Supplementary Materials for

Hidden RNA pairings counteract the “first-come, first-served” splicing principle to drive stochastic choice in Dscam1 splice variants

Haiyang Dong, Bingbing Xu, Pengjuan Guo, Jian Zhang, Xi Yang, Lei Li, Ying Fu, Jilong Shi, Shixin Zhang, Yanda Zhu, Yang Shi, Fengyan Zhou, Lina Bian, Wendong You, Feng Shi, Xiaofeng Yang, Jianhua Huang, Haihuai He, Yongfeng Jin*

*Corresponding author. Email: jinyf@zju.edu.cn

Published 26 January 2022, Scie. Adv. 8, eabm1763 (2022)
DOI: 10.1126/sciadv.abm1763

The PDF file includes:

Supplementary Text
Tables S1 to S3
Legend for table S4
Figs. S1 to S15

Other Supplementary Material for this manuscript includes the following:

Table S4
Supplementary Text

Lineage-specific expansion of balancer RNA secondary structures

We next explored how balancer RNA secondary structures evolved during fly Dscam1 evolution. Detailed genomic structure and phylogenetic analyses revealed that the upstream balancer sequences 6.1 (Ubs6.1) and the downstream balancer sequences 6.40 (Dbs6.40) were conserved in all 20 Schizophora species, spanning ~100 million years (Fig. 3 and fig. S4A). However, phylogenetic analyses suggest that Dbs6.45 elements might have been lost in melanogaster groups following divergence of the melanogaster and willistoni groups approximately 30 million years ago. For example, Drosophila virilis contains two balancer sequences: one is orthologous to D. melanogaster Dbs6.40, and the other is located in the intron upstream of exon 6.45 (fig. S4A). Similarly, the Dbs6.45 sequence might have specifically degenerated in D. mojavensis, and Dbs6.48 sequences might have specifically emerged in the Dacinae subfamily. Thus, we revealed lineage-specific and conserved balancer RNA secondary structures during fly Dscam1 evolution.

We found that the number and distribution of the downstream balancer sequences were not random, but rather were selected under evolutionary pressure. First, the number of balancer sequences is largely correlated with the size of the exon 6 cluster. For example, we found at least four balancer sequences in the exon 6 cluster of the housefly Dscam1, which is consistent with the larger exon 6 cluster containing 65 exons. It is possible that Dbs6.40-homologue sequences were simultaneously duplicated with duplication-mediated exon expansion (27, 28). Moreover, the downstream balancer sequences are unevenly distributed, with the vast majority located in the back region of the exon 6 cluster (Fig. 3 and fig. S4B). The correlation of the balancer number with exon 6 cluster size and their unique distribution might be associated with their function.
**Exon 6.40 is repressed by balancer RNA secondary structures in WT flies**

The exon 6.40 balancer sequence Dbs6.40 is located in a region similar to that of the selector sequence upstream of exon 6.40. Dbs6.40 competitively pairs with the docking site and Ubs6.1 balancer sequence. Although exon 6.40 inclusion is dependent on balancer RNA secondary structures in mutant flies containing docking-site deletions (Fig. 2C), this does not necessarily mean that exon 6.40 is activated in WT flies. By contrast, balancer RNA secondary structures may cooperate with docking site–selector interactions to form strong, multiple-domain RNA secondary structures (fig. S7A, left panel). In this complex structure, exon 6.40 is located in the internal middle loop, and thus the inclusion of exon 6.40 would be inhibited by balancer RNA secondary structures employing a “looping-out” mechanism (16, 18). Indeed, we observed a significant increase in the inclusion frequency of exon 6.40 compared to the WT when we introduced disruptive mutations into Ubs6.1 (Dscam\textsuperscript{BM1–BM3}; fig. S7C, lane 2–4). By contrast, the inclusion frequency of exon 6.40 was significantly decreased in flies containing mutations in Ubs6.1 that strengthened the balancer RNA secondary structure (Dscam\textsuperscript{BM4}; fig. S7C, lane 5). These data indicate that balancer RNA secondary structures likely inhibit the inclusion of exon 6.40 in WT flies.

Furthermore, we found that deleting Dbs6.40 dramatically reduced exon 6.40 inclusion (Dscam\textsuperscript{BM5}; fig. S7C, lane 6); this is consistent with disruption of both the docking site and selector for exon 6.40 and Dbs6.40–Ubs6.1 pairing interactions. Similar results were obtained for Dscam\textsuperscript{BM6} flies. Conversely, Dscam\textsuperscript{BM7} flies exhibited a significant but modest reduction in exon 6.40 inclusion compared to WT flies (fig. S7C, lane 7). The point mutations in Dscam\textsuperscript{BM7} did not affect docking site–selector 6.40 pairing strength (fig. S7B). Therefore, this reduction might have been due to exon repression within the loop, caused by increased strength of the Dbs6.40–Ubs6.1 pair (fig. S7C, lane
8). Taken together, these data indicate that exon 6.40 inclusion is mediated by competition between docking site–selector 6.40 pairing and Dbs6.40–Ubs6.1 balancer pairing.
Table S1. All species used in this study were listed according to class, order, family, genus, and species. Three letter abbreviations used in sequence alignments were indicated in parentheses following species names for certain species. Dscam1 genomic DNAs obtained are indicated in the Genbank column. The number of Dscam1 exon 6 copies in each species is shown on the right.

| Class    | Order    | Family         | Genus     | Species (abbreviation) | GenBank            | Exon 6 |
|----------|----------|----------------|-----------|------------------------|--------------------|--------|
| Insecta  | Diptera  | Drosophilidae  | Drosophila| D. melanogaster (Dme)  | SIXD01000003.1     | 48     |
| Insecta  | Diptera  | Drosophilidae  | Drosophila| D. yakuba (Dya)        | AAEU02000101.1     | 41     |
| Insecta  | Diptera  | Drosophilidae  | Drosophila| D. erecta (Der)        | QMER02000007.1     | 45     |
| Insecta  | Diptera  | Drosophilidae  | Drosophila| D. ananassae (Dan)     | QMES02000004.1     | 46     |
| Insecta  | Diptera  | Drosophilidae  | Drosophila| D. pseudoobscura (Dps)| AAFS01000662.1     | 49     |
| Insecta  | Diptera  | Drosophilidae  | Drosophila| D. persimilis (Dpe)   | QMET02000002.1     | 49     |
| Insecta  | Diptera  | Drosophilidae  | Drosophila| D. willistoni (Dwi)   | AAQB01006282.1     | 49     |
| Insecta  | Diptera  | Drosophilidae  | Drosophila| D. albomicans (Dal)   | ACVY01105852.1     | 53     |
| Insecta  | Diptera  | Drosophilidae  | Drosophila| D. majavensis (Dmo)   | AAPU01010180.1     | 50     |
| Insecta  | Diptera  | Drosophilidae  | Drosophila| D. virilis (Dvi)      | QMEO02000215.1     | 52     |
| Insecta  | Diptera  | Drosophilidae  | Drosophila| D. grimshawi (Dgr)    | AAPT01021484.1     | 52     |
| Insecta  | Diptera  | Drosophilidae  | Scaptomyza| S. flava (SfI)        | RKRM01000564.1     | 50     |
| Insecta  | Diptera  | Tephritidae    | Bactrocera| B. tryoni (Btr)       | JHQJ01000220.1     | 56     |
| Insecta  | Diptera  | Tephritidae    | Bactrocera| B. dorsalis (Bdo)     | JFBF01000093.1     | 57     |
| Insecta  | Diptera  | Tephritidae    | Bactrocera| B. latifrons (BlA)    | MIMC01000186.1     | 57     |
| Insecta  | Diptera  | Tephritidae    | Bactrocera| B. oleae (Bol)        | JXPT01026309.1     | 57     |
| Insecta  | Diptera  | Tephritidae    | Zeugodacus| Z. cucurbitae (Zcu)   | JRNW01020448.1     | 57     |
| Insecta  | Diptera  | Tephritidae    | Ceratitis| C. capitata (Cca)     | AOHK02003278.1     | 57     |
| Insecta  | Diptera  | Diopsidae      | Teleopsis| T. dalmanni (Tda)     | NLCU01019975.1     | 55     |
| Insecta  | Diptera  | Muscidae       | Musca     | M. domestica (Mdo)    | AQPM01055346.1     | 64     |
### Table S2 Specific primers used for sgRNA and mutant screening

| Mutants | Screening primers | sgRNA Primers |
|---------|-------------------|---------------|
| **Dscam**<sup>M1</sup> | **Dme-Dock-M1-F1:** CCGTTACCCACATTTGCGG | **Ds-4.12-sg-F1:** TTCGAAAAATGATTACCCAGCCATG |
| | **Dme-Dock-R1:** ATGGGAAACCTTGCCGAAACTCTG | **Ds-4.12-sg-R1:** AAACCATGGCTGGTATACTATTTT |
| **Dscam**<sup>M2</sup> | **Dme-6.5SE-F1:** TGGACCCCGGTGGACTAGCG | **Ds-6.5-sg-F1:** TTCGGCATATCCCATGTAAGAAAGC |
| | **Dme-6.5SE-R1:** TCTAGACCACACTTATCCC | **Ds-6.5-sg-R1:** AAACGTTTTCTACATGGGATAGC |
| **Dscam**<sup>M3</sup> | **Dme-Dock-F1:** GCTACAGTGCCGAAACAAATC | **Ds-1Ea-sg-F1:** TTCGGACCTCCCATGAGTAG |
| **Dscam**<sup>M4</sup> | **Dme-Dock-R1:** ATGGGAAACCTTGCCGAAACTCTG | **Ds-1Ea-sg-R1:** AAACGTTTCTACATGGGATAGC |
| **Dscam**<sup>M5</sup> | **Dme-Dock-F1:** GCTACAGTGCCGAAACAAATC | **Ds-1Ea-sg-F1:** TTCGGACCTCCCATGAGTAG |
| **Dscam**<sup>M6</sup> | **Dme-Dock-R1:** ATGGGAAACCTTGCCGAAACTCTG | **Ds-1Ea-sg-R1:** AAACGTTTCTACATGGGATAGC |
| **Dscam**<sup>M7</sup> | **Dme-Dock-F1:** GCTACAGTGCCGAAACAAATC | **Ds-1Ea-sg-F1:** TTCGGACCTCCCATGAGTAG |
| **Dscam**<sup>M8</sup> | **Dme-Dock-R1:** ATGGGAAACCTTGCCGAAACTCTG | **Ds-1Ea-sg-R1:** AAACGTTTCTACATGGGATAGC |
| **Dscam**<sup>M9</sup> | **Dme-Dock-F1:** GCTACAGTGCCGAAACAAATC | **Ds-1Ea-sg-F1:** TTCGGACCTCCCATGAGTAG |
| **Dscam**<sup>M10</sup> | **Dme-Dock-R1:** ATGGGAAACCTTGCCGAAACTCTG | **Ds-1Ea-sg-R1:** AAACGTTTCTACATGGGATAGC |
| Fig. S1 | **Dscam**<sup>Mok.1</sup> | **Dme-Dock-F2:** CCGAAAAGAATTGTTGCTGTAAG | **Ds-1Ea-sg-F1:** TTCGGACCTCCCATGAGTAG |
| | | **Dme-Dock-R1:** ATGGGAAACCTTGCCGAAACTCTG | **Ds-1Ea-sg-R1:** AAACGTTTCTACATGGGATAGC |
| Fig. 2 | **Dscam**<sup>Adock1</sup> | **Dme-Dock-F2:** CCGAAAAGAATTGTTGCTGTAAG | **Ds-6.5-sg-F1:** TTCGGGATATCCCATGTAAGAAAGC |
| | | **Dme-Dock-R1:** ATGGGAAACCTTGCCGAAACTCTG | **Ds-6.5-sg-R1:** AAACGTTTCTACATGGGATAGC |
| | **Dscam**<sup>Adock2</sup> | **Dme-Dock-F2:** CCGAAAAGAATTGTTGCTGTAAG | **Ds-6.5-sg-F1:** TTCGGGATATCCCATGTAAGAAAGC |
| | | **Dme-Dock-R1:** ATGGGAAACCTTGCCGAAACTCTG | **Ds-6.5-sg-R1:** AAACGTTTCTACATGGGATAGC |
| | **Dscam**<sup>Adock3</sup> | **Dme-Dock-F2:** CCGAAAAGAATTGTTGCTGTAAG | **Ds-6.5-sg-F1:** TTCGGGATATCCCATGTAAGAAAGC |
| | | **Dme-Dock-R1:** ATGGGAAACCTTGCCGAAACTCTG | **Ds-6.5-sg-R1:** AAACGTTTCTACATGGGATAGC |
| | **Dscam**<sup>Adock4</sup> | **Dme-Dock-F2:** CCGAAAAGAATTGTTGCTGTAAG | **Ds-6.5-sg-F1:** TTCGGGATATCCCATGTAAGAAAGC |
| | | **Dme-6.1-R1:** TGGGATAGGCGATGGCGATTA | **Ds-6.5-sg-R1:** AAACGTTTCTACATGGGATAGC |
| Fig. 4 | **Dscam**<sup>Adock3M1</sup> | **Dme-Dock-F2:** CCGAAAAGAATTGTTGCTGTAAG | **Ds-4.12-sg-F1:** TTCGAAAAATGATTACCCAGCCATG |
| | | **Dme-6.1-R1:** TGGGATAGGCGATGGCGATTA | **Ds-4.12-sg-R1:** AAACCATGGCTGGTATACTATTTT |
| | **Dscam**<sup>Adock3M2</sup> | **Dme-Dock-F2:** CCGAAAAGAATTGTTGCTGTAAG | **Ds-6.5-sg-F1:** TTCGGGATATCCCATGTAAGAAAGC |
| | | **Dme-6.1-R1:** TGGGATAGGCGATGGCGATTA | **Ds-6.5-sg-R1:** AAACGTTTCTACATGGGATAGC |
| | **Dscam**<sup>Adock3M3</sup> | **Dme-Dock-F2:** CCGAAAAGAATTGTTGCTGTAAG | **Ds-4.12-sg-F1:** TTCGAAAAATGATTACCCAGCCATG |
| | | **Dme-6.1-R1:** TGGGATAGGCGATGGCGATTA | **Ds-4.12-sg-R1:** AAACCATGGCTGGTATACTATTTT |
| | | | **Ds-6.5-sg-R1:** AAACGTTTCTACATGGGATAGC | **Ds-6.5-sg-R1:** AAACGTTTCTACATGGGATAGC |
| Fig. 5 |   |   |   |
|--------|---|---|---|
| **ADbs6.40** | Dme-6.40-F1: CGAATTTGCCCCCTAACCTTGCG | Dme-6.40-R1: CTCGAAAAGGTCGGAGCTTCGTC | Ds-6.40-sg-F1: TCGCAAGACACTTGGAACAGCGT |
|        | Ds-6.40-sg-R1: AAAACGGCTTCCACTGGTTCTG | Ds-6.40-sg-F2: TCGGATATAATGTAATATTGTT | Ds-6.40-sg-R2: AAAACACCATATTACATTTATC |
| Dscam**E*40-48 | Dme-6.40-F1: CGAATTTGCCCCCTAACCTTGCG | Dme-6.40-R1: AGGATTTGCGATCGGGATCAAGT | Ds-6.48-sg-F1: TCGGCGCTCCTACATCTGCG |
|        | Ds-6.36-sg-F1: TCGGTTAGGTTCAGACCTCTG | Ds-6.36-sg-R1: AAAACCAGAGGTCTGGAACCTGAC | Ds-6.36-sg-R2: AAAACCAGAGGTCTGGAACCTGAC |
| Dscam**E*40-48 | Dme-6.40-F1: CGAATTTGCCCCCTAACCTTGCG | Dme-6.48-R1: AGGATTTGCGATCGGGATCAAGT | Ds-6.36-sg-F1: TCGGCGCTCCTACATCTGCG |
|        | Ds-6.36-sg-R1: AAAACCAGAGGTCTGGAACCTGAC | Ds-6.36-sg-R2: AAAACCAGAGGTCTGGAACCTGAC |
| **Dscam** | Dme-6.1-F1: CAAATCCAAACCTGCGCTATC | Dme-6.1-R1: TGGGATAGGCAGAATGGGT | Ds-6.1-sg-F1: TCGGCTTCTATTCACAGAAGCTAG |
| **RM1** | Ds-6.1-sg-R1: AAAACACTGGCTCTGGAAATGAGC | Ds-6.1-sg-R2: AAAACACTGGCTCTGGAAATGAGC |
| Dscam**RM2** | Dme-Dock-F2: CCGAAAAGAATTTGTGCTGCTAG | Dm-6.5-sg-F1: TCGGCTATCCACATGGAAGAAC |
| Dscam**RM3** | Dme-6.1-R1: TGGGATAGGCAGAATGGGT | Ds-6.5-sg-R1: AAAACGCTTCTACATGGGGGATAGC |
| Dscam**RM4** | Ds-6.40-F1: GAGTACTTGGCTCTCCACAGC | Ds-6.40-R1: TCGGCGCTCCTACATCTGCG |
| Dscam**RM6** | Dm-6.40-R1: TGGGATAGGCAGAATGGGT | Ds-6.40-R1: TCGGCGCTCCTACATCTGCG |
| Dscam**RM7** | Dm-In2-F1: CTTTCTAACTGCGCTGAAACTAG | Dm-6.1-R1: TGGGATAGGCAGAATGGGT |
| Dscam**In1** | Dm-In1-F1: CTTGCAACTTGGCTCTGATTAG | Dm-6.1-R1: TGGGATAGGCAGAATGGGT | Ds-6.1-sg-F1: TCGGCTTCTATTCACAGAAGCTAG |
| Dscam**In2** | Dm-6.1-R1: TGGGATAGGCAGAATGGGT | Ds-6.1-sg-R1: AAAACATGGCTTCTGGAAATGAGC |
| Dscam**In2** | Dm-Dock-F2: CCGAAAAGAATTTGTGCTGCTAG | Ds-6.1-sg-R1: AAAACATGGCTTCTGGAAATGAGC |
| **AEExon 6.1-6.6** | Dme-6.6-R1: GCAAGCTCAAGCTCAGCAATCT | Ds-6.1-sg-F1: TCGGCTTCTATTCACAGAAGCTAG |
|        | Ds-6.1-sg-R1: AAAACATGGCTTCTGGAAATGAGC | Ds-6.1-sg-R1: AAAACATGGCTTCTGGAAATGAGC |
|        | Ds-6.1-6.16-R1: GACATACGAGATTCAGGGAGCT | Ds-6.1-sg-F1: TCGGCTTCTATTCACAGAAGCTAG |
| **AEExon 6.1-6.16** | Dme-Dock-F2: CCGAAAAGAATTTGTGCTGCTAG | Ds-6.1-sg-R1: AAAACATGGCTTCTGGAAATGAGC |
|        | Dm-6.16-R1: GACATACGAGATTCAGGGAGCT | Ds-6.1-sg-R1: AAAACATGGCTTCTGGAAATGAGC |
### Specific primers used for RT-PCR

| Primers | 5'-3' sequence | Assay |
|---------|-----------------|-------|
| Ds-5-F  | GCT ACC AGT GCC GCA A CCA AA AC AT C | RT-PCR |
| Ds-7-R  | AG TCC CA C G TTT TCC GC TTC | RT-PCR |
| Ds-7-F  | AA CA TAT C TC G GT CAC GC | RT-PCR |
| Ds-8-R  | GT GC TGT GGT GGT | RT-PCR |
| Ds-5-F-seq | TC TGC GC CAG CAG TGC TTA TAA A GAG A | Multiplex high-throughput sequencing |
| Ds-7-R-seq | GT CG TGG CAG CAG A GT GAT G TTA AAG A G | Multiplex high-throughput sequencing |
| Ds-7-RT  | TACT GGA CAG TGA AT C CGG GGG GGG | Reverse transcription |
| Mutants | Mutation sequences* | Assay |
|---------|---------------------|-------|
| **Dscam**<sup>M1</sup> | gtctcattgtctgaccattccccatccccacactgagccagaacttcatttc | Docking site mutation |
| **Dscam**<sup>M2</sup> | atctcttcattttctctgctgatgtggtatgggtactgcgaatgc | 6.3 selector mutation |
| **Dscam**<sup>M3</sup> | gtctaaatctgtctgcactgatgtggtatgggtactgcgaatgc | Docking site mutation |
| **Dscam**<sup>M4</sup> | gtctaaatctgtctgcactgatgtggtatgggtactgcgaatgc | Docking site mutation |
| **Dscam**<sup>M5</sup> | gtctaaatctgtctgcactgatgtggtatgggtactgcgaatgc | Docking site mutation |
| **Dscam**<sup>M6</sup> | gtctaaatctgtctgcactgatgtggtatgggtactgcgaatgc | Docking site mutation |
| **Dscam**<sup>M7</sup> | gtctaaatctgtctgcactgatgtggtatgggtactgcgaatgc | Docking site mutation |
| **Dscam**<sup>M8</sup> | gtctaaatctgtctgcactgatgtggtatgggtactgcgaatgc | Docking site mutation |
| **Dscam**<sup>M9</sup> | gtctaaatctgtctgcactgatgtggtatgggtactgcgaatgc | Docking site mutation |
| **Dscam**<sup>M10</sup> | gtctaaatctgtctgcactgatgtggtatgggtactgcgaatgc | Docking site mutation |
| **Fig. S1** | Dscam<sup>Ms6.1</sup> | ccaagtctttgatagaattcattcaattcgaattcttctttccttaagcttggagctattactttgaatccagaggtcactgtctctttgcaacttggtgttggctatcttcacttcttattctcctttgaactgcgctcaagtgactcacttcagatccacttccatatcttcaccaattggcattttactatgtaaggctattaacagttttgaactatatttcgcttcattcagcgcatctggccaggaagaagctcttagttcgaagtcggcccaagttcccaataactctgaccagcagcagtttcccacgagtcagccccctcttatgccctgcccaggtttatccagctccccttttttag... | 6.1 selector mutation |
| **Fig. 2** | Dscam<sup>ΔDock1</sup> | Fragment from the docking site to the entire exon 6.1 deletion |
| Dscam<sup>ΔDock2</sup> | gtctaaatctgtctgcactgatgtggtatgggtactgcgaatgc | Docking site and its downstream sequence deletion |
| Dscam<sup>ΔDock3</sup> | gtctaaatctgtctgcactgatgtggtatgggtactgcgaatgc | Docking site and its downstream sequence deletion |
| Dscam<sup>ΔDock4</sup> | gtctaaatctgtctgcactgatgtggtatgggtactgcgaatgc | Docking site and its downstream sequence deletion |
| **Dscam**<sup>ΔDock3M1</sup> | Fragment from the docking site to the entire exon 6.1 deletion |
| **Dscam**<sup>ΔDock3M2</sup> | Fragment from the docking site to the entire exon 6.1 deletion |

*Note: The mutation sequences are presented in DNA format.*
**Dcam**

- **Fig. S9**

  - Insertion mutation upstream of exon 6.1

  - Insertion mutation upstream of exon 6.1

- **Dbs6.40**

  - Fragment from exon 6.40-6.48 deletion but not include Dbs6.40

- **Dbs6.40-Exon4**

  - Fragment from exon 6.40-6.48 deletion including Dbs6.40

- **Dbs6.40**

  - Dbs6.40 deletion

- **Dcam exon 6.4**

  - Dcam exon 6.4 deletion (Fig. 2) and Ubs6.1 mutation

- **Exon 6.1**

  - Exon 6.1 deletion

  - Ubs6.1 deletion and 5’ splice site mutation

  - Ubs6.1 and 5’ splice site mutation

- **Exon 6.48**

  - Dbs6.40 deletion

- **Exon 6.6**

  - Dbs6.40 deletion

- **Exon 6.6**

  - Dbs6.40 deletion

- **Exon 6.7**

  - Dbs6.40 deletion

- **Exon 6.8**

  - Dbs6.40 deletion

- **Exon 6.9**

  - Fragment from exon 6.1-6.6 deletion
*Lowercase letters denote intron sequences and uppercase letters denote exon sequences. Sequences in red font represent mutation sequences. These sequences in red font and strikethrough represent the deleted sequences.

**Table S4 Data of the inclusion frequency of variable exon 6s. Please see separate file attached.**
Fig. S1. *In vivo* mutagenesis verifies cluster-wide docking site-selector base pairing in exon 6 cluster, Related to Fig. 1. (A) Quantitation of the RT-PCR diagram from the head of wild-type and mutant flies (*Dscam*<sup>M1</sup>, *Dscam*<sup>M2</sup>, and *Dscam*<sup>M12</sup>). (B) Inclusion frequency for exon 6 variants of disruptive mutations (*Dscam*<sup>M1</sup>, *Dscam*<sup>M2</sup>) and compensatory double mutation (*Dscam*<sup>M12</sup>). (C) RT-
PCR analysis shows no obvious exon 6 skipping in point mutations of the docking site. (D) Effects of point mutations of the docking site on exon 6 inclusion. Percent inclusion is shown for each exon 6 variant (from exons 6.1 to 6.48) in Dscam<sup>M3-M10</sup> mutant flies. All data on high-throughput sequencing of RT-PCR products are expressed as mean ± s.d. from three independent experiments. (E) The schematic of mutation of the predicted selector sequence 6.1. The RNA secondary structure between the docking site and selector sequence 6.1 was predicted in the previous study (14). Mutations introduced into dsRNA were indicated in Dscam<sup>M6.1</sup> flies. (F) RT-PCR analysis shows no obvious exon 6 skipping in Dscam<sup>M6.1</sup> flies. (G) Mutation of 6.1 exon selector sequence resulted in increased exon 6.1 inclusion compared to wild-type. The data suggested that the predicted selector sequence upstream of exon 6.1 was incorrect. (H, I) The revised docking site-selector base pairings are shown. The revised selector sequences of exon 6.22 and 6.38 are evolutionarily conserved in Drosophila species.
### Correlation between inclusion frequency and secondary structure strength

| Exon 6.x | r     | p value | Exon 6.x | r     | p value |
|----------|-------|---------|----------|-------|---------|
| 6.1 *    | 0.09  | 0.82388 | 6.25     | 0.79  | 0.01190 |
| 6.2      | 0.61  | 0.07846 | 6.26     | 0.96  | 0.00004 |
| 6.3      | 0.73  | 0.02603 | 6.27     | 0.73  | 0.02684 |
| 6.4      | 0.71  | 0.03086 | 6.28     | 0.71  | 0.03242 |
| 6.5      | 0.83  | 0.00514 | 6.29     | 0.73  | 0.02582 |
| 6.6      | 0.80  | 0.00900 | 6.30     | 0.90  | 0.00086 |
| 6.7      | 0.88  | 0.00179 | 6.31     | 0.73  | 0.02486 |
| 6.8      | 0.63  | 0.07078 | 6.32     | 0.91  | 0.00076 |
| 6.9      | 0.56  | 0.11762 | 6.33     | 0.87  | 0.00217 |
| 6.10     | 0.90  | 0.00105 | 6.34     | 0.91  | 0.00069 |
| 6.11     | ---   | ---     | 6.35     | 0.44  | 0.23278 |
| 6.12     | 0.90  | 0.00095 | 6.36     | 0.30  | 0.42880 |
| 6.13     | 0.89  | 0.00112 | 6.37     | 0.90  | 0.00090 |
| 6.14     | 0.55  | 0.12419 | 6.38     | Previous 0.64 0.06264 |
| 6.15     | 0.89  | 0.00140 | Revised 0.97 0.00002 |
| 6.16     | 0.91  | 0.00074 | 6.39     | 0.38  | 0.30712 |
| 6.17     | 0.70  | 0.03569 | 6.40     | 0.74  | 0.02197 |
| 6.18     | 0.66  | 0.05232 | 6.41 *   | 0.06  | 0.87310 |
| 6.19     | 0.89  | 0.00148 | 6.42     | 0.44  | 0.23982 |
| 6.20     | 0.20  | 0.60112 | 6.43 *   | -0.09 | 0.81823 |
| 6.21 *   | -0.31 | 0.41305 | 6.44     | 0.59  | 0.09678 |
| 6.22     | Previous -0.51 0.15647 | 6.45 *   | -0.11  | 0.77904 |
| Revised  | 0.87  | 0.00217 | 6.46 *   | -0.89  | 0.00123 |
| 6.23     | 0.83  | 0.00513 | 6.47     | 0.38  | 0.31087 |
| 6.24     | 0.81  | 0.00809 | 6.48     | 0.67  | 0.04774 |

**Fig. S2.** The correlation coefficients of exon 6 selection with the strength of predicted RNA pairing for each exon 6 in the docking site mutation flies, Related to Fig. 1.

Exon 6s with no or poor positive correlation between its frequency and base-pairing strength are marked by a red asterisk. Based on these data, we revised the selector sequence for exon 6.22 and exon 6.38 (in red).
Fig. S3. The exon 6.1 sequences are required for exon 6.40 inclusion in the absence of the docking site, Related to Fig. 2. (A) Effect of deletions on exon 6 inclusion in various tissues. The inclusion frequency of exon 6 increased as the deleted docking site and its downstream sequence increased. (B) The log₂ fold change of exon 6 inclusion frequency of DscamΔDock1–ΔDock4 flies compared to wild type. The boxed area is magnified in the inset showing the log₂ fold change on the
first exon 6 and exon 6.40. In the absence of the docking site, the first exon 6 remarkably increased (panel i). Deleting the docking site but not exon 6.1 resulted in significant increase of exon 6.40 inclusion frequency (\textit{Dscam}^{\Delta Dock1-\Delta Dock3}, panel ii), while knocking out the fragment spanning the docking site and exon 6.1 resulted in a decrease in exon 6.40 inclusion (\textit{Dscam}^{\Delta Dock4}, panel ii). This result suggests that the exon 6.1 sequence is required for high exon 6.40 inclusion frequencies in the absence of the docking site.
Fig. S4. Evolutionarily conserved signatures of the balancer RNA base pairings in exon 6 cluster of fly *Dscam1*, Related to Fig. 3. (A) The arrangement of conserved and lineage-specific balancer sequences across *Drosophila* species. Symbols used are the same as in Fig. 3. A phylogenetic tree depicting the relationships is shown on the left. Above are conserved sequences for *Drosophila*
species. Individual upstream balancer sequence (marked by red triangle) was predicted to pair with downstream balancer sequences (marked by red sector). The red dashed arrow represents the balancer base pairing interactions. Phylogenetic analysis suggests that Dbs6.45 elements might have been lost in melanogaster and obscura groups after the divergence of melanogaster and willistoni group near 30 million years ago. Similarly, Dbs6.45 elements might have specifically degenerated in D. mojavensis. (B) The balancer RNA secondary structures are conserved across fly species. The Ubs6.1-Dbs6.40 base pairings are shown. The sequences that make up the core of the stem are highlighted in blue. Nucleotide structural covariation in Ubs6.1-Dbs6.40 is highlighted in red.
Fig. S5. Evolution of balancer RNA secondary structures in exon 6 cluster of **Dscam1**, Related to Fig. 3. (A–C) Evolutionarily conserved balancer RNA secondary structures are shown in exon 6
cluster of *Dscam1*. The core regions of the balancer RNA base pairing are highlighted in blue (See Table S1 for abbreviations). Nucleotide structural covariations that maintain the structural integrity of the Ubs6.1-Dbs6.48 base pairing are shaded in red. (D) The alignment of four downstream balancer sequences in *M. domestica*. The most frequent nucleotides are highlighted in magenta. The core regions of the downstream balancer sequences are complementary to the upstream balancer sequence.
Fig. S6. Balancer base pairing enhances distal exon 6 inclusion, Related to Fig. 5. (A) The percent inclusion of exon 6 variants in Ubs6.1 or Dbs6.40 mutant flies (Dscam\textsuperscript{BM1–BM7}). The schematic diagram of the mutations is shown in the inset. (B) The schematic that the coordination of balancer base pairing (Ubs6.1-Dbs6.40) and docking site-selector base pairing promotes the inclusion of exon 6.41 and its downstream exon 6 variants is shown.
Fig. S7. Exon 6.40 inclusion mediated by competition between docking site-selector 6.40 pairing and Dbs6.40-Ubs6.1 balancer pairing, Related to Fig. 5. (A) The schematic of exon 6.40 inclusion mediated by competition between docking site-selector 6.40 pairing and Dbs6.40-Ubs6.1 balancer pairing. Dbs6.40 was located at the region nearly to the selector sequence of exon 6.40. (B) Predicted balancer (upper) and docking site-selector (lower) base pairings for the wild type and Dbs6.40 mutant flies (Dscam<sup>Bm5–Bm7</sup>; mutated nucleotides are shown in red). (C) The exon 6.40 inclusion frequency in Dscam<sup>Bm1–Bm7</sup> flies. Disrupting or weakening Ubs6.1-Dbs6.40 base pairing increased exon 6.40 inclusion frequency (Dscam<sup>Bm1–Bm3</sup>), while enhancing the balancer RNA secondary structure decreased exon 6.40 inclusion (Dscam<sup>Bm4</sup>). Dbs6.40 deletion or partial disruption almost destroyed the exon 6.40 inclusion (Dscam<sup>Bm5</sup>, Dscam<sup>Bm6</sup>). Moreover, we observed a significant but modest reduction in exon 6.40 inclusion in Dscam<sup>Bm7</sup>. All data are expressed as means ± s.d. of three independent experiments. *P < 0.05; **P < 0.01; ***P < 0.001; NS, not significant (Student’s t-test, two-tailed).
Fig. S8. Heat map of changes in the frequency of the exon 6 inclusion in various developmental stages and tissues of balancer mutants, Related to Fig. 5.

The log2 fold change of the inclusion frequency of variable exons in various developmental stages and tissues of mutant flies (Dscam\textsuperscript{BM2}, Dscam\textsuperscript{BM4}, Dscam\textsuperscript{BM5}, and Dscam\textsuperscript{BM7}) compared to wild-type. The red triangles indicate the increased inclusion frequency in a proximity-correlated manner, while the blue triangles indicate the decreased inclusion frequency.
Fig. S9. Effect of mimicking balancer base pairing and deleting variable exon 6s on exon 6 inclusion, Related to Fig. 5. (A) Schematic of knock-in mutant flies. we mimicked balancer base pairing through artificially inserting a Ubs-like element into the intron upstream of exon 6.1 of D. melanogaster Dscam1. The dashed arrow marks RNA pairing interaction. (B) The predicted balancer RNA secondary structures. Mutations are introduced into dsRNA to potentially form a stronger two-domain RNA secondary structure to enhance the distal exon 6s inclusion (Dscam1<sup>in1</sup>, Dscam1<sup>in2</sup>). (C) Effect of mimicking balancer RNA secondary structure on the inclusion of exon 6.46 and downstream exon 6s. Data are expressed as mean ± s.d. from three independent experiments.
Deletion of exon 6.1-containing fragment led to a remarkable preference for choice of the 5’ proximal exon 6s in ΔExon6.1, ΔExon6.1-6.6 and ΔExon6.1-6.16 mutant flies (A), while we have not observed a preference for the choice of the 5’ proximal exon 6s in the non-exon 6.1 fragment knockout mutant flies (ΔExon6.2-6.10, ΔExon6.10-6.20, and ΔExon6.21-6.30) (B). “**” in ΔExon6.10-6.20 represents abnormal exon 6.21 reduction due to the deletion of partial 6.21 selector sequence. The red striped frame represents the deleted variable exon. All data are expressed as means ± s.d. of three independent experiments.

Fig. S10. Effect of deleting variable exon 6 region on exon 6 inclusion, Related to Fig. 5. (A, B)
Fig. S11. Multiple balancer base pairings confer long-range enhancement of back exon 6
inclusion in *D. virilis*, Related to Fig. 6. (A) Schematic of multiple balancer base pairings in *D. virilis*. The balancer base pairing may cooperate with docking site-selector base pairing to potentially form multiple-domain RNA secondary structures, thereby enhancing the inclusion of back exon 6s. (B) Comparison of RNA secondary structures among *D. melanogaster* (Dme), *D. virilis* (Dvi), and CK(Dme/Dvi). (C) RT-PCR analysis shows no obvious exon 6 skipping in CK(Dme/Dvi) and mutant flies (*Dscam* CM1–CM4, *Dscam* CM24). (D) The percent inclusion of exon 6 variants in CK(Dme/Dvi) and mutant flies. The inset shows a schematic diagram of the balancer RNA secondary structure mutations. The green column indicates the inclusion frequency of *D. melanogaster* variable exon 6 in the chimera mutants, and the purple column indicates the inclusion frequency of *D. virilis* variable exon 6 in the chimera mutants. All data are expressed as means ± s.d. of three independent experiments.
Fig. S12. RNA secondary structures are required for *Drosophila* nervous system development, Related to Figs. 7 and 8. (A) Schematic diagram of class I neurons (vpda). Representative images of dendrites self-repulsion of class I (vpda) neurons of different genotypes. Yellow arrowheads indicate self-crossing of class I dendrites. Scale bars, 50 µm. (B) Representative images of class I-III neurons in *Dscam*ΔDock1 and *Dscam*ΔUbs1 and their corresponding dendrite tracings were shown. The splicing pattern of *Dscam1* variable exon 6 of wild-type and mutants were shown at the bottom, and the inclusion of exon 6.1 was marked in red. (C-E) Correlation analysis between splicing preference of exon 6.1 and three neuronal defect phenotypes.
Fig. S13. Compensation effect of docking site-selector and balancer base pairing on the preference for exon 6 choice, Related to Fig. 9. (A) Statistical analysis revealed striking positive linear correlations between the log₂ fold change in the inclusion frequency of exon 6 and the distance to the targeted exon 6 variant in docking site deletion mutants ($Dscam^{\Delta Dock1}$–$Dscam^{\Delta Dock3}$). (B) Striking negative linear correlations between the log₂ fold change and the distance to the targeted exon 6 variant in balancer deletion mutants ($Dscam^{BM1}$–$Dscam^{BM3}$). (C) Schematic diagram of the compensation effect of docking site-selector and balancer base pairing on the preference for exon 6 choice.
Fig. S14. Similar balancer RNA secondary structures and architecture were found in Dscam1 exon 9 clusters. (A) Sequence alignment of Dscam1 exon 9 from Drosophila species. Dme: D. melanogaster (SIXD0100003.1); Dsi: D. simulans (NIFY0100002.1); Dse: D. sechellia (AAKO01000452.1); Dya: D. yakuba (AAEU02000101.1, AAEU02000102.1); Der: D. erecta (QMER0200007.1); Dwi: D. willistoni (AAQB01006282.1); Dvi: D. virilis (QMEO02000215.1); Dgr: D. grimshawi (AAPT01021484.1). Above are conserved sequences in Drosophila species. Upstream balancer sequence (marked by red saddle shapes) was supposed to pair with downstream balancer sequence (marked by red semicircles). The introns are not drawn to scale. The red dashed arrow represents the balancer base pairing. (B) The balancer RNA secondary structures are conserved across Drosophila species.
Fig. S15. Similar balancer RNA secondary structures and architecture were found in *Drosophila MRP1* gene. (A) Sequence alignment of *MRP1* exon 8 from *Drosophila* species. Dme: *D. melanogaster* (AE014134.5); Dsi: *D. simulans* (CM000361.1); Dse: *D. sechellia* (CH480831.1); Dya: *D. yakuba* (CM000157.2); Der: *D. erecta* (CH954177.1); Dan: *D. ananassae* (CH902620.1); Dps: *D. pseudoobscura* (CH379061); Dpe: *D. persimilis* (CH479247.1). Above are conserved sequences in *Drosophila* species. Upstream balancer sequence (marked by red saddle shapes) was supposed to pair with downstream balancer sequence (marked by red semicircles). The introns are not drawn to scale. The red dashed arrow represents the balancer base pairing. Previous study had verified the docking site-selector base pairing of *Drosophila MRP1* gene (40). (B) The balancer RNA secondary structures of *MRP1* gene are conserved across *Drosophila* species. The sequences that make up the core of the stem are highlighted in blue.