(*B*)** The heatmap of the expression of the splicing isoforms. Based on the exon-exon junction reads from RNA-seq data, we calculated the splice isoforms in different developmental stages. These data indicate that the 3′ splice site of exon 2 can be spliced with the most proximal alternative 5′ splice site and the three most distal alternative 5′ splice sites, and the 3′ splice site of exon 3 can be spliced with the two most proximal alternative 5′ splice sites (J1-6, Left), while the 3′ splice site of exon 2 cannot be spliced with the two intermediate splice sites, and the 3′ splice site of exon 3 cannot be spliced with the three most distal sites (J7-11, Right).


A
Dme
RE
CE
(A)n
BamHI
SalI
Minigene
12730±230 bp
Complementary
Dme AGCTCCTTGGTGATTGATATT
Dsi AGCTCCTTGGTGATTGATATT
Dsi AGCTCCTTGGTGATTGATATT
Dya AGCTCCTTGGTGATTGATATT
Der AGCTCCTTGGTGATTGATATT
Dan AGCTCCTTGGTGTTTGATATT
Dpe AGCTCCTTGGTGTTTGATATT
Dwi GGCTCCTTGGTGTTTGATGCA
Dmo GGCTCCTTGGTGTTTGATATT
Dvi AGCTCCTTGGTGTTTGATATT
Dgr AGCTCCTTGGTGTTTGATATT
Dme CATATCAAACACCAAGGA
Dsi CATATCAAACACCAAGGA
Dsi CATATCAAACACCAAGGA
Dya CATATCAAACACCAAGGA
Der AGTATCAAACACCAAGGA
Dan AATATCAAACACCAAGGA
Dpe CATATCAAACACCAAGGA
Dwi GCTATCAAACACCAAGGA
Dmo AATATCAAACACCAAGGA
Dvi AATATCAAACACCAAGGA
Dgr AATATCAAACACCAAGGA
B
4.1 4.2 4.3 4.4 4.5 4.6 4.7 4.8 4.9 4.10
embryos 0-2hr
embryos 2-4hr
embryos 4-6hr
embryos 6-8hr
embryos 8-10hr
embryos 10-12hr
embryos 12-14hr
embryos 14-16hr
embryos 16-18hr
embryos 18-20hr
embryos 20-22hr
embryos 22-24hr
L1 stage larvae
L2 stage larvae
L3 stage larvae
pupae WPP
pupae WPP+12hr
pupae WPP+24hr
pupae WPP+2days
pupae WPP+3days
pupae WPP+4days
♂ eclosion+1day
♂ eclosion+5day
♂ eclosion+30day
♀ eclosion+1day
♀ eclosion+5day
♀ eclosion+30day
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
C
12696 nt
M1
M2
5'
3'
D
WT
M1
M2
M12
PCR
BamHI
SalI
E
Inclusion / total(%)
100
80
60
40
20
0
WT
M1
M2
M12

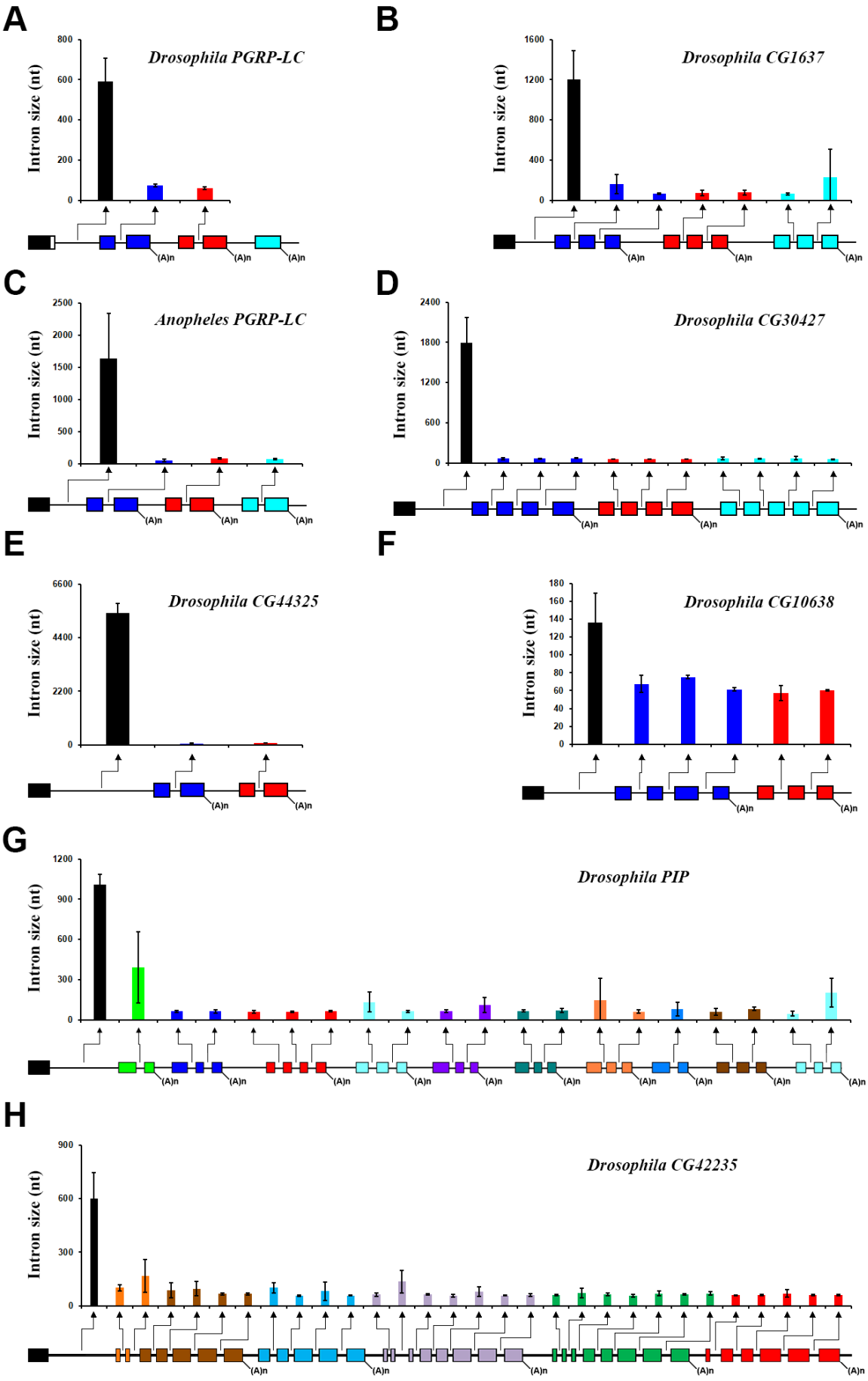
Supplemental Figure S8. RNA pairings were validated at the 3' **variable region of *Drosophila pip* pre-mRNA. (*A*)** Overview of 3' variable region of *Drosophila pip*. Symbols used are the same as in Fig. 1. This gene contained ten tandemly arrayed cassettes with 2-4 exons at 3' variable region, which are alternatively spliced to the common exon 3. Predicted RNA pairing interactions see Supplemental Fig. S8. **(*B*)** The heatmaps of the expression of the splice isoforms in different developmental stages. **(*C*)** Schematic diagrams of minigene constructs to verify the predicted RNA pairings of *pip* pre-mRNA. Mutations introduced into dsRNA are indicated on the left or right mutated sequences (M1, M2). A green arrow is depicted as activating the inclusion of the proximal exon. **(*D*)** Effects of RNA pairing on the inclusion of the 3' variable exon is validated by disruptive single mutations (M1, M2) and compensatory double mutations (M12). **(*E*)** Effects of disruptive and compensatory mutations on the inclusion of alternative cassette. Data were expressed as percentage of mean means ± SD.



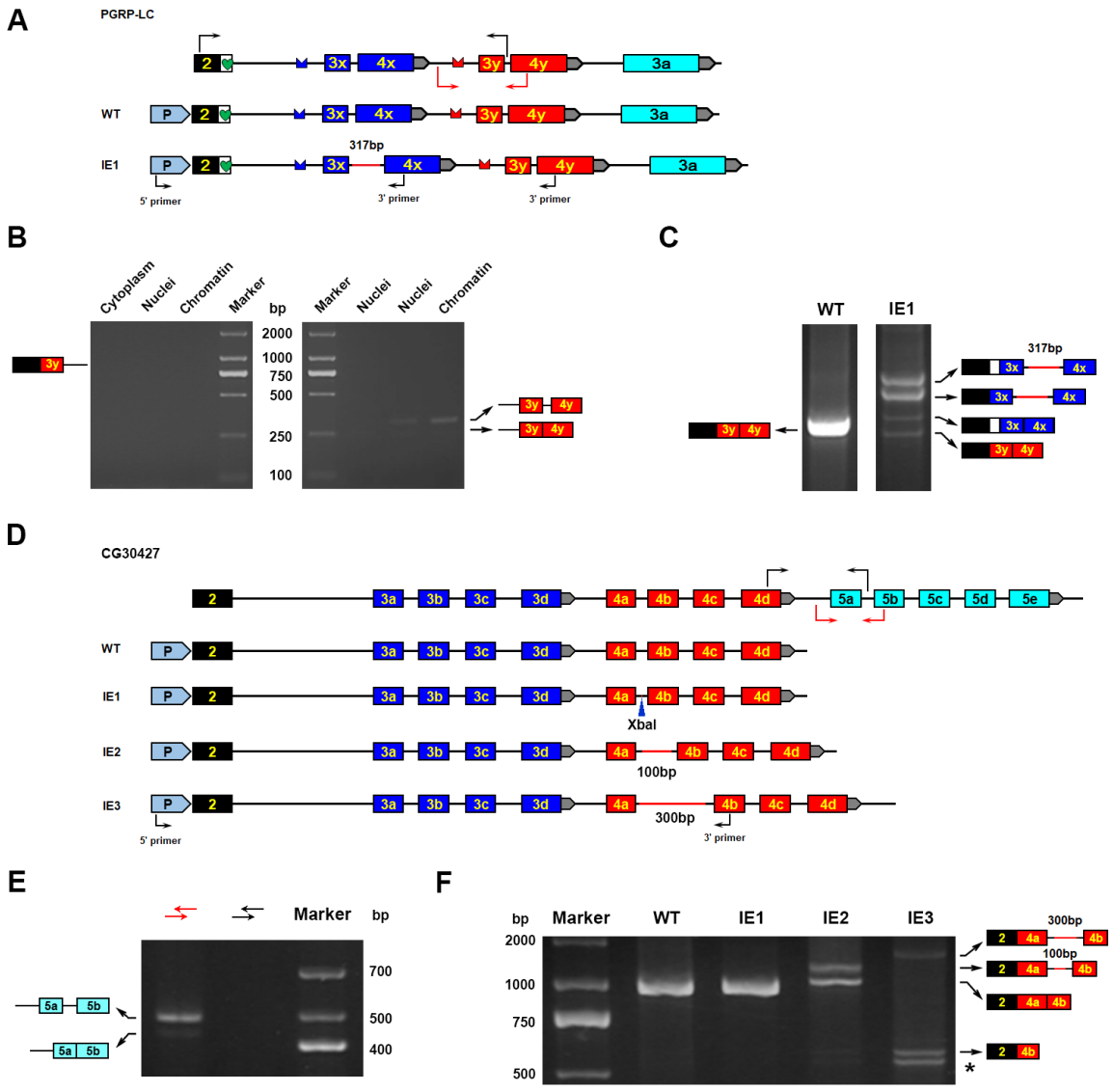
Supplemental Figure S9. The intronic RNA pairing is conserved in *Drosophila pip* 3' tandem cluster. Abbreviations used see Supplemental Table S1. The intronic RNA pairings between the selector sequences and the docking site were predicted using Mfold program[1].



Supplemental Figure S10. Evolutionary dynamics of 3' variable region of *Drosophila pip*. Exons and introns are not drawn to scale. 8-11 tandemly arrayed domains with 2-4 exons at the 3' variable region are alternatively spliced to the common exon 3 of the *Pip* gene in mutually exclusive mode. These secondary structures are conserved across all *Drosophila Pips* investigated (Supplemental Fig. S9).


A
Drosophila PGRP-LC
Intron size (nt)
800
600
400
200
0
(A)n
B
Drosophila CG1637
Intron size (nt)
1600
1200
800
400
0
(A)n
C
Anopheles PGRP-LC
Intron size (nt)
2500
2000
1500
1000
500
0
(A)n
D
Drosophila CG30427
Intron size (nt)
2400
1800
1200
600
0
(A)n
E
Drosophila CG44325
Intron size (nt)
6600
4400
2200
0
(A)n
F
Drosophila CG10638
Intron size (nt)
180
160
140
120
100
80
60
40
20
0
(A)n
G
Drosophila PIP
Intron size (nt)
1200
900
600
300
0
(A)n
H
Drosophila CG42235
Intron size (nt)
900
600
300
0
(A)n

Supplemental Figure S11. Short introns within 3' tandem clusters are evolutionarily conserved. Genome-wide exon-intron architecture of pre-mRNAs was investigated in *Drosophila* and mosquito species. Statistics analysis of genes revealed that the short intron is a highly conserved feature in *Drosophila* and mosquito species, although the remaining intron sequences were highly divergent. The data shows the average size of the introns. Data were expressed as mean ± SD.



Supplemental Figure S12. The small size of introns within 3' tandem multi-exon clusters are required for proper alternative splicing. (*A*) Overview of constructs used to test the effect of the size of introns within 3' tandem multi-exon clusters on proper alternative splicing of *PGRP-LC* gene. **(*B*)** The intron between 3y and 4y was spliced earlier than constitutive introns in *PGRP-LC* gene. The arrows in the same color represent a pair of primers used to amplify PCR products after reverse transcription of RNA extracted from cytoplasm, nuclei and chromatin, respectively. **(*C*)** Increasing intron length within 3' tandem clusters affects normal splicing of *PGRP-LC* gene. The red line and the number above indicate that the intron length between 3x and 4x is extended by 317 bp, and primers (black arrows in Supplemental Fig. S11A) were used to amplify PCR products after transfection of WT and IE1 constructs. **(*D*)** Overview of constructs

used to test the effect of the size of introns within 3' tandem multi-exon clusters on proper alternative splicing of *CG30427* gene. The red line and the number above mean that the intron length between 4a and 4b is extended by 100 and 300 bp, respectively, and red triangle shows an artificial restriction site obtained from point mutation of intron between 4a and 4b. **(*E*)** The intron between 5a and 5b is spliced earlier than constitutive introns in *CG30427* gene. The arrows in the same color represent a pair of primers used to amplify PCR products after reverse transcription of RNA extracted from nuclei. **(*F*)** Increasing intron length within 3' tandem clusters affects normal splicing of *CG30427* gene. The asterisk indicates non-specific amplification.

Supplemental reference

Zuker M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* **31**: 3406-3415.