MOLECULAR BIOLOGY

Trans-splicing facilitated by RNA pairing greatly expands sDscam isoform diversity but not homophilic binding specificity

Shouqing Hou[1]†, Guo Li[1]†, Bingbing Xu[1], Haiyang Dong[1], Shixin Zhang[1], Ying Fu[1], Jilong Shi[1], Lei Li[1], Jiayan Fu[1], Feng Shi[1], Yijun Meng[2], Yongfeng Jin[1]*

The *Down syndrome cell adhesion molecule 1* (*Dscam1*) gene can generate tens of thousands of isoforms via alternative splicing, which is essential for nervous and immune functions. Chelicerates generate approximately 50 to 100 shortened Dscam (sDscam) isoforms by alternative promoters, similar to mammalian protocadherins. Here, we reveal that trans-splicing markedly increases the repository of sDscamβ isoforms in *Tetranychus urticae*. Unexpectedly, every variable exon cassette engages in trans-splicing with constant exons from another cluster. Moreover, we provide evidence that competing RNA pairing not only governs alternative cis-splicing but also facilitates trans-splicing. Trans-spliced sDscam isoforms mediate cell adhesion ability but exhibit the same homophilic binding specificity as their cis-spliced counterparts. Thus, we reveal a single *sDscam* locus that generates diverse adhesion molecules through cis- and trans-splicing coupled with alternative promoters. These findings expand understanding of the mechanism underlying molecular diversity and have implications for the molecular control of neuronal and/or immune specificity.

INTRODUCTION

Neuronal self-avoidance refers to the tendency of neurites from the same neuron to avoid one another. This mechanism is conserved in both vertebrates and invertebrates and plays a vital role in the assembly of neural circuits (*1*–*3*). In *Drosophila*, neuronal self-avoidance is mediated by the *Drosophila Down syndrome cell adhesion molecule 1* (*Dscam1*) locus, which potentially encodes 38,016 distinct isoforms via stochastic mutually exclusive alternative splicing (*1*, *4*, *5*). Individual neurons express a unique set of distinct Dscam1 isoforms in a stochastic but biased manner (*4*, *6*–*8*). These Dscam1 isoforms engage in highly isoform-specific homophilic interactions (*9*, *10*). The vast diversity of Dscam1 isoforms is sufficient to confer a unique molecular identity on each neuron, thereby allowing neurons to discriminate between self and nonself neurons (*1*, *11*). Genetic analyses have shown that Dscam1 isoform diversity is required for self-avoidance and self-/non–self-discrimination (*12*–*16*).

However, mammalian *Dscam* genes do not generate the extensive diversity of their insect *Dscam1* counterparts (*17*, *18*). Instead, another single genomic locus, the clustered *Pcdh*s of the cadherin superfamily, performs an analogous function in mammals (*1*, *2*, *5*, *19*). This genomic locus contains three tandemly arranged gene clusters—*Pcdh*α, *Pcdh*β, and *Pcdh*γ—encoding a total of 50 to 60 protocadherin (Pcdh) proteins (*20*, *21*). In contrast to *Dscam1*, differential expression of the Pcdh isoforms is achieved through a combination of stochastic promoter choice and alternative cis-splicing (*22*–*24*). However, clustered Pcdh isoform proteins, like fly Dscam1 proteins (*25*, *26*), undergo highly specific homophilic binding, which is mediated by a mechanism coupling nonspecific cis and specific trans interactions (*25*–*29*). Knockout deletion analyses indicate that Pcdh isoform diversity is essential for normal self-avoidance of dendrites and axons (*30*, *31*), which suggests that mammalian clustered Pcdhs and fly Dscam1 evolved analogous processes for neuronal self-avoidance.

We recently identified a family of shortened *Dscam* (*sDscam*) genes with tandemly arrayed 5′ cassettes in the subphylum Chelicerata (*32*, *33*). Chelicerata genomes encode approximately 100 *sDscam* isoforms via alternative promoters, except in *Tetranychus urticae* (*32*, *33*). These *sDscam*s contain tandemly arrayed cassettes encoding one or two immunoglobulin (Ig) domains in the variable 5′ region and thus can be subdivided into the *sDscam*α and *sDscam*β subfamilies. It is interesting that Chelicerata *sDscam*s show remarkable organizational resemblance to vertebrate clustered Pcdhs, occupying the 5′ variable region and 3′ constant region (*17*, *32*, *33*). Recent studies have demonstrated that sDscamαs and sDscamβs of *Mesobuthus martensii* engage in isoform-specific homophilic binding and interact in trans conformation via antiparallel Ig1 self-binding (*34*). Different sDscam isoforms exhibit promiscuous cis interactions via membrane-proximal fibronectin type III domains, which are independent of the trans interactions. Thus, Chelicerata sDscams appear to behave similarly to vertebrate clustered Pcdhs in many respects.

Mites belong to subphylum Chelicerata, which is the second largest group of terrestrial animals. This clade includes members with a wide range of lifestyles, from parasitic to predatory to herbivorous, and includes scabies mites and allergy-causing dust mites, which pose major risks to human health (*35*, *36*). The spider mite *T. urticae*, a cosmopolitan agricultural pest with a wide host plant range and strong resistance to pesticide, causes substantial damage and losses to yields (*37*–*40*). We previously identified a single genomic locus containing three variable *sDscam* clusters in *T. urticae*, with 15, 14, and 4 copies of *sDscam*β*1*, *sDscam*β*2*, and *sDscam*β*3*, respectively (*33*). Thus, the total number of sDscam isoforms is exceptionally low in *T. urticae* compared to the approximately 100 isoforms in other Chelicerata species investigated, as estimated from the number of Ig7s or orthologs (*32*, *33*). Moreover, the *sDscam*α subfamily

[1]MOE Laboratory of Biosystems Homeostasis and Protection and Innovation Center for Cell Signaling Network, College of Life Sciences, Zhejiang University, Hangzhou, Zhejiang ZJ310058, P. R. China. [2]College of Life and Environmental Sciences, Hangzhou Normal University, Hangzhou, Zhejiang ZJ310018, P. R. China.
*Corresponding author. Email: jinyf@zju.edu.cn
†These authors contributed equally to this work.

is not present in *T. urticae*. Whether these mites have evolved other mechanisms to compensate for their low diversity of sDscam isoforms remains unknown.

In this study, we reveal that trans-splicing markedly expands the sDscamβ isoform repertoire in *T. urticae*. We are surprised to find that every variable exon cassette engages in trans-splicing with constant exons from another cluster. Moreover, we provide evidence that competing RNA pairs govern alternative cis- and trans-splicing. Cell aggregation assays indicate that trans-spliced sDscam isoforms mediate cell adhesion activity but share the same homophilic binding specificity as their cis-spliced counterparts. Thus, we reveal a single extreme *sDscam* locus that generates marked diversity in molecular adhesion through alternative cis- and trans-splicing coupled with alternative promoters and combinatorial homophilic recognition. These findings help to elucidate the cell identities and molecular control mechanisms underlying neuronal and/or immune specificity.

RESULTS

Trans-splicing markedly increases the sDscam isoform repertoire

To characterize the genomic structure of *sDscam* genes in *T. urticae*, we conducted RNA sequencing (RNA-seq) analyses using a publicly available *T. urticae* genome. Genome-wide analyses confirmed that all 5′ clustered cassettes of sDscams were located in this single chromosomal locus. It is interesting that an additional 3′ exon homolog was located downstream of the large common exon of *sDscamβ1* (Fig. 1, A and B). However, we found no 5′ variable region. Phylogenetic analyses revealed that *T. urticae sDscam*s arose from two duplications of one ancestral gene (fig. S1, A and B). In addition, a comparison of 5′ cassette- and gene-based clustering showed that *sDscam* and cassette duplications occurred alternately during mite evolution (fig. S1, B and C). RNA-seq analyses revealed that this 3′ exon could be spliced to the exon in the 5′ variable region of *sDscamβ1* (Fig. 1C). To confirm these results, we used reverse transcription polymerase chain reaction (RT-PCR) with exon-specific primers (table S2) to systematically validate the possible combinations of the 5′ variable exon cassette and alternative 3′ exon (Fig. 1, B and D). Together, these data indicate that the alternative 3′ exon may have evolved to generate a greater number of sDscam isoforms.

We were surprised to identify chimeric transcripts containing *sDscamβ1* variable cassettes and the *sDscamβ2* constant exon through RNA-seq analyses (Fig. 1C). The *sDscamβ1* and *sDscamβ2* gene clusters were located on the same chromosome but were transcribed in different directions (Fig. 1B). These transcripts can be explained by intermolecular trans-splicing (*41*). The trans-splicing juncture of chimeric mRNA uses canonical cis splice sites. To confirm these exon junctions, we examined the chimeric transcripts using RT-PCR. The results demonstrated the existence of trans-splicing products between *sDscamβ1* variable cassettes and *sDscamβ2* constant exons for all members tested (Fig. 1H and fig. S2). Furthermore, transcripts composed of *sDscamβ2* variable cassettes and *sDscamβ1* constant exons were identified (Fig. 1E and fig. S2). All four constant exons could be spliced to both the intra- and intergenic variable cassettes (Fig. 1, C to J, and fig. S2). Thus, this *sDscam* locus may produce 132 isoforms in *T. urticae* through a combination of alternative promoters with alternative cis- and trans-splicing. On the basis of the exon junctions inferred from the RNA-seq data, we estimated that approximately 60% of *sDscam* mRNA isoforms arose from trans-splicing. Further analyses demonstrated that these constant and variable regions exhibited different specific expression patterns in various development stages and different stresses (fig. S3, A and B). Collectively, these results indicate that trans-splicing can markedly expand the diversity of *sDscam* transcript isoforms in *T. urticae*.

To examine whether these trans-splicing patterns are conserved in other *Dscam* genes, we investigated trans-spliced isoforms in other Chelicerata species. We detected trans-splicing isoforms between *sDscamβ2* variable exons and *sDscam*β5 constant exons in *Ixodes scapularis*, albeit with very low frequency, which suggests that alternative trans-splicing functions in a species-specific manner. Given the exceptionally low number of cassette repeats in the 5′ variable region in *T. urticae* compared to other species, we speculate that alternative trans-splicing evolved to compensate for the low number of *sDscam* isoforms.

Intronic competing RNA pairing mediates alternative cis-splicing

To identify the cis-elements involved in regulating the selection of the 5′ alternative cassette, we used comparative sequence analysis to search for sequences that are conserved among *sDscamβ1*–*β4*s. It is interesting that sequence alignment revealed one conserved element (Ds, docking site) within an intron upstream of the individual constant region (Fig. 2A and fig. S4A). By probing with the docking site sequence, we identified reverse complementary sequences in the intron immediately downstream of the individual variable cluster (Fig. 2A and fig. S4B). Coincidently, a statistical survey of *sDscamβ1*–*β4* genes revealed that the intronic distance between the 5′ splice site and the selector sequence was small [43 ± 10 nucleotides (nt)] and relatively conserved (fig. S4, C and D). Thus, although the distance between the alternative 5′ splice site of the variable region and constitutive 3′ splice site of the constant exon is very large and highly variable, the base-pairing interaction between the docking site and selector sequence shortens the effective distance to approximately 120 nt. The predicted architecture of base pairing between the docking site and selector sequence in *sDscamβ1*–*β4* is analogous to the model of competing RNA structures that governs the internal mutually exclusive splicing of *Drosophila Dscam1* and *14-3-3ξ* pre-mRNA (*42*–*44*). We also found that there were moderate correlations between the biased expression and the predicted thermodynamic stabilities of docking site-selector pairings (fig. S6). Therefore, we propose that the selection of the *sDscam* 5′ splice isoform is regulated through intronic competing base pairing. To explore how these cis-elements and their base pairing mediate the selection of the 5′ alternative cassette of *sDscamβ*, we first generated a minigene construct containing the 5′ alternative cassette β1V13 to the constant region under the inducible metallothionein promoter (Fig. 2B). Splice isoforms containing the 5′ alternative cassette β1V13 or β1V14 were detected through transfection experiments using heterologous *Drosophila* S2 cells (Fig. 2, C and D). This system is well suited for analyzing cis-elements involved in the selection of 5′ alternative cassettes.

Next, we tested the effects on 5′ alternative cassette selection of disruptive and compensatory mutations within the predicted stem structure. We found that mutation of the selector sequence downstream of the alternative cassette β1V13 (Ss13 mutation, M1) destroyed β1V13 cassette inclusion (Fig. 2D and fig. S7, A and B), which indicates that β1V13 cassette inclusion is dependent on the Ss13 element. Conversely, Ss14 mutation (M2) eliminated β1V14

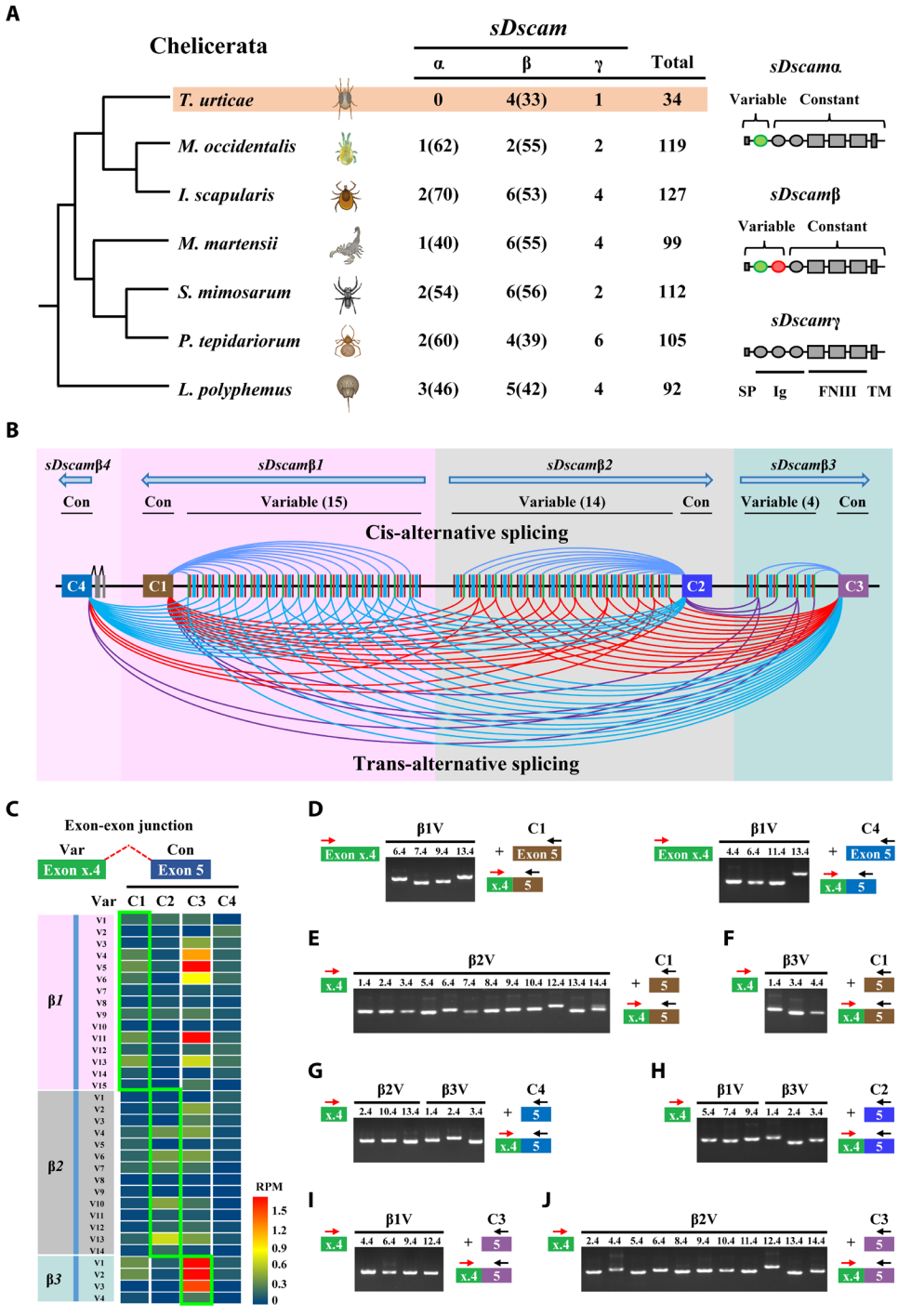

Fig. 1. A genomic locus generating extensive sDscam isoforms via multiple promoters and cis- and trans-alternative splicing in *T. urticae*. See also figs. S1 and S2. (**A**) Phylogenetic distribution of sDscam and isoform members in chelicerates. The variables Ig1s and Ig2s are indicated by green and red circles, respectively. Data from other species are referenced from our previous study (*32*). (**B**) Schematic of an *sDscam* locus. The 5′ untranslated region of *sDscamβ4* is represented by a gray rectangle. The arrow indicates transcriptional direction. Cis- and trans-spliced isoforms are represented by blue lines (above) and other colored lines (below), respectively. The color connections are supported by RNA-seq and RT-PCR data. Var, variable; Con (C), constant. (**C**) Quantification of the cis- and trans-spliced isoforms. RPM, reads per million. (**D**) Validation of alternative combinations of 5′ and 3′ alternative exons. Because of the low expression of variable exons, nested PCR was required to amplify the products; only the primers used in the second PCR are depicted (table S2). (**E** to **J**) Evidence of trans-splicing between different genes. These combinations included *sDscamβ2* and *sDscamβ1* (E), *sDscamβ3* and *sDscamβ1* (F), *sDscamβ2*–β3 and *sDscamβ4* (G), *sDscamβ1*/*sDscamβ3* and *sDscamβ2* (H), *sDscamβ1* and *sDscamβ3* (I), and *sDscamβ2* and *sDscamβ3* (J).

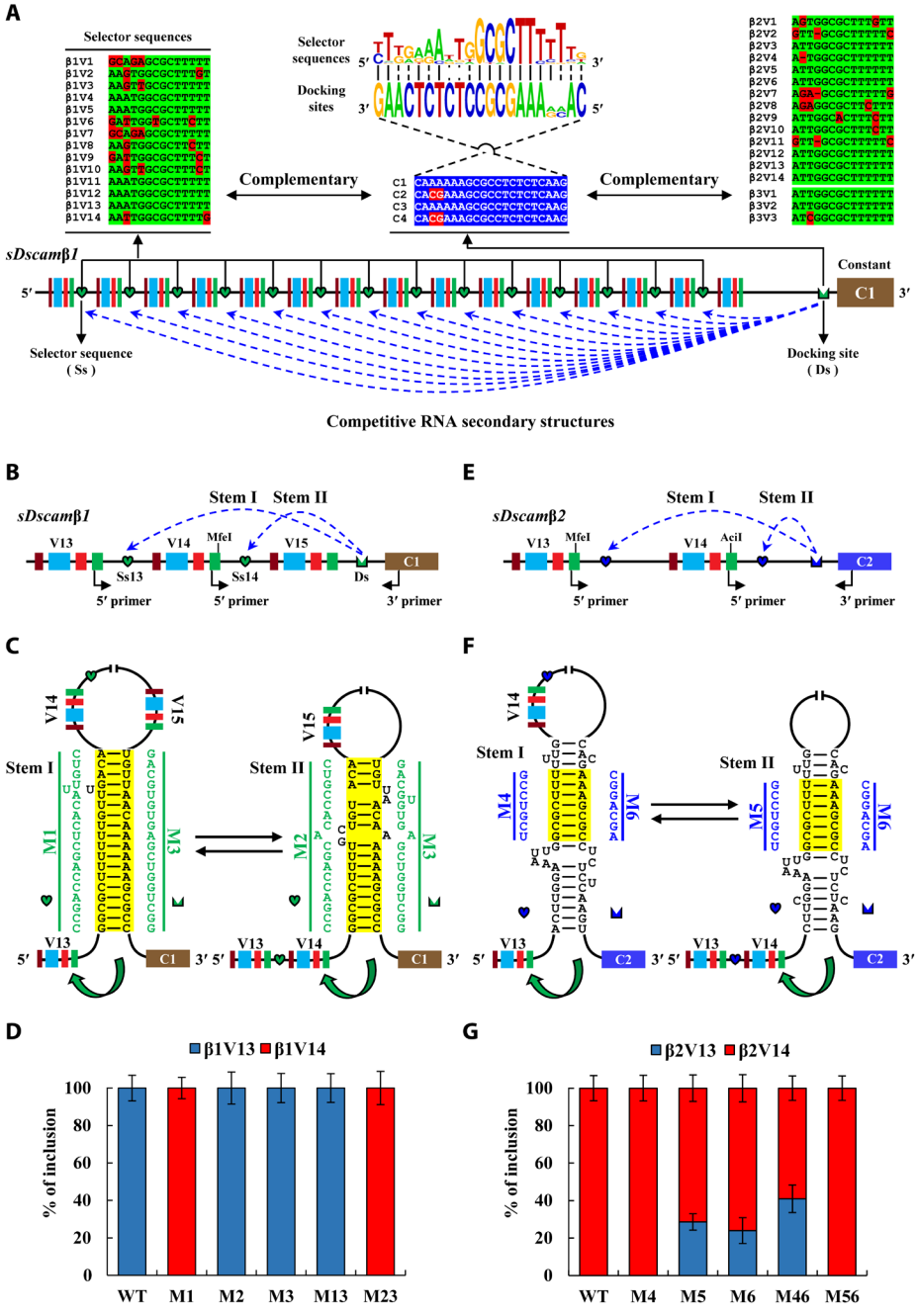

Fig. 2. Intronic competing RNA pairings mediating the selection of the 5′ alternative cassette of *sDscam*βs. See also figs. S4 to S8. (**A**) Arrangement of cis intronic elements in the 5′ variable region. The symbols used are the same as those in Fig. 1B. The sequences shown above are consensus intronic sequences. The most identical nucleotides in the selector sequences and docking sites are shown in green and blue, respectively. Sequences of the same color are highly similar. The downstream docking site (marked by crowns) was reverse complementary to the upstream selector sequences (marked by hearts) in a competitive mode. For the predicted intronic RNA pairings, see fig. S5. (**B** and **E**) Schematic diagrams of minigene constructs used to assess the effects of RNA secondary structure on 5′ alternative cassette selection. (**C** and **F**) Predicted competing RNA pairings. Mutations introduced into the double-stranded RNA (dsRNA) are indicated on the left or right mutated sequences (M1–M3, M4–M6). Green arrows depict the activated inclusion of the alternative exon. (**D** and **G**) Validation of the effects of competing RNA pairings by disruptive single mutations (M1–M3, M4–M6) and compensatory double mutations (M1 + M3: M13; M2 + M3: M23; M4 + M6: M46; M5 + M6: M56). Quantitation of gel data. Data are expressed as means ± SD from three independent experiments. See also figs. S7 and S8.

cassette inclusion almost completely (Fig. 2D and fig. S7, A and B), which reveals that β1V14 cassette inclusion is dependent specifically on the Ss14 element. Ds mutation (M3), which disrupted its pairing with Ss13 or Ss14, exhibited similar cassette inclusion to wild-type (WT) control (Fig. 2D). It is curious that a structure-restoring double mutation (M1 + M3, M13) increased the efficiency of β1V13 cassette inclusion to almost 100% (Fig. 2D). Conversely, double compensatory mutation of Ss14 and Ds (M2 + M3, M23) led to the inclusion

of predominantly β1V14 cassette, as β1V13 cassette was almost entirely excluded (Fig. 2D). Collectively, these data indicate that the selection of β1V13 cassette and β1V14 cassette depends on intronic competitive base pairing. Similar results were obtained when disruptive and compensatory mutations were introduced into the sDscamβ2 and sDscamβ3 minigenes (Fig. 2, E to G, and figs. S7 and S8). We also noticed that the relative expression of β2V13 and β2V14 in minigenes differed from the endogenous expression (Figs. 1C and 2G), which was likely attributed to usage of the heterologous expression system and minigene constructs containing only partial *sDscam* sequence. Together, our data indicate that competing RNA pairing plays an important role in regulating alternative selection of the 5′ variable region in this highly complex *sDscam* locus.

Competing intermolecular RNA pairing facilitates alternative trans-splicing

As noted above, extensive isoforms were generated through trans-splicing between the variable exon and constant region of various *sDscam*βs. Previous studies have indicated that RNA secondary structures can facilitate trans-splicing (*45*, *46*). It is conceivable that the docking site could be predicted to pair with almost any selector sequence from another gene cluster (fig. S5), which could bring the pre-mRNAs into physical contact and thus facilitate trans-splicing. We also found moderate correlations between the expression frequency of trans-splicing variants and the strength of docking site–selector pairings (fig. S6). To determine whether the *sDscam*β pre-mRNAs can form double-stranded RNA intermediates that facilitate trans-splicing (Fig. 3A), we cotransfected a plasmid-borne 5′ construct carrying the *sDscamβ1* variable exon 4 and partial intron 4 containing the selector sequence, and a 3′ construct carrying the *sDscamβ2* constant exon 5 and partial neighboring intron 4 containing the docking site, into *Drosophila* S2 cells. Then, we detected trans-spliced products between exon 4 of *sDscamβ1* and exon 5 of *sDscamβ2* (Fig. 3B). The combination of the *sDscamβ1* 5′ construct and various 3′ constructs from the constant regions of *sDscamβ2*, *sDscamβ3*, and *sDscamβ4* was designed to test the effects of intermolecular RNA secondary structures on trans-splicing through the disrupting and structure-restoring of double mutations (Fig. 3, C and D). In the WT control, 5′ exons of β1V13 cassette were efficiently trans-spliced to the constant exon 5 of *sDscamβ2* (Fig. 3E, I, lane 1). When the selector sequence (Ss) was mutated (M1), trans-splicing products were nearly undetectable (Fig. 3E, I, lane 2). Similar outcomes were obtained when the docking site sequence (Ds) was mutated (M2; Fig. 3E, I, lane 3). These observations reveal that trans-splicing depends on these cis-elements.

A structure-restoring double mutation (M1 + M2, M12) restored the efficiency of the inclusion of trans-splicing products to the level of the WT (Fig. 3E, I, lane 4). These observations demonstrate that these elements enhance trans-splicing through the formation of base pairs. Similar results were obtained for disruptive and compensatory mutations in the combinations *sDscamβ1* and *sDscamβ3* (II) or *sDscamβ1* and *sDscamβ4* (III; Fig. 3, D and E). Meanwhile, similar results were obtained from combining various 5′ constructs from the variable regions of *sDscamβ2* and *sDscamβ3* with 3′ constructs from the constant region of *sDscamβ4* (Fig. 4, A to C). As the common selector sequences paired competitively with various docking sites, and vice versa, intermolecular RNA secondary structures promoted trans-splicing between *sDscamβ1*–*β4* transcripts in a competitive manner. Together, these data on disruptive and compensatory mutations demonstrate that intermolecular RNA secondary structures are required for the trans-splicing of *sDscamβ1*–*β4* transcripts.

To preclude the possibility that the chimeric mRNAs detected above were artificially generated through homology-driven template switching during RT-PCR, we split the enhanced green fluorescent protein (EGFP) coding sequence (CDS) into two halves that were separately fused with intronic sequences from β1V13 cassette and *sDscamβ2* on two plasmids (Fig. 4D) (*47*). The upstream exon (EG) was followed by the intron 4 sequence of β1V13 cassette, whereas the downstream exon (FP) was fused with intron 4 of *sDscamβ2*. In such a system, EGFP can be expressed only when the two exons are trans-spliced. In the WT control, fluorescence was visible under a fluorescence microscope (Fig. 4E). When the selector sequence (Ss) was mutated (EGm), EGFP products were undetectable (Fig. 4, E and F). A similar outcome was obtained when the docking site (Ds) was mutated (mFP; Fig. 4, E and F). However, a structure-restoring double mutation (EGm + mFP) recovered the efficiency of EGFP expression to the level of the WT (Fig. 4, E and F). These results confirm that intermolecular RNA secondary structures play a critical role in trans-splicing.

Cis- and trans-spliced isoforms mediate homophilic binding activity

To examine whether trans- and cis-spliced sDscamβ isoforms mediate homophilic binding and the possible mechanism behind this process, we expressed sDscamβ proteins in Sf9 cells using an insect baculovirus expression system in an aggregation assay (Fig. 5A), as described previously (*34*). We constructed most of the *sDscam*β isoforms that can be formed through cis- and trans-splicing (Fig. 5, B and C, and fig. S9A). These isoforms were examined after the infection of Sf9 cells with baculovirus vectors encoding individual sDscamβ C-terminal–mCherry fusion proteins. Cell aggregation was then visualized under a fluorescence microscope, and the sizes of cell aggregates were quantified. We performed systematic analyses of homophilic binding for 86 of the 132 sDscamβ proteins (30 of 33 cis-spliced sDscamβs and 56 of 99 trans-spliced sDscamβs). We found that 24 of the 33 cis-spliced sDscamβ isoforms formed homophilic aggregates when assayed individually (Fig. 5C and fig. S9A). Of the tested isoforms generated through trans-splicing, 47 of 56 formed homophilic aggregates (Fig. 5C and fig. S9A). Together, these findings show that both cis- and trans-spliced sDscamβ proteins can mediate cell aggregation.

The size of cell aggregates varied greatly among individual cis- and trans-spliced isoforms according to the results of the quantitative assay (Fig. 5C and fig. S9, B and C). The presence of naturally occurring trans-spliced isoforms between *sDscamβ1*–*β4* allows for finer dissection of the mechanisms through which variable and constant domains contribute to cell aggregation. We observed marked differences in aggregation ability among isoforms with the same variable region but different constant regions (Fig. 5C). For example, sDscamβ1V7C1 exhibited strong cell aggregation, whereas sDscamβ1V7C3 exhibited little cell aggregation. These data indicate that the distinct constant domains generated through trans-splicing may influence homophilic trans-binding activity.

The variability in aggregation activity among sDscamβ isoforms is likely due to differences in the expression, membrane localization, or intrinsic trans-binding affinities of individual isoforms (*26*). However, immunostaining revealed that both sDscamβ2V2C2 and β3V1C2, which does not mediate cell aggregation, and sDscamβ1V11C1,

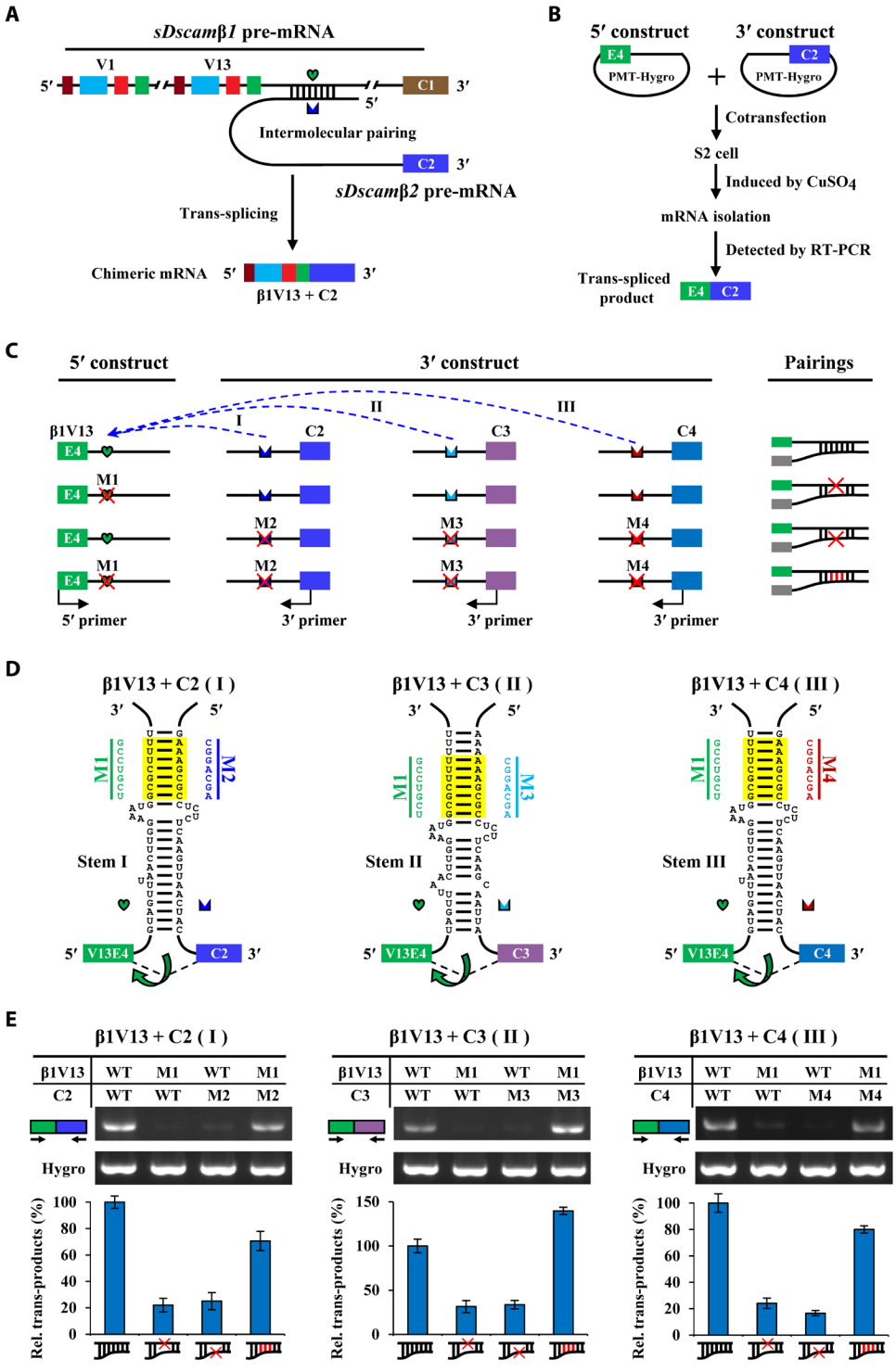

Fig. 3. Intronic RNA pairings facilitating trans-splicing of *sDscam*β variables with different constants. (**A**) Potential model of *sDscam*β trans-splicing facilitated by intermolecular RNA secondary structures. (**B**) Schematic of a trans-splicing system in *Drosophila* S2 cells. (**C**) Schematic diagrams using minigene constructs to assess the effects of RNA base pairing on trans-splicing between different *sDscam*βs. For the predicted intronic RNA pairings between different *sDscam*βs, see fig. S5. (**D**) Predicted intermolecular base pairings of pre-mRNAs. Mutations introduced into dsRNA are indicated on the left or right mutated sequences (M1–M4). Green arrows depict the activation of trans-splicing between different *sDscam*βs. (**E**) Validation of the effects of RNA pairings on trans-splicing by disruptive single mutations (M1–M4) and compensatory double mutations (M1 + M2: M12; M1 + M3: M13; M1 + M4: M14). Data were quantified below their gels. Data are expressed as means ± SD from three independent experiments.

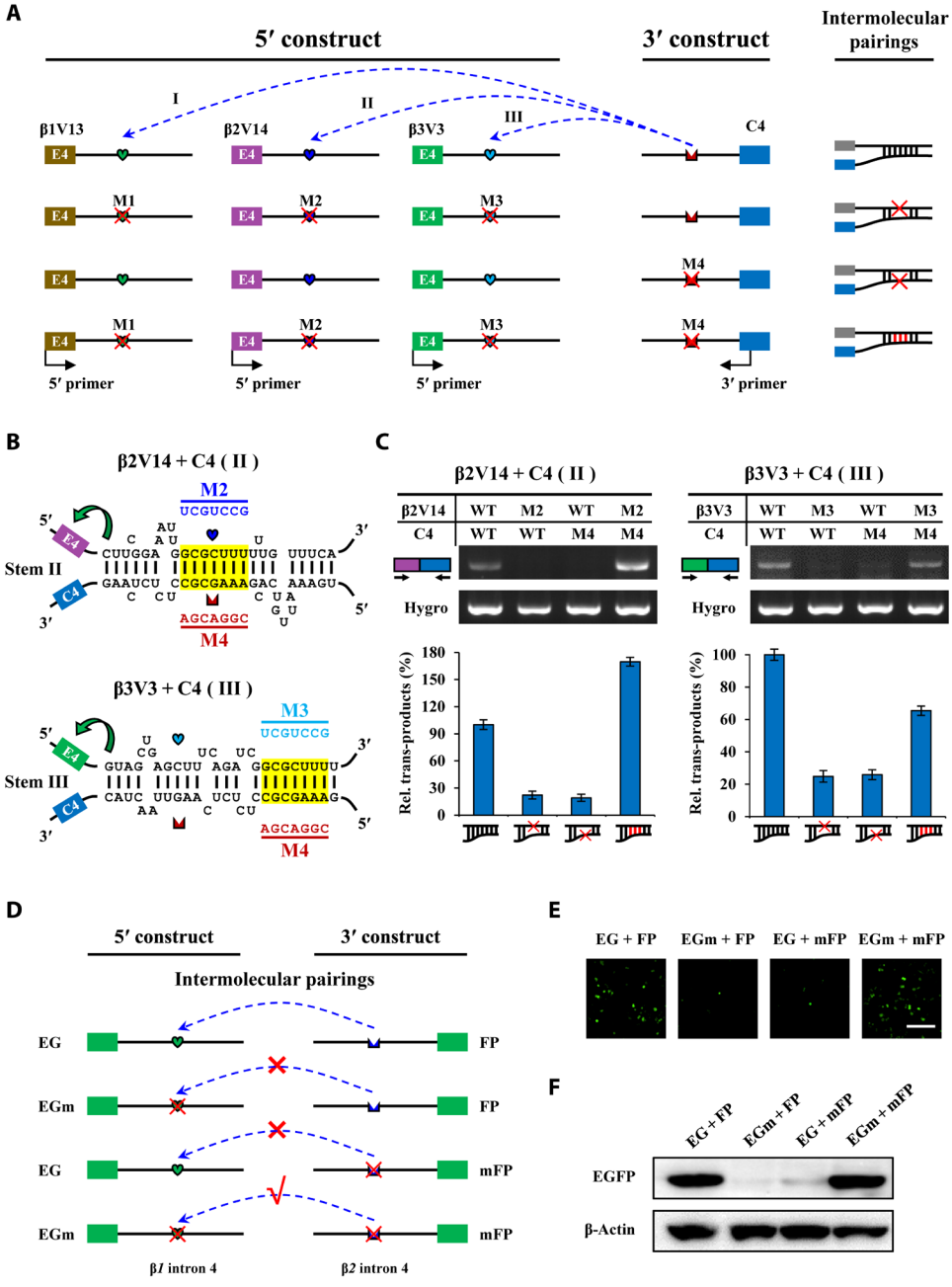

Fig. 4. RNA pairings enhancing trans-splicing of the *sDscam*β constant with different variables. (**A**) Schematic diagrams of minigene constructs used to assess the effects of RNA secondary structure on trans-splicing of *sDscam*β*4*. (**B**) Predicted competing RNA pairings. Mutations introduced into dsRNA are indicated on the upper or lower mutated sequences (M2, M3, M4). Stem I has been validated in Fig. 3. Green arrows depict the activation of trans-splicing between different *sDscam*βs. (**C**) Validation of the effects of RNA pairings on trans-splicing by disruptive single mutations (M2, M3, M4) and compensatory double mutations (M24: M2 + M4; M34: M3 + M4). Data are expressed as means ± SD from three independent experiments. (**D**) Validation of trans-splicing at the protein level. The CDS of EGFP was split into two halves (EG and FP), followed by intronic sequences of *sDscam*β. EG: exon EG with 1 to 154 nt of intron 4 of the β1V13 cassette; FP: exon FG with 272 to 892 nt of intron 4 of β2. (**E**) Fluorescent photos of S2 cells transfected with WT and mutant plasmids containing CDS of EG/FP fused with intronic sequences from the β1V13 cassette and β2 (scale bars, 100 μm). (**F**) Detection of trans-spliced products by Western blotting using anti-EGFP antibody.

β3V1C3, and β1V13C4, which engages in homophilic interactions, were present on the surface of Sf9 cells (fig. S10A). Moreover, some isoforms (i.e., sDscamβ1V10C1, β2V2C2, β3V2C3, and β2V6C4), which does not mediate cell aggregation, was expressed at a similar level to isoforms that mediate cell aggregation (fig. S10B). Therefore, we hypothesize that different outcomes of cell aggregation mediated by individual sDscamβ isoforms might be attributed to differences in intrinsic affinities among isoforms.

To further identify the regions of *T. urticae* sDscamβ proteins responsible for homophilic interactions, we performed systematic

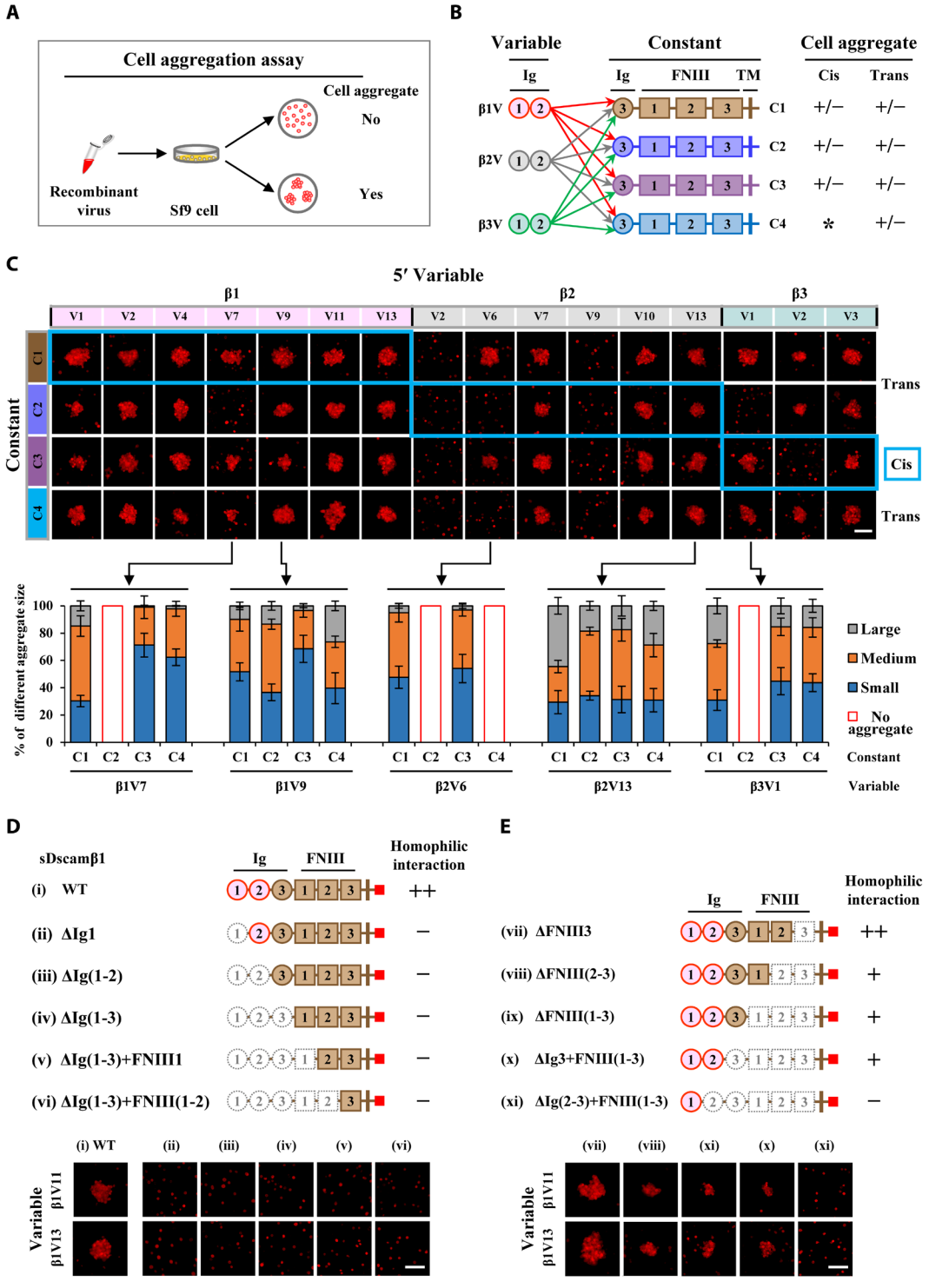

Fig. 5. Cluster-wide analysis of homophilic binding of trans- and cis-spliced sDscamβ isoforms in *T. urticae*. See also fig. S9. (**A**) Schematic diagram of cell aggregation experiments. The mCherry-tagged sDscamβ proteins were expressed in Sf9 cells to test their ability to form cell aggregates. (**B**) Schematic diagram of the combination between the 5′ variable region and 3′ constant regions of *sDscamβ* to form cis- and trans-spliced isoforms. The results of homophilic binding are summarized on the right-hand side. * indicates the lack of the 5′ variable region. Cis, cis-spliced isoforms; Trans, trans-spliced isoforms. (**C**) Homophilic binding of 64 cis- and trans-sDscamβ isoforms. Data quantitation of representative isoforms is shown. Data are expressed as means ± SD from three independent experiments. These data indicate that constant domains influence homophilic trans-binding ability. See also fig. S9. (**D**) A series of N-terminal truncations of the extracellular domain of sDscamβ fused to mCherry were subjected to cell aggregation assays. All sDscam truncations lacking the N-terminal Ig1 domain failed to form cell aggregates. (**E**) A series of domain deletion truncations were performed starting from the membrane-proximal FNIII3 domain. These data indicate that homophilic trans-binding is associated with constant extracellular domains of sDscamβ (scale bars, 100 μm).

aggregation assays with the series sDscamβ1V11C1/β1V13C1, in which extracellular domains were successively deleted starting with N-Ig1 or membrane-proximal FNIII3 (Fig. 5, D and E). No cell aggregation was observed when the Ig1 domain was deleted from sDscamβ, which suggests that the Ig1 domain is essential for homophilic interactions (Fig. 5D). Conversely, aggregation was apparent when up to four extracellular domains were deleted from the membrane-proximal FNIII3 domain of sDscamβ (Fig. 5E). These domain-deleted isoforms of sDscamβ have been located on the cell membrane (fig. S10A), thus precluding the possibility of the lack of membrane localization. Together, these findings demonstrate that sDscamβ-mediated cell surface recognition and binding depends on variable regions and that different constant regions generated through trans-splicing can affect cell aggregation capacity.

sDscamβs exhibit N-terminal variable domain–specific binding that is independent of cis- or trans-splicing

To determine the process through which cis- and trans-spliced sDscamβs engage in specific homophilic interactions, we conducted a series of experiments to test the binding specificity of pairwise sDscamβ isoform combinations. Each protein was expressed with mCherry or GFP fused to its C terminus to provide an easily observable assay of cell homophilic aggregates (Fig. 6A). To determine the stringency of recognition specificity, we generated pairwise sequence identity heatmaps of the variable regions (Fig. 6B). Using these heatmaps, we identified sDscamβ pairs with greater than 80% pairwise sequence identity in their variable Ig1-Ig2 domains. We hypothesized that if the two most closely related sDscamβs could not bind to each other, it would be impossible for two distantly related sDscamβs to recognize each other. Unlike other Chelicerata species, closely related pairs of mite sDscamβs originate from different gene clusters. For example, sDscamβ1V9C1 and sDscamβ2V10C2 share 97.6% amino acid sequence identity within their Ig1-Ig2 domains.

Pairwise sDscamβ isoform combinations were created. Among the 170 sDscamβ pairs with different variable Ig1-Ig2 domains tested here, only self-pairs on the matrix diagonals displayed intermixing of mCherry- and GFP-expressing cells; all nonself pairs were fully segregated into homophilic aggregates of red and green cells (Fig. 6, C and D, and fig. S11, A and B). Even the pair β1V9C1/β2V10C2, with 97.6% sequence similarity of the Ig1-Ig2 domain, formed segregated red and green aggregates (Fig. 6D, panel IV, and fig. S11B). By contrast, the 48 sDscamβ pairs with the same variable Ig1-Ig2 domains exhibited intermixing of mCherry- and GFP-expressing cells (Fig. 6E and fig. S11C). Thus, trans-spliced sDscams, which share the same homophilic Ig domain as their cis-spliced counterparts, did not increase homophilic specificity. These data indicate that homophilic specificity depends on N-variable Ig domains and is independent of the constant regions derived from cis- or trans-splicing.

Binding specificity of trans- and cis-spliced sDscamβ isoforms is mediated by the Ig1 domain

Previous studies have shown that the first Ig domain is the primary determinant of trans interaction specificity in scorpion sDscamαs and sDscamβs (*34*). However, unlike in other Chelicerata species, we have not found *sDscamα* subfamily containing only a variable Ig domain in *T. urticae* (Fig. 1A). To examine whether variable Ig1 is responsible for trans interaction specificity between sDscamβs, we constructed a series of chimeras exhibiting variable Ig domain swapping within a single clustered gene or between different genes (Fig. 7, A to C). In all sDscamβ pairs tested, swapping the Ig1 domain of a given sDscamβ for that of another led to a shift in binding specificity (Fig. 7, A to C). By contrast, swapping the Ig2 domain or constant region of sDscamβ resulted in no change in specificity. Further analyses of domain shuffling showed that only isoforms with the same Ig1 domain could recognize each other among all pairs investigated (Fig. 7, A to C). These data indicate that the Ig1 domain of sDscamβ is necessary and sufficient to determine its adhesive specificity, at least among the isoforms investigated here.

We used homology modeling to generate homophilic binding complexes between sDscamβs, and the Ig1 domain interacted in an antiparallel orientation (Fig. 7D and fig. S12). This result confirms that the first Ig domain of *T. urticae* sDscamβ determines its trans homophilic interaction specificity between proteins on apposing cell surfaces, as has been reported for scorpion sDscamαs and sDscamβs. Thus, the specific loss of *sDscamα* counterparts in *T. urticae* did not affect the function of the Ig1 domain of sDscamβ in determining trans interaction specificity.

Coexpression of multiple trans- and cis-spliced sDscamβ isoforms expands homophilic specificity

Previous studies on Pcdh isoforms have revealed that recognition specificity is diversified through the coexpression of multiple isoforms (*25*, *26*). Meanwhile, our recent research showed that scorpion sDscamαs and sDscamβs can produce combinatorial recognition specificity (*34*). To test the possibility that recognition specificity is diversified by cis- and trans-spliced isoforms, we coexpressed multiple sDscamβ isoforms with different N-variable domains. Sf9 cells were coinfected with trans- and cis-spliced sDscamβ isoforms, which were tagged with mCherry and GFP, respectively. In all cases, cells that coexpressed the same set of sDscamβ1 isoforms formed intermixed yellow aggregates (Fig. 8A). By contrast, cells coexpressing a set of two cis-spliced sDscamβ1 isoforms formed separate nonadhering aggregates with cells expressing a different set of two sDscamβ1s. Further coimmunoprecipitation experiments confirmed the interaction between two different isoforms coexpressed in Sf9 cells (fig. S13A). Similar results were obtained for each of the trans- and cis-spliced sDscamβ pairs shown in Fig. 8 (B to D). However, cells that coexpressed two sDscamβ isoforms formed mixed aggregates containing cells expressing each of the two sDscamβs with the same variable region but different constant regions (Fig. 8E). These results suggest that a single mismatched sDscamβ isoform that differs in its N-variable domain can interfere with combinatorial homophilic interactions, whereas sDscamβ that differs in its constant domain cannot.

We further coexpressed distinct sets of three sDscamβ isoforms and analyzed their ability to mediate homophilic specificity in cells with various numbers of mismatches (Fig. 8F). We found that cells that expressed mismatched isoforms with different N-variable domains generally formed separate red and green aggregates, and only cells that expressed identical isoform combinations formed robust mixed yellow aggregates (Fig. 8F). Cells that coexpressed three sDscamβ isoforms coaggregated with cells that expressed a different set of three sDscamβ isoforms containing the same variable region but different constant regions (Fig. 8F). These data indicate that trans-spliced sDscamβs share the same combinatorial homophilic specificity as their cis-spliced counterparts (Fig. 8G and fig. S13, B and C).

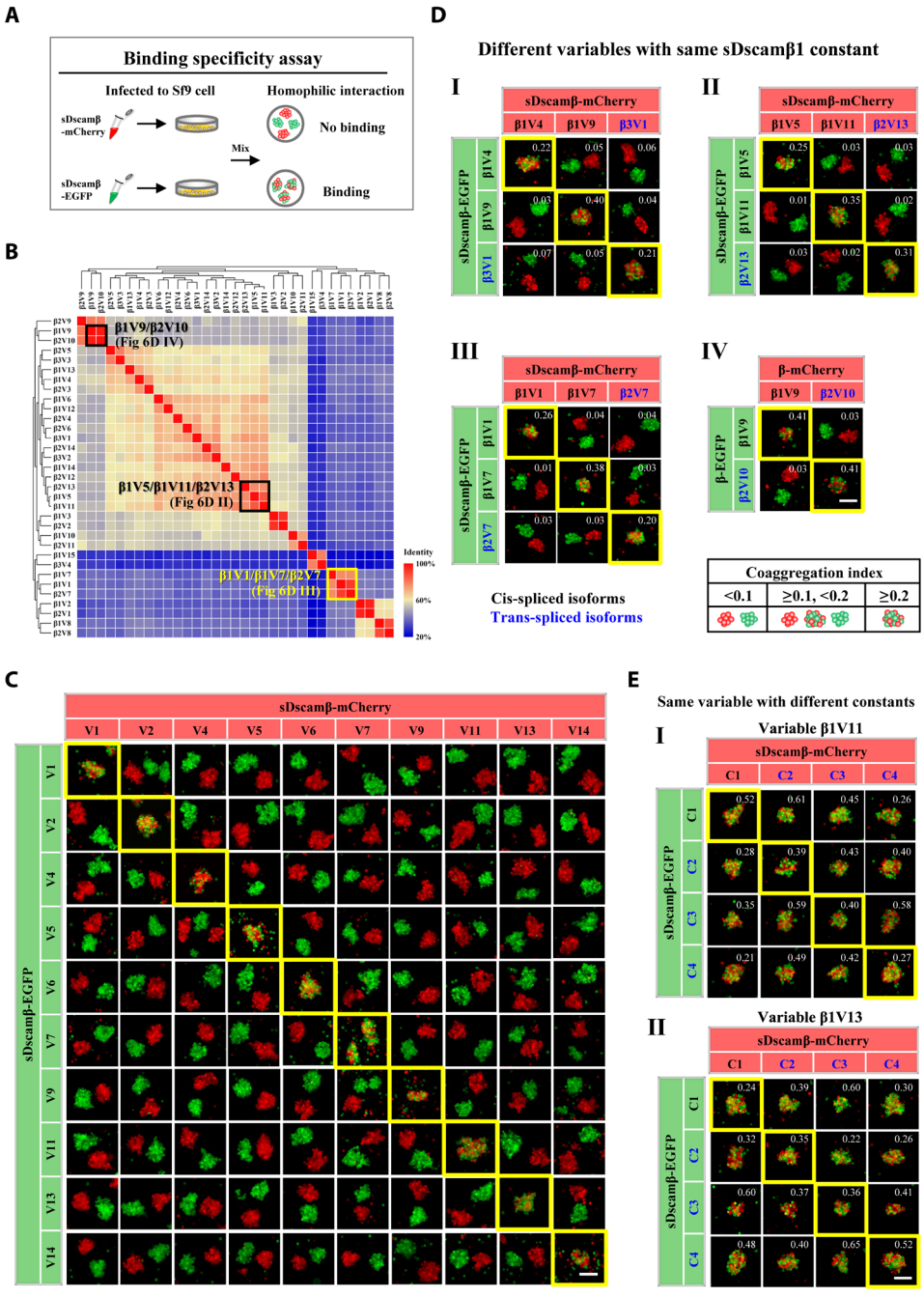

Fig. 6. *T. urticae* sDscamβ isoforms exhibiting N-variable domain–specific binding. See also fig. S11. (**A**) Schematic diagram of the interaction specificity assay. Cells expressing mCherry- or EGFP-tagged sDscamβ isoforms were mixed and analyzed for homophilic or heterophilic binding. The state of cell aggregation includes red-green cell coaggregation or segregation. (**B**) Heatmap of pairwise amino acid sequence identity of the variable region of sDscamβ isoforms and their clustering relationships. Subsets of the isoforms within the boxed region were assayed in (D). See also fig. S11 (A and B). (**C**) Cis-spliced sDscamβ1 isoforms displaying strict binding specificity. (**D**) Cis- and trans-spliced sDscamβ1 isoforms with 50 to 97.6% sequence identity for nonself pairs in their variable regions exhibiting strict trans homophilic specificity. (**E**) Cis- and trans-spliced sDscamβ pairs with the same variable Ig1-Ig2 domains displaying red-green cell coaggregation. Mean coaggregation indices were quantified and illustrated by numbers in the corresponding fluorescent photos (scale bars, 100 μm).

DISCUSSION

We found that trans-splicing markedly expands the sDscamβ isoform repertoire of *T. urticae*. We were surprised to find that every variable exon cassette engages in trans-splicing with constant exons from another cluster. Moreover, we provide evidence that intronic competing RNA pairings govern alternative cis- and trans-splicing. Note that these trans-spliced sDscamβ isoforms mediate cell adhesion activity while sharing the same homophilic binding specificity as their cis-spliced counterparts. Thus, we have identified a single extreme *sDscam* locus that generates broad adhesion molecular

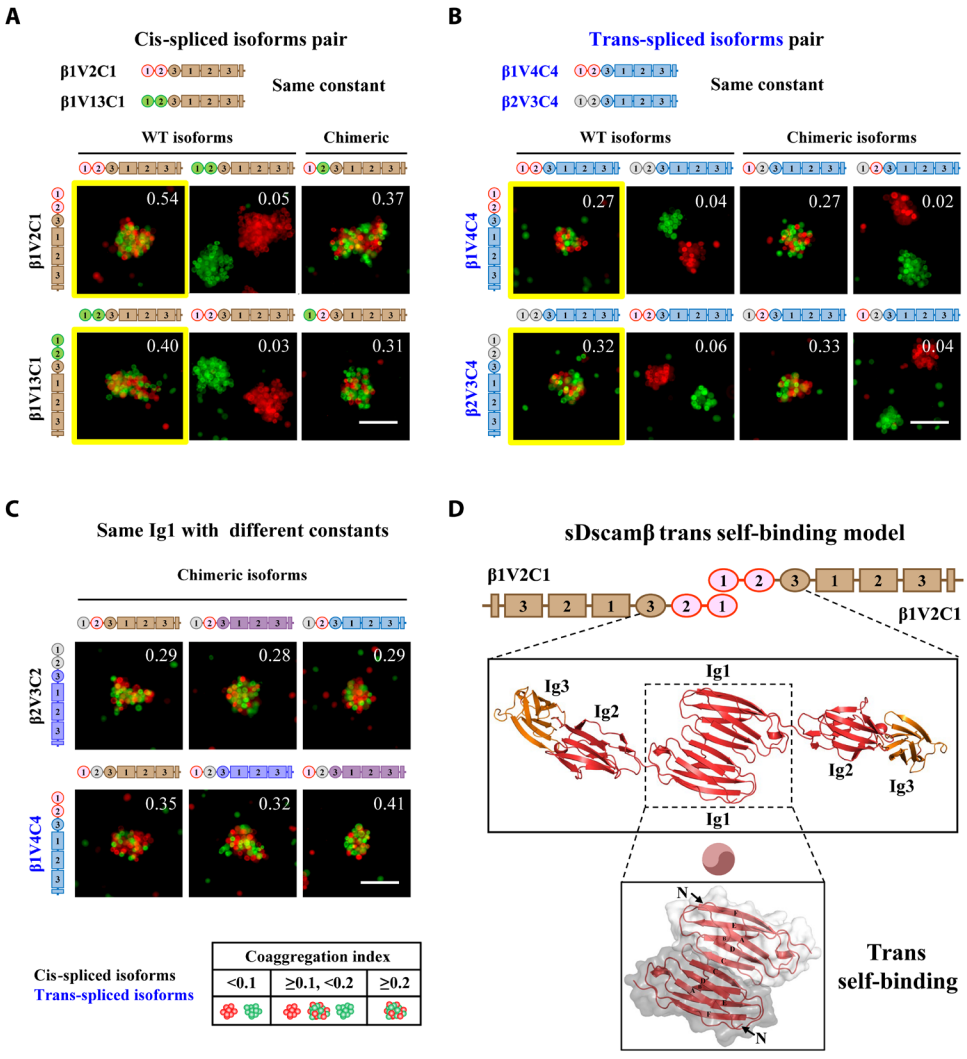

Fig. 7. Homophilic interaction specificity depends on the variable Ig1 domain. See also fig. S12. (**A** and **B**) Domain-specific recognition of the N-variable mediated by the sDscamβ Ig domain shuffled isoform. Domain-shuffled chimeras of sDscamβ isoforms and their parental counterparts were assayed for their binding specificity. Chimeras in which the Ig1 domain was replaced by the corresponding domain swapped binding specificity, whereas the Ig2 replacement did not. (**C**) sDscamβ1 pairs with the same variable Ig1 domain do not display recognition specificity. Mean coaggregation indices are shown in the top right corner of each representative image (scale bars, 100 μm). (**D**) Schematic diagram of trans interactions of sDscamβ. Structural modeling shows that the Ig1 domain of sDscamβ interacts in an antiparallel manner.

diversity through alternative cis- and trans-splicing coupled with alternative promoters and combinatorial homophilic recognition units. Below, we discuss the combinatorial mechanism of sDscam isoform diversity and the potential significance of *sDscam* trans-splicing in neuronal circuits, with particular emphasis on comparison to vertebrate Pcdhs.

A combinatorial mechanism of sDscam molecular diversity

Our present findings indicate that trans-splicing between distinct *sDscam* gene clusters produces a previously unidentified set of chimeric transcripts at a high frequency in *T. urticae*. The frequency of trans-spliced isoforms was estimated to be up to 60% based on the exon junctions predicted from the RNA-seq data. However, the occurrence rate of trans-splicing was obviously underestimated, as we calculated trans-splicing only between pairs of distinct *sDscam*βs. Intracluster trans-splicing likely also occurred, but we could not distinguish whether these isoforms resulted from cis- or trans-splicing. Our data demonstrate that every variable exon cassette engages in trans-splicing with constant exons from another cluster belonging to the same *sDscam* locus. Thus, extensive *sDscam* isoforms are produced through a combination of alternative promoter choice and alternative cis- and trans-splicing processes (Fig. 9A). The frequent occurrence of trans-splicing in this mite reflects the intricate mechanism of *sDscam* pre-mRNA splicing, which compensates for the exceptionally low number of Dscam isoforms in *T. urticae*.

We explored the processes through which chimeric transcripts were frequently produced from the *sDscam* locus in *T. urticae*. We propose that three nonexclusive mechanisms might be involved. First, the unusual genomic organization of *sDscam* may facilitate trans-splicing between different transcripts. In this case, the nascent transcripts generated from a single locus are geometrically close before leaving their transcription sites (Fig. 9B). A similar genomic architecture frequently occurs in genes that undergo trans-splicing, such as *mod* (*mdg4*) and *lola* in flies (*41*, *47*–*49*). This possibility is

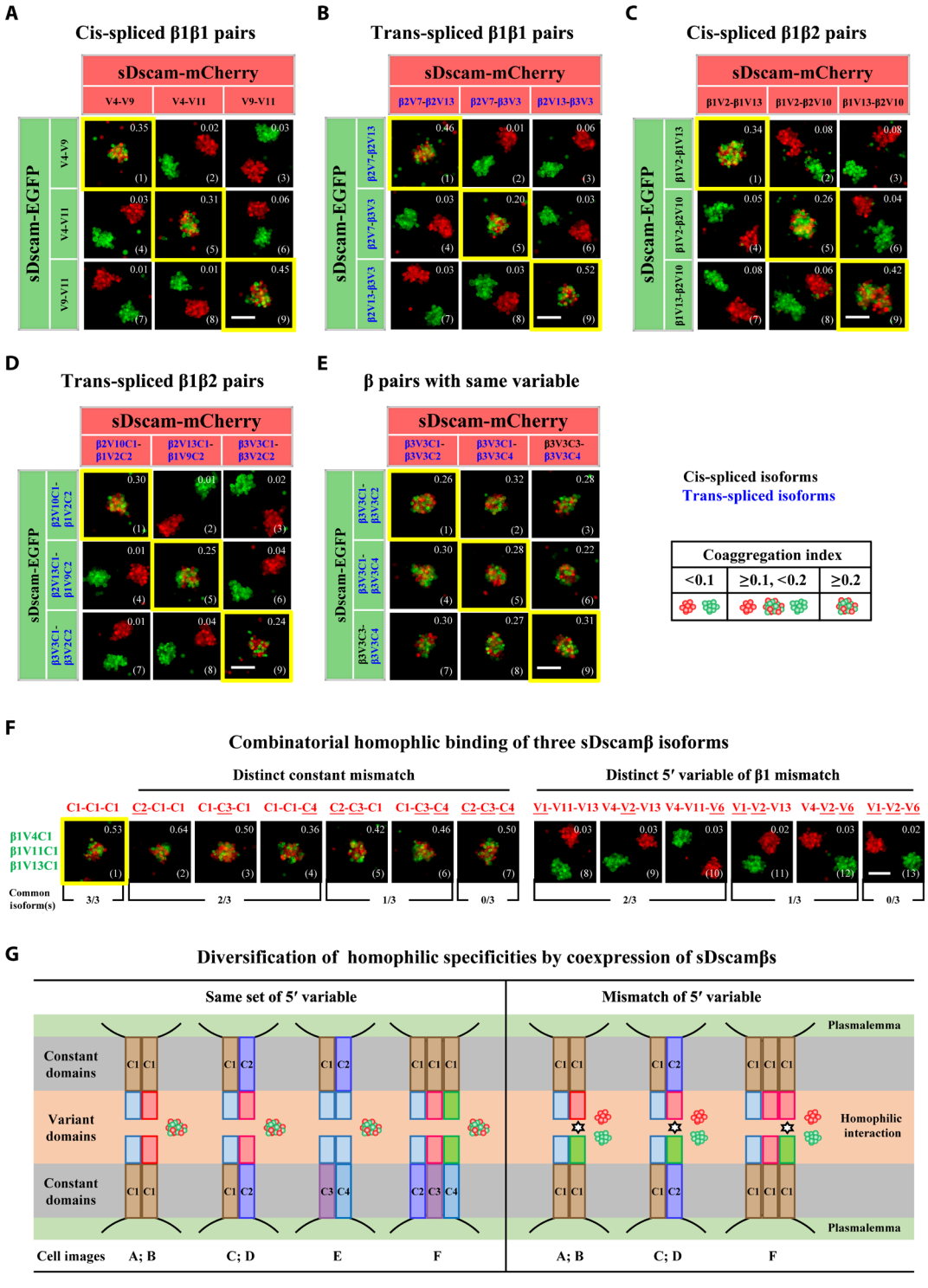

Fig. 8. Combinatorial homophilic specificity resulting from coexpression of distinct cis- and trans-spliced sDscamβ isoforms. See also fig. S13. (**A** to **D**) Cells coexpressing different combinations of differentially tagged sDscamβ isoform pairs were mixed and assayed for their coaggregation. β1 cis-spliced isoforms (A), β1 trans-spliced isoforms (B), β1/β2 cis-spliced isoforms (C), and β1/β2 trans-spliced isoforms (D) were measured. (**E**) The combination of cis- and trans-spliced sDscamβ pairs with the same variable Ig1-Ig2 domains did not exhibit the combinatorial homophilic specificity. (**F**) Analysis of the interaction of cells coexpressing three different GFP tags with cells expressing the same or different groups of mCherry tags. The underline marks the mismatched isoforms between the two cell groups. Mean coaggregation indices for (A) to (F) are shown in the top right corner of each representative image (scale bars, 100 μm). (**G**) Schematic diagram of the outcome of combinatorial homophilic specificity. The diagram shown here does not reflect cis multimers.

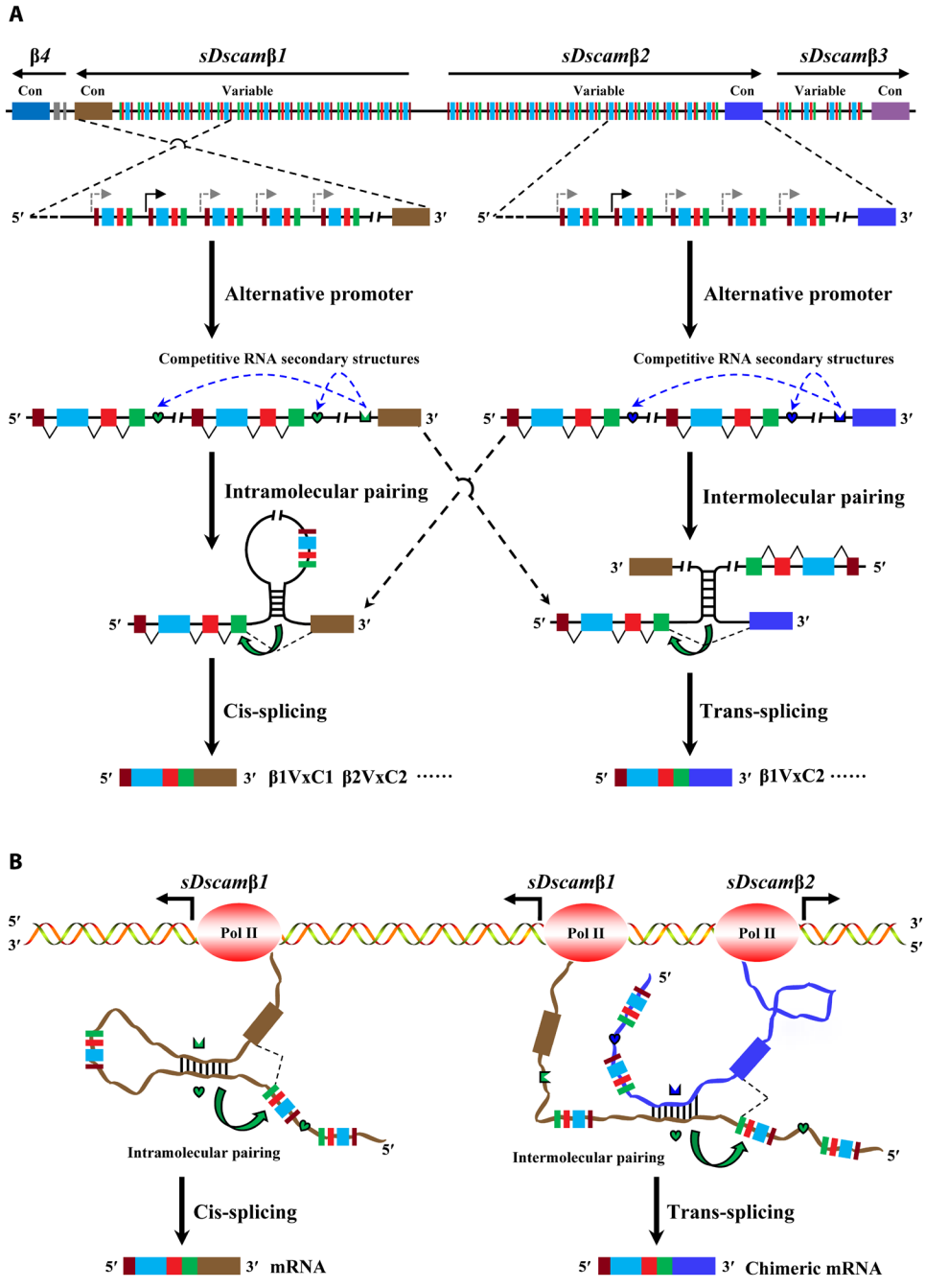

Fig. 9. A single complex *sDscam*β locus generating a marked adhesion molecular diversity. (**A**) Generation of extensive *sDscam* isoforms through a combination of alternative promoter choices and cis- and trans-alternative splicing. On the left is the schematic representation generating cis-spliced sDscamβ isoform diversity. This alternative cis-splicing process is mediated by a competing RNA secondary structure between the docking site and selector sequences. On the right is a schematic representation of the generation of trans-spliced sDscamβ isoforms. This trans-splicing process is facilitated by intronic intermolecular RNA secondary structures. (**B**) Schematic representation of sDscamβ diversity mediated by cotranscriptional RNA folding and alternative cis- and trans-splicing. These nascent transcripts generated from this single locus are geometrically close to each other before leaving their transcription sites, which facilitates trans-splicing between different *sDscam*β transcripts.

supported by the recent observation that cross-strand chimeric RNAs were generated through the fusion of bidirectional transcripts in humans (*50*). We suggest that base pairing between these convergent sense-antisense transcripts promotes trans-splicing, as reported in *Caenorhabditis elegans eri-6/7* (*45*). Second, the compatibility of splice sites between the constant and variable exon cassettes of different *sDscam*β*1*–β*4* transcripts may contribute to efficient trans-splicing between these regions. Third, and perhaps most important, intermolecular base pairing between the docking site and selector sequence can bring these regions into spatial proximity, thereby facilitating trans-splicing between two distinct *sDscam*β*1*–β*4* transcripts (Fig. 9B). Thus, the extensive occurrence of trans-splicing in the mite *sDscam*

locus is driven by the specific evolution of intermolecular base pairing. Many lines of evidence have shown that RNA secondary structures can enhance trans-splicing between different transcripts (*45*, *51*, *52*). However, the present study provides the clearest bioinformatics-based and experimental evidence of a process through which complex intermolecular base-pairing interactions mediate alternative trans-splicing.

Meanwhile, this study demonstrates that competing base pairing between the docking site and selector sequence mediates complex 5′ alternative splicing. This competing base-pairing system was initially identified in the exon 6 cluster of fly *Dscam1* (*42*) and was believed to be unique. Similar structural codes have recently been revealed in several exon clusters, including *Drosophila 14-3-3ξ*, the exon 4 and 9 clusters of *Drosophila Dscam1*, *srp*, *Branchiostoma MRP*, and human *dynamin 1* and *CD55* genes (*43*, *44*, *53*–*57*). This docking site–mediated base-pairing process also regulates alternative splicing at the 3′ end, such as in *Drosophila PGRP-LC*, *CG42235*, and *Pip* (*58*). In the present study, we found that base pairing of the docking site and selector sequence mediated alternative cis- and trans-splicing in the 5′ variable region. Therefore, competing base pairing is a widespread mechanism of regulating mutually exclusive splicing and other RNA processing events.

The significance of trans-splicing and chimeric sDscam transcripts

Note that trans-spliced sDscams share the same homophilic Ig domain as their cis-spliced counterparts. The cell aggregation assay in the present study showed no differences in binding specificity between cis- and trans-spliced sDscams pairs with identical N-variable Ig1 domains. However, extracellular Ig3-FNIII3, transmembrane (TM), and cytoplasmic domains encoded by *sDscam* constant regions have distinct sequences and likely play vital roles. First, the present study shows that homophilic binding capacity can be strongly affected by the constant region (Fig. 5). A similar result was obtained for sDscam of the scorpion *M. martensii* (*34*), in which the constant region affects the formation and size of cell aggregates. Biophysical experiments showed that cis interactions of cPcdh are generally promiscuous but with a preference for the formation of heterologous cis dimers (*59*). If this finding holds true in the sDscam family, we would expect that differences in the constant region would affect cis interactions. Therefore, the different constant regions produced through trans-splicing might influence the strength of downstream signaling. Alternatively, given the extraordinary diversity of trans-spliced isoforms, we posit that the alternative function (i.e., immunity) of mite sDscams depends on a receptor-ligand interaction that may involve different constant domains. For example, the constant domains of an antibody support specific recognition and response functions in adaptive immunity (*60*, *61*).

Second, the distinct TM domains formed through trans-splicing may lead to differences in protein localization. For example, TM domains mediate Dscam1 protein targeting and then alter dendritic elaboration and axonal arborization (*62*–*65*). TM1-containing Dscam1 is targeted toward dendrites and mainly regulates dendritic development, whereas TM2-containing Dscam1 is mainly expressed in axons and mediates axonal arborization. Last, the cytoplasmic domain of sDscam regulates intracellular signaling pathways. Several lines of experimental evidence have shown that specific cytoplasmic domains of Dscam1 are essential to neural development (*64*, *66*, *67*). Conversely, the cytoplasmic domain of Pcdhs mediates homophilic interactions and intracellular trafficking. Despite the relatively weak cellular adhesion observed in Pcdh-γs mediated by the variable extracellular domain, the constant C-terminal cytoplasmic domain of Pcdh-γs regulates dendrite arborization through the binding and inhibition of focal adhesion kinase (*68*, *69*). Moreover, diverse cytoplasmic domains of Pcdh-γs are critical to late endosome and lysosome trafficking during synapse development (*70*–*72*). Therefore, homophilic interactions between cis- and trans-spliced sDscams may provide a basis for signal transduction in neuronal development and circuit formation. The functional significance of extensive trans-splicing of sDscams remains to be determined.

Comparison of fly Dscam1 and mite sDscams to mammalian Pcdhs

Extensive Dscam diversity is unique to arthropods, which use two main mechanisms to produce isoform diversity. Insect *Dscam* genes use exclusive alternative splicing to generate distinct isoforms, whereas chelicerate *Dscam*s use alternative promoters. Unlike other chelicerates, mites encode all clustered *sDscam* isoforms from a single genomic locus. Thus, mite *Dscam*s and mammalian *Pcdh*s have remarkable parallels: Both encode notably diversified isoforms from a single locus covering three tandemly arranged gene clusters, and both are organized as a tandem array in the 5′ variable region (fig. S14). For both genes, each variable cassette is generally preceded by a given promoter (*23*, *24*, *33*). In addition, transcription of the *Pcdh* gene clusters is regulated via long-range chromatin-looping interactions (*2*, *73*), and we speculate that *sDscam* genes are transcribed through a similar regulatory mechanism. Moreover, the present results combined with our previous data indicate that clustered *sDscam*s exhibit N-variable Ig-specific homophilic binding in a manner similar to *Pcdh*s (*34*). Last, given the remarkable parallels among and complementary phylogenetic distribution of mite sDscam, fly Dscam1, and mammalian Pcdhs, we suggest that mite sDscam may play a similar role to mammalian Pcdhs and fly Dscam1 in self-/non–self-discrimination and other neuronal functions (*1*, *5*, *30*, *31*, *74*).

Despite their overall similarities, mite *sDscam* and mammalian *Pcdh* genes differ in at least two major aspects of their mechanisms underlying molecular diversity (fig. S14). The first difference is that the variable cassette of mite clustered *sDscam*βs is generally composed of four exons, whereas each variable cassette of the clustered *Pcdh* gene is composed of a single large exon. This multi-exon organization may increase sDscam isoform diversity through alternative cis-splicing combination of variable exons or through trans-splicing between different genes. Therefore, mite clustered *sDscam*s appear to have more complex splicing patterns in the 5′ variable region than clustered *Pcdh*s. The second major difference is related to the splicing of the 5′ variable region. Previous studies of the *Pcdh* locus have shown that every variable exon engages in trans-splicing with constant exons from another cluster, albeit at a very low level (*23*, *24*), which suggests that the occurrence of chimeric mRNA does not reflect the primary mechanism of *Pcdh* pre-mRNA splicing. The lack of apparent intermolecular base pairing between different Pcdh transcripts may explain their low trans-splicing frequency compared to mite sDscam. Together, the insights obtained and framework developed in this study help to clarify the mechanisms of molecular diversity and trans-splicing.

MATERIALS AND METHODS

Cell lines and cell cultures

Spodoptera frugiperda 9 (Sf9) cells (a gift from J. Chen, Zhejiang Sci-Tech University) were cultured in Sf-900 II SFM (Gibco, 10902088)

supplemented with 10% heat-inactivated fetal bovine serum (Gibco, 10099141) and 1% penicillin-streptomycin (Gibco, 15140163) at 27°C.

Drosophila Schneider 2 (S2) cells (male) were maintained in Schneider's Drosophila medium (Gibco, 21720-024) supplemented with 10% heat-inactivated fetal bovine serum and 1% penicillin-streptomycin (Gibco, 15140163) at 27°C.

Animals

Two-spotted spider mites (*T. urticae*) (a gift from X. Hong, Nanjing Agricultural University) were used in this study.

Availability of genome and RNA-seq data

We investigated *T. urticae* in Chelicerata (*40*). The source of *T. urticae* genome sequence (CAEY00000000.1) used in this study was obtained from the National Center for Biotechnology Information (NCBI) (www.ncbi.nlm.nih.gov/). For *Dscam* candidate validation, we selected 45 publicly available RNA-seq data corresponding to various developmental stages, longevity, and stress treatment (table S1).

Annotation and identification of *Dscam* genes

Dscam genes of Mesostigmata *Metaseiulus occidentalis*, Trombidiformes *Ixodes scapularis*, Araneae *Stegodyphus mimosarum* and *Parasteatoda tepidariorum*, Scorpiones M. martensii, and Merostomata *Limulus polyphemus* have been previously described (*32*). Sequences of Dscam homologs of *T. urticae* were annotated by cross-species BLAST searches using the available annotated Dscam sequences (https://blast.ncbi.nlm.nih.gov/Blast.cgi). These Dscam candidate homologs were further validated further using publicly available RNA-seq datasets. All Dscam candidates were confirmed by phylogenetic analysis using MEGA X and then analyzed by predicting protein domains using InterPro (www.ebi.ac.uk/interpro/), SMART (http://smart.embl-heidelberg.de/), and PROSITE (https://prosite.expasy.org/prosite.html).

Analysis of RNA-seq data

Exon junctions

Using an in-house computational program, we calculated exon-exon spliced junctions within or between genes to investigate sequencing evidence (*33*). Briefly, the exonic sequences covering all possible junctions of variable exons were first created, and a given number of reads were assigned to an exon-exon junction using 10-nt positions per exon in a pair (table S1). For example, the 180-nt exonic sequence includes 90-nt upstream and 90-nt downstream junctions for the 100-nt RNA-seq reads. Next, all RNA-seq reads were mapped to the exonic sequences created above, and perfectly mapped RNA-seq reads covering the exon-exon junctions were kept. Because of the high sequence similarity of the *T. urticae sDscamβ1* constant (C1) and *sDscamβ3* constant (C3), a match length of at least 52 nt is required except for the 10-nt positions. For example, on the one hand, the length of RNA-seq reads matching the 3′ end of the query sequence is at least 10 nt. In addition, because 52 nt of the 90-nt exonic sequences was identical and could not be used to distinguish C1 from C3, the length of the remaining matching variable regions in 100-nt RNA-seq reads is 38 nt. On the other hand, the length of RNA-seq reads matching the 5′ end of the query sequence is at least 10 nt. Because C1 and C3 have a 52-nt sequence in common, the remaining length in the 100-nt RNA-seq reads used to distinguish the C1 and C3 constant regions is also 38 nt. On the basis of these analyses, we used a 128-nt full query sequence.

Analysis of differential and biased expression

The expression of *sDscam* genes in *Wolbachia* infection, various developmental stages, longevity, acaricide treatment, and feeding was analyzed using RNA-seq data from publicly accessible samples (table S1). To quantify the expression level of each *sDscam* gene from the replicates, we calculated the values of the reads in the constant exonic region for each sample. Alternative exons encoding Ig1 were selected to calculate the expression level of the replicates for each 5′ variable cassette. Considering the short length of the alternative exons, RNA-seq reads were divided into 25-nt segments for mapping using Bowtie 2 software, and only perfectly mapped fragments were retained for expression level calculations. In addition, the read counts of a 25-nt fragment from multiple loci were split by the number of loci, and then each locus was assigned equally for expression level calculations. To eliminate the effects of identical sequences among exon duplications on expression calculations, expression profiles were generated using 25-nt fragmented RNA-seq datasets, as previously reported (*33*).

Reverse transcription PCR

Whole bodies from *T. urticae* were collected for RNA preparation. Total RNA was isolated using TRIzol reagent (Invitrogen, 15596026) and reverse-transcribed using oligo(dT) primer and SuperScript III RTase (Invitrogen, 18080-093). RT-PCR was performed with initial denaturation at 94°C for 3 min, followed by 30 to 35 cycles of denaturation at 94°C for 30 s, annealing at 60° to 65°C for 30 s, and extension at 72°C for 15 s, with a final extension at 72°C for 10 min.

Phylogenetic analysis

The nucleotide sequences for all 5′ variable cassettes and constant exons of *sDscam* were translated into amino acid sequences, and the resulting sequences were aligned. Genetic distances for each sequence were estimated using MEGA X software.

Sequence alignments and RNA pairing predictions

The alignments of the conserved regions between distinct variable exon cluster of *T. urticae sDscam* were done using the Clustal Omega program (www.ebi.ac.uk/Tools/msa/clustalo/). The consensus sequences of the docking site and selector sequences were derived using WebLogo (http://weblogo.berkeley.edu/logo.cgi). The intronic RNA pairings between the selector sequences and the docking site were predicted using Mfold (www.unafold.org/mfold/applications/rna-folding-form.php). Because of the limitations of the Mfold program, only the docking site, the selector sequences, and their flanking sequences were used as input for Mfold.

Plasmid construction of *sDscam*

Minigene construction for cis-splicing system

Genomic DNA isolated from *T. urticae* was used as a template, and PCR was carried out with primers (table S3) and PrimeSTAR DNA Polymerase (TaKaRa, R045Q) to obtain the corresponding DNA segments encompassing variable exon clusters, constant exons, and intervening sequences (Fig. 2, B and E, and figs. S7, A and C, and S8A). WT minigene DNAs were cloned into the pEASY-blunt zero cloning vector (TransGen Biotech, CB501-01). The minigene constructs were further cloned behind the metallothionein promoter in the pMT/V5-His B vector.

Minigene construction for trans-splicing system

Variable exon accompanied with downstream intron and constant exon accompanied with the upstream intron of different gene clusters

were amplified by PCR (table S3) from *T. urticae* genomic DNA. Moreover, two sets of PCR products were inserted into a modified pMT/V5-His B vector with hygromycin B and P copia promoter (a gift from Y. Xu, Wuhan University) (*75*), which were cotransfected into S2 cells (as described below). In addition, intron 4 sequences of β1V13 and C2 were amplified from genomic DNA. The CDS of EGFP was amplified from the pEGFP N1 vector (a gift from N. Zhou, Zhejiang University) and split into two halves (EG and FP) after nucleotide G489 (*47*). Minigene constructs containing the EG/FP CDS followed by the β1V13/C2 intron 4 sequences were constructed in the modified pMT/V5-His B vector, respectively (Fig. 4D).

Disruptive and compensatory mutations of RNA elements

Site-directed mutagenesis was conducted on both the docking site and selector sequences to disrupt the RNA secondary structure in the pEASY-blunt zero cloning vector. Structure-restoring double mutations of RNA elements were performed to restore the RNA stem structure based on the schematic diagrams. Primer (table S3) sequences used for PCR amplification will be provided upon request.

Plasmid construction for sDscam isoform expression

Extracellular domains and TM domains were predicted using PROSITE (https://prosite.expasy.org/prosite.html) and SMART (http://smart.embl-heidelberg.de/). DNA fragments encoding isoforms lacking the cytoplasmic domain or partially lacking the extracellular domain were amplified by PCR using cDNA isolated from *T. urticae*. PCR products were cloned into the pEASY-blunt zero cloning vector (TransGen), followed by recombination to ligate the DNA fragments with the pFastBacHTB-mCherry/EGFP/Myc/HA expression vector using the pEASY-Uni Seamless Cloning and Assembly Kit (TransGen, CU101-01), respectively. pFastBacHTB-mCherry/EGFP/Myc/HA vectors were generated by inserting mCherry/EGFP/Myc/HA DNA sequences into pFastBacHTB vectors (a gift from X. Wu, Zhejiang University) by overlapping PCR. To obtain the pFastBac-Dual Myc-mCherry/HA-GFP vector, sequences encoding Myc/HA peptides were synthesized and annealed to form a double strand and then cloned into the pFastBac-Dual vector, and the mCherry/GFP peptides were then inserted behind another promoter of pFastBac-Dual vector (a gift from J. Chen, Zhejiang Sci-Tech University). All recombinant vectors were confirmed by DNA sequencing (*34*). Primer sequences used for PCR amplifications are listed in tables S4 to S7.

Minigene transfection

For plasmids used in S2 cells, minigene constructs were transfected into 50 to 70% confluent S2 cell lines using Lipofectamine 3000 Reagent (Invitrogen) according to the manufacturer's protocol, and $CuSO_4$ was added after 5 hours to induce plasmid expression. Cells were harvested after 48 hours of treatment. In experiments where two minigenes were cotransfected, the two plasmids were mixed together before being mixed with transfection reagents, and these mixtures were then transfected into 50 to 70% confluent cells using Lipofectamine 3000 Reagent (Invitrogen, L3000015). After 48 hours of $CuSO_4$ treatment, cells were harvested.

Quantification of mRNA splice isoforms

We assayed the RNA splice isoform ratio using RT-PCR followed by exon-specific restriction digestion. Total RNA was isolated from S2 cell lines transfected with the *T. urticae sDscam*β construct. RT-PCR products were then digested by exon-specific restriction enzymes. Images were captured using a charge-coupled device camera, and quantification of mRNA isoforms was achieved by comparing the integrated optical density of the detected bands measured by the GIS 1D Gel Image System (Tanon, version 3.73).

Recombinant baculovirus production

Baculoviruses of sDscam isoforms were produced by the Bac-to-Bac Baculovirus Expression System (Gibco, 10359016). The process was as follows: The pFastBac plasmid containing the *sDscam* segment was transformed into DH10Bac competent cells (Biomed, BC112), blue-white screening was used to obtain positive colonies, and then the recombinant bacmid DNA was identified by PCR with M13 primer and *sDscam*-specific primer. Recombinant bacmid was transfected into 50 to 70% confluent Sf9 cells using Lipofectamine 3000 Reagent. P1 viral stock was collected 6 to 8 days after transfection. To amplify the titer of the virus, 50 to 70% confluent sf9 cells were infected with the p1 virus to obtain the P2 viral stock. All baculoviruses were stored at 4°C, protected from light, or stored at −80°C for long-term storage (*34*).

Cell aggregation assays

Sf9 cells were infected with recombinant P2 viruses of mCherry- or GFP-tagged sDscam isoforms and incubated in six-well plates at 27°C for 3 days. To pretreat the six-well plates for cell aggregation assays, unused six-well plates were incubated overnight at 4°C with 1% bovine serum albumin (BSA) in 1× Hanks' balanced salt solution (HBSS; 1:10; Gibco, 14185052) and washed three times with 1× HBSS, and finally, 2 ml of ice-cold 1× HCMF [4-(2-Hydroxyethyl) piperazine-1-ethanesulfonic acid (HEPES)] (1:10; Leagene Biotechnology, CC0073) was added to each well. Infected cells were collected and centrifuged at 1000 rpm for 5 min and then resuspended with 1 ml of ice-cold 1× HCMF. Four hundred microliters of cell suspension from each sample was transferred to each well of pretreated six-well plates for single fluorescence cell aggregation assays, and 200 μl of cell suspension from each sample was transferred jointly for binding specificity assays. Cell suspensions in six-well plates should be gently mixed at 27°C in a gyratory shaker (IKA KS260) at 60 rpm for 30 min. Last, images were captured using a Nikon Ti-S inverted fluorescence microscope (*34*, *76*).

Quantification of cell aggregates using MATLAB

Quantitative analysis of cell aggregates was carried out using an in-house computational program written in MATLAB. The "aggregation" and "no aggregation" were distinguished by the number of pixels in each object, with objects smaller than 300 pixels (~3 cells) being classified as no aggregation to exclude the individual large cell or dividing cell, and objects larger than 300 pixels being classified as aggregation. The percentage of cell aggregation was calculated by dividing the number of aggregation objects by the number of all objects in each image. For aggregate size quantification, objects between 300 pixels and 1000 pixels (3 to 10 cells) were categorized as "small," objects between 1000 pixels and 3000 pixels (10 to 30 cells) were categorized as "medium," and objects larger than 3000 pixels (>30 cells) were categorized as "large." The number of aggregates for each size category was then counted. Images used for quantification were obtained from three independent cell aggregation experiments (*34*).

Immunofluorescence

Sf9 cells were seeded onto coverslips (WHB Scientific, WHB-6-CS) in six-well plates that were precoated with 1 mM poly-L-lysine (Sigma-Aldrich, P6282). P2 viral stocks of sDscam proteins inserted

with the c-Myc tag between FNIII3 and the TM domain and the mCherry tag at the C terminus were transfected into 50 to 70% confluent Sf9 cells (fig. S10A). After 72 hours, the cells were fixed and washed, and after being blocked with 5% BSA, the cells were then incubated with anti-Myc tag monoclonal antibody (1:4000; EarthOx, catalog no. E022050-01) overnight at 4°C. Cells were washed three times with Dulbecco's phosphate-buffered saline (D-PBS) before being incubated with goat anti-mouse IgG (H+L) Dylight488 (1:5000; EarthOx, catalog no. E032210-01) for 1 to 2 hours at room temperature. Last, the nuclei were stained with Hoechst (2 μg/ml; Invitrogen, Hoechst 33342) for 15 to 30 min. Last, cells were imaged using a laser scanning confocal microscope LSM800 (Carl Zeiss) (*34*).

Heatmap analysis of the *sDscam*β variable region

Multisequence alignments of the *sDscam*β variable region were carried out using Clustal Omega (www.ebi.ac.uk/Tools/msa/clustalo/), and the sequence similarity heatmap was generated by TBtools (*77*).

Binding specificity assay for cells expressing single or multiple sDscam isoform(s)

After 3 days of infection, Sf9 cells expressing differentially tagged sDscam isoforms of *T. urticae* were mixed. Coexpression of multiple sDscam isoforms was applied in an appropriate ratio to roughly guarantee approximately equal surface expression. Images were captured using a Nikon Ti-S inverted fluorescence microscope, capturing red and green fluorescence, and merged by Nikon software, and aggregates containing red cells only, green cells only, and both red and green cells were analyzed for binding specificity (*34*).

Calculation of coaggregation index

The coaggregation index was calculated according to a previous study on delta protocadherins (d-Pcdhs) (*78*). Fluorescence images were analyzed using a custom code written in MATLAB (*34*). An image with completely red/green segregated cells would have a very low coaggregation index (<0.1). In contrast, an image containing intermixed red and green cells would achieve a high coaggregation index (≥0.2), while an image with partially intermixed red and green cells would have an intermediate index (≥0.1 and <0.2).

Homology modeling and protein-protein docking

Ig1-Ig3 homology models of sDscam were built using RoseTTAFold (https://robetta.bakerlab.org/). The Ig1-Ig3 domain was then used by the M-ZDOCK server (https://zdock.umassmed.edu/m-zdock/) for homologous dimer docking. Last, the PyMOL package (https://pymol.org/2/) was used to visualize the models.

Antibodies

The primary antibodies were used in isoform coimmunoprecipitation: anti-HA (hemagglutinin) tag rabbit polyclonal antibody (1:50; EarthOx, catalog no. E022180-01, RRID:AB_2811272). Coimmunoprecipitation samples were probed with anti-HA tag mouse monoclonal antibody (1:5000; EarthOx, catalog no. E022010-01) and anti-Myc tag mouse monoclonal antibody (1:5000; EarthOx, catalog no. E022050-01).

Western blotting primary antibodies were used for relative quantification of sDscam: anti-mCherry tag mouse monoclonal antibody (1:5000; EarthOx, catalog no. E022110-01, RRID:AB_2687920) and anti–β-actin mouse monoclonal antibody (1:5000; Abcam, catalog no. ab8224, RRID:AB_449644). Western blotting primary antibodies were used for relative quantification of EGFP: anti-EGFP tag mouse monoclonal antibody (1:5000; EarthOx, catalog no. E022030-01) and anti–β-actin mouse monoclonal antibody (1:5000; Abcam, catalog no. ab8224, RRID:AB_449644). Last, secondary antibody was used for all Western blots: horseradish peroxidase AffiniPure goat anti-mouse IgG (1:8000; EarthOx, catalog no. E030110-01, RRID:AB_2572419).

Coimmunoprecipitation

Sf9 cells were infected with recombinant viruses containing HA- or Myc-tagged sDscam and incubated in six-well plates at 27°C for 3 days. Infected cells were washed three times with ice-cold D-PBS, collected by centrifugation at 1000*g* for 5 min at 4°C, and then homogenized in immunoprecipitation lysis buffer (Thermo Fisher Scientific, 87787) supplemented with 100× phenylmethylsulfonyl fluoride (PMSF; Beyotime, ST505) and 100× ProteinSafe Protease Inhibitor Cocktail (TransGen, DI111-01).

The supernatant was incubated with anti-HA tag rabbit polyclonal antibody (1:50) at 4°C overnight while mixing. After washing Pierce Protein A/G Magnetic Beads (Thermo Fisher Scientific, 88802) according to guidelines, the antigen sample/antibody mixture was added to the prewashed magnetic beads and incubated at 4°C for 1 to 3 hours while mixing. Subsequently, the beads were collected using a magnetic stand and washed three times with beads wash buffer. Next, the collected beads were mixed with 80 μl of 5× Protein Loading Dye (Sangon Biotech, C508320-0001) and heated at 96° to 100°C for 10 min. Last, the beads were magnetically separated and the supernatant was stored at −80°C or used for Western blotting.

Western blot

For relative expression quantification, infected cells were lysed in radioimmunoprecipitation assay lysis buffer (strong) (Cowin Biosciences, CW2333S) supplemented with 100× PMSF (Beyotime, ST505) and 100× ProteinSafe Protease Inhibitor Cocktail (TransGen, DI111-01). The protein lysate was centrifuged to remove debris. The supernatant was mixed with 5× Protein Loading Dye (Sangon Biotech, C508320-0001) and heated at 96° to 100°C for 10 min. Then, the mixed sample was centrifuged at 13,000 rpm for 20 min at 4°C. The sample was separated by 10% Precast-Glgel Tris-Glycine PAGE (Sangon Biotech, C651101-0001) and transferred to polyvinylidene difluoride (PVDF) membranes (Millipore, IPVH00010). After probing with the respective antibodies, the PVDF membranes were finally analyzed with SuperSignal West Femto Maximum Sensitivity Substrate (Thermo Fisher Scientific, 34095).

Statistical analysis

The effect was considered statistically significant at $P < 0.05$. To examine significant differences in cell aggregation mediated by sDscam isoforms, statistical significance was calculated using IBM SPASS Statistics V22.0 (Student's *t* tests). Similarly, Mann-Whitney *U* tests were used to examine the statistical significance of cell aggregation size among sDscam isoforms using IBM SPASS Statistics V22.0 (*34*).

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at https://science.org/doi/10.1126/sciadv.abn9458

View/request a protocol for this paper from *Bio-protocol*.

REFERENCES AND NOTES

1. S. L. Zipursky, W. B. Grueber, The molecular basis of self-avoidance. *Annu. Rev. Neurosci.* **36**, 547–568 (2013).
2. G. Mountoufaris, D. Canzio, C. L. Nwakeze, W. V. Chen, T. Maniatis, Writing, reading, and translating the clustered protocadherin cell surface recognition code for neural circuit assembly. *Annu. Rev. Cell Dev. Biol.* **34**, 471–493 (2018).
3. J. R. Sanes, S. L. Zipursky, Synaptic specificity, recognition molecules, and assembly of neural circuits. *Cell* **181**, 1434–1435 (2020).
4. D. Schmucker, J. C. Clemens, H. Shu, C. A. Worby, J. Xiao, M. Muda, J. E. Dixon, S. L. Zipursky, Drosophila Dscam is an axon guidance receptor exhibiting extraordinary molecular diversity. *Cell* **101**, 671–684 (2000).
5. S. L. Zipursky, J. R. Sanes, Chemoaffinity revisited: Dscams, protocadherins, and neural circuit assembly. *Cell* **143**, 343–353 (2010).
6. S. K. Miura, A. Martins, K. X. Zhang, B. R. Graveley, S. L. Zipursky, Probabilistic splicing of Dscam1 establishes identity at the level of single neurons. *Cell* **155**, 1166–1177 (2013).
7. G. Neves, J. Zucker, M. Daly, A. Chess, Stochastic yet biased expression of multiple Dscam splice variants by individual cells. *Nat. Genet.* **36**, 240–246 (2004).
8. W. Sun, X. You, A. Gogol-Doring, H. He, Y. Kise, M. Sohn, T. Chen, A. Klebes, D. Schmucker, W. Chen, Ultra-deep profiling of alternatively spliced Drosophila Dscam isoforms by circularization-assisted multi-segment sequencing. *EMBO J.* **32**, 2029–2038 (2013).
9. W. M. Wojtowicz, J. J. Flanagan, S. S. Millard, S. L. Zipursky, J. C. Clemens, Alternative splicing of Drosophila Dscam generates axon guidance receptors that exhibit isoform-specific homophilic binding. *Cell* **118**, 619–633 (2004).
10. W. M. Wojtowicz, W. Wu, I. Andre, B. Qian, D. Baker, S. L. Zipursky, A vast repertoire of Dscam binding specificities arises from modular interactions of variable Ig domains. *Cell* **130**, 1134–1145 (2007).
11. D. Hattori, S. S. Millard, W. M. Wojtowicz, S. L. Zipursky, Dscam-mediated cell recognition regulates neural circuit formation. *Annu. Rev. Cell Dev. Biol.* **24**, 597–620 (2008).
12. D. Hattori, E. Demir, H. W. Kim, E. Viragh, S. L. Zipursky, B. J. Dickson, Dscam diversity is essential for neuronal wiring and self-recognition. *Nature* **449**, 223–227 (2007).
13. M. E. Hughes, R. Bortnick, A. Tsubouchi, P. Baumer, M. Kondo, T. Uemura, D. Schmucker, Homophilic Dscam interactions control complex dendrite morphogenesis. *Neuron* **54**, 417–427 (2007).
14. B. J. Matthews, M. E. Kim, J. J. Flanagan, D. Hattori, J. C. Clemens, S. L. Zipursky, W. B. Grueber, Dendrite self-avoidance is controlled by Dscam. *Cell* **129**, 593–604 (2007).
15. P. Soba, S. Zhu, K. Emoto, S. Younger, S. J. Yang, H. H. Yu, T. Lee, L. Y. Jan, Y. N. Jan, Drosophila sensory neurons require Dscam for dendritic self-avoidance and proper dendritic field organization. *Neuron* **54**, 403–416 (2007).
16. D. Hattori, Y. Chen, B. J. Matthews, L. Salwinski, C. Sabatti, W. B. Grueber, S. L. Zipursky, Robust discrimination between self and non-self neurites requires thousands of Dscam1 isoforms. *Nature* **461**, 644–648 (2009).
17. Y. Jin, H. Li, Revisiting Dscam diversity: Lessons from clustered protocadherins. *Cell. Mol. Life Sci.* **76**, 667–680 (2019).
18. D. Schmucker, B. Chen, Dscam and DSCAM: Complex genes in simple animals, complex animals yet simple genes. *Genes Dev.* **23**, 147–156 (2009).
19. T. Yagi, Molecular codes for neuronal individuality and cell assembly in the brain. *Front. Mol. Neurosci.* **5**, 45 (2012).
20. Q. Wu, T. Maniatis, A striking organization of a large family of human neural cadherin-like cell adhesion genes. *Cell* **97**, 779–790 (1999).
21. Q. Wu, T. Zhang, J. F. Cheng, Y. Kim, J. Grimwood, J. Schmutz, M. Dickson, J. P. Noonan, M. Q. Zhang, R. M. Myers, T. Maniatis, Comparative DNA sequence analysis of mouse and human protocadherin gene clusters. *Genome Res.* **11**, 389–404 (2001).
22. S. Ribich, B. Tasic, T. Maniatis, Identification of long-range regulatory elements in the protocadherin-alpha gene cluster. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 19719–19724 (2006).
23. B. Tasic, C. E. Nabholz, K. K. Baldwin, Y. Kim, E. H. Rueckert, S. A. Ribich, P. Cramer, Q. Wu, R. Axel, T. Maniatis, Promoter choice determines splice site selection in protocadherin alpha and gamma pre-mRNA splicing. *Mol. Cell* **10**, 21–33 (2002).
24. X. Wang, H. Su, A. Bradley, Molecular mechanisms governing Pcdh-gamma gene expression: Evidence for a multiple promoter and cis-alternative splicing model. *Genes Dev.* **16**, 1890–1905 (2002).
25. D. Schreiner, J. A. Weiner, Combinatorial homophilic interaction between gamma-protocadherin multimers greatly expands the molecular diversity of cell adhesion. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 14893–14898 (2010).
26. C. A. Thu, W. V. Chen, R. Rubinstein, M. Chevee, H. N. Wolcott, K. O. Felsovalyi, J. C. Tapia, L. Shapiro, B. Honig, T. Maniatis, Single-cell identity generated by combinatorial homophilic interactions between α, β, and γ protocadherins. *Cell* **158**, 1045–1059 (2014).
27. R. Rubinstein, C. A. Thu, K. M. Goodman, H. N. Wolcott, F. Bahna, S. Mannepalli, G. Ahlsen, M. Chevee, A. Halim, H. Clausen, T. Maniatis, L. Shapiro, B. Honig, Molecular logic of neuronal self-recognition through protocadherin domain interactions. *Cell* **163**, 629–642 (2015).
28. J. Brasch, K. M. Goodman, A. J. Noble, M. Rapp, S. Mannepalli, F. Bahna, V. P. Dandey, T. Bepler, B. Berger, T. Maniatis, C. S. Potter, B. Carragher, B. Honig, L. Shapiro, Visualization of clustered protocadherin neuronal self-recognition complexes. *Nature* **569**, 280–283 (2019).
29. K. M. Goodman, R. Rubinstein, H. Dan, F. Bahna, S. Mannepalli, G. Ahlsen, C. Aye Thu, R. V. Sampogna, T. Maniatis, B. Honig, L. Shapiro, Protocadherin *cis*-dimer architecture and recognition unit diversity. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E9829–E9837 (2017).
30. J. L. Lefebvre, D. Kostadinov, W. V. Chen, T. Maniatis, J. R. Sanes, Protocadherins mediate dendritic self-avoidance in the mammalian nervous system. *Nature* **488**, 517–521 (2012).
31. G. Mountoufaris, W. S. V. Chen, Y. Hirabayashi, S. O'Keeffe, M. Chevee, C. L. Nwakeze, F. Polleux, T. Maniatis, Multicluster Pcdh diversity is required for mouse olfactory neural circuit assembly. *Science* **356**, 411–414 (2017).
32. G. Cao, Y. Shi, J. Zhang, H. Ma, S. Hou, H. Dong, W. Hong, S. Chen, H. Li, Y. Wu, P. Guo, X. Shao, B. Xu, F. Shi, Y. Meng, Y. Jin, A chelicerate-specific burst of nonclassical Dscam diversity. *BMC Genomics* **19**, 66 (2018).
33. Y. Yue, Y. Meng, H. Ma, S. Hou, G. Cao, W. Hong, Y. Shi, P. Guo, B. Liu, F. Shi, Y. Yang, Y. Jin, A large family of Dscam genes with tandemly arrayed 5′ cassettes in Chelicerata. *Nat. Commun.* **7**, 11252 (2016).
34. F. Zhou, G. Cao, S. Dai, G. Li, H. Li, Z. Ding, S. Hou, B. Xu, W. You, G. Wiseglass, F. Shi, X. Yang, R. Rubinstein, Y. Jin, Chelicerata sDscam isoforms combine homophilic specificities to define unique cell recognition. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 24813–24824 (2020).
35. H. Hammad, M. Chieppa, F. Perros, M. A. Willart, R. N. Germain, B. N. Lambrecht, House dust mite allergen induces asthma via Toll-like receptor 4 triggering of airway structural cells. *Nat. Med.* **15**, 410–416 (2009).
36. D. H. Paris, T. R. Shelite, N. P. Day, D. H. Walker, Unresolved problems related to scrub typhus: A seriously neglected life-threatening disease. *Am. J. Trop. Med. Hyg.* **89**, 301–307 (2013).
37. W. Dermauw, N. Wybouw, S. Rombauts, B. Menten, J. Vontas, M. Grbic, R. M. Clark, R. Feyereisen, T. Van Leeuwen, A link between host plant adaptation and pesticide resistance in the polyphagous spider mite Tetranychus urticae. *Proc. Natl. Acad. Sci. U.S.A.* **110**, E113–E122 (2013).
38. N. Bensoussan, M. E. Santamaria, V. Zhurov, I. Diaz, M. Grbic, V. Grbic, Plant-herbivore interaction: Dissection of the cellular pattern of tetranychus urticae feeding on the host plant. *Front. Plant Sci.* **7**, 1105 (2016).
39. T. Van Leeuwen, J. Vontas, A. Tsagkarakou, W. Dermauw, L. Tirry, Acaricide resistance mechanisms in the two-spotted spider mite Tetranychus urticae and other important Acari: A review. *Insect Biochem. Mol. Biol.* **40**, 563–572 (2010).
40. M. Grbic, T. Van Leeuwen, R. M. Clark, S. Rombauts, P. Rouze, V. Grbic, E. J. Osborne, W. Dermauw, P. C. Ngoc, F. Ortego, P. H.-Crespo, I. Diaz, M. Martinez, M. Navajas, E. Sucena, S. Magalhaes, L. Nagy, R. M. Pace, S. Djuranovic, G. Smagghe, M. Iga, O. Christiaens, J. A. Veenstra, J. Ewer, R. M. Villalobos, J. L. Hutter, S. D. Hudson, M. Velez, S. V. Yi, J. Zeng, A. Pires-daSilva, F. Roch, M. Cazaux, M. Navarro, V. Zhurov, G. Acevedo, A. Bjelica, J. A. Fawcett, E. Bonnet, C. Martens, G. Baele, L. Wissler, A. Sanchez-Rodriguez, L. Tirry, C. Blais, K. Demeestere, S. R. Henz, T. R. Gregory, J. Mathieu, L. Verdon, L. Farinelli, J. Schmutz, E. Lindquist, R. Feyereisen, Y. Van de Peer, The genome of Tetranychus urticae reveals herbivorous pest adaptations. *Nature* **479**, 487–492 (2011).
41. C. J. McManus, M. O. Duff, J. Eipper-Mains, B. R. Graveley, Global analysis of trans-splicing in Drosophila. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 12975–12979 (2010).
42. B. R. Graveley, Mutually exclusive splicing of the insect Dscam pre-mRNA directed by competing intronic RNA secondary structures. *Cell* **123**, 65–73 (2005).
43. Y. Yang, L. Zhan, W. Zhang, F. Sun, W. Wang, N. Tian, J. Bi, H. Wang, D. Shi, Y. Jiang, Y. Zhang, Y. Jin, RNA secondary structure in mutually exclusive splicing. *Nat. Struct. Mol. Biol.* **18**, 159–168 (2011).
44. Y. Yue, Y. Yang, L. Dai, G. Cao, R. Chen, W. Hong, B. Liu, Y. Shi, Y. Meng, F. Shi, M. Xiao, Y. Jin, Long-range RNA pairings contribute to mutually exclusive splicing. *RNA* **22**, 96–110 (2016).
45. S. E. Fischer, M. D. Butler, Q. Pan, G. Ruvkun, Trans-splicing in C. elegans generates the negative RNAi regulator ERI-6/7. *Nature* **455**, 491–496 (2008).
46. R. Kamikawa, Y. Inagaki, M. Tokoro, A. J. Roger, T. Hashimoto, Split introns in the genome of Giardia intestinalis are excised by spliceosome-mediated trans-splicing. *Curr. Biol.* **21**, 311–315 (2011).
47. J. L. Gao, Y. J. Fan, X. Y. Wang, Y. Zhang, J. Pu, L. Li, W. Shao, S. Zhan, J. Hao, Y. Z. Xu, A conserved intronic U1 snRNP-binding sequence promotes trans-splicing in Drosophila. *Genes Dev.* **29**, 760–771 (2015).
48. T. Horiuchi, E. Giniger, T. Aigaki, Alternative trans-splicing of constant and variable exons of a Drosophila axon guidance gene, lola. *Genes Dev.* **17**, 2496–2501 (2003).
49. M. Gabler, M. Volkmar, S. Weinlich, A. Herbst, P. Dobberthien, S. Sklarss, L. Fanti, S. Pimpinelli, H. Kress, G. Reuter, R. Dorn, Trans-splicing of the mod(mdg4) complex locus is conserved between the distantly related species Drosophila melanogaster and D. virilis. *Genetics* **169**, 723–736 (2005).

50. Y. Wang, Q. Zou, F. Li, W. Zhao, H. Xu, W. Zhang, H. Deng, X. Yang, Identification of the cross-strand chimeric RNAs generated by fusions of bi-directional transcripts. *Nat. Commun.* **12**, 4645 (2021).
51. E. L. Lasda, T. Blumenthal, Trans-splicing. *Wiley Interdiscip. Rev. RNA* **2**, 417–434 (2011).
52. W. Shao, Q. Y. Zhao, X. Y. Wang, X. Y. Xu, Q. Tang, M. Li, X. Li, Y. Z. Xu, Alternative splicing and trans-splicing events revealed by analysis of the Bombyx mori transcriptome. *RNA* **18**, 1395–1407 (2012).
53. Y. Jin, H. Dong, Y. Shi, L. Bian, Mutually exclusive alternative splicing of pre-mRNAs. *Wiley Interdiscip. Rev. RNA* **9**, e1468 (2018).
54. W. Hong, J. Zhang, H. Dong, Y. Shi, H. Ma, F. Zhou, B. Xu, Y. Fu, S. Zhang, S. Hou, G. Li, Y. Wu, S. Chen, X. Zhu, W. You, F. Shi, X. Yang, Z. Gong, J. Huang, Y. Jin, Intron-targeted mutagenesis reveals roles for Dscam1 RNA pairing architecture-driven splicing bias in neuronal wiring. *Cell Rep.* **36**, 109373 (2021).
55. M. Suyama, Mechanistic insights into mutually exclusive splicing in dynamin 1. *Bioinformatics* **29**, 2084–2087 (2013).
56. K. Hatje, R. U. Rahman, R. O. Vidal, D. Simm, B. Hammesfahr, V. Bansal, A. Rajput, M. E. Mickael, T. Sun, S. Bonn, M. Kollmar, The landscape of human mutually exclusive splicing. *Mol. Syst. Biol.* **13**, 959 (2017).
57. Y. Yue, S. Hou, X. Wang, L. Zhan, G. Cao, G. Li, Y. Shi, P. Zhang, W. Hong, H. Lin, B. Liu, F. Shi, Y. Yang, Y. Jin, Role and convergent evolution of competing RNA secondary structures in mutually exclusive splicing. *RNA Biol.* **14**, 1399–1410 (2017).
58. H. Pan, Y. Shi, S. Chen, Y. Yang, Y. Yue, L. Zhan, L. Dai, H. Dong, W. Hong, F. Shi, Y. Jin, Competing RNA pairings in complex alternative splicing of a 3′ variable region. *RNA* **24**, 1466–1480 (2018).
59. K. M. Goodman, P. S. Katsamba, R. Rubinstein, G. Ahlsén, F. Bahna, S. Mannepalli, *H. Dan*, R. Sampogna, L. Shapiro, B. Honig, How clustered protocadherin binding specificity is tuned for neuronal self/non-self-recognition. bioRxiv 2021.07.22.453400 [Preprint]. 18 December 2021. https://doi.org/10.1101/2021.07.22.453400.
60. J. Stavnezer, J. E. Guikema, C. E. Schrader, Mechanism and regulation of class switch recombination. *Annu. Rev. Immunol.* **26**, 261–292 (2008).
61. Y. Feng, N. Seija, J. M. Di Noia, A. Martin, AID in antibody diversification: There and back again: (Trends in Immunology 41, 586-600; 2020). *Trends Immunol.* **42**, 89 (2021).
62. L. Shi, H. H. Yu, J. S. Yang, T. Lee, Specific Drosophila Dscam juxtamembrane variants control dendritic elaboration and axonal arborization. *J. Neurosci.* **27**, 6723–6728 (2007).
63. Z. Yang, S. U. Huh, J. M. Drennan, H. Kathuria, J. S. Martinez, H. Tsuda, M. C. Hall, J. C. Clemens, Drosophila Vap-33 is required for axonal localization of Dscam isoforms. *J. Neurosci.* **32**, 17241–17250 (2012).
64. J. Wang, X. Ma, J. S. Yang, X. Zheng, C. T. Zugates, C. H. Lee, T. Lee, Transmembrane/juxtamembrane domain-dependent Dscam distribution and function during mushroom body neuronal morphogenesis. *Neuron* **43**, 663–672 (2004).
65. H. Liu, S. Pizzano, R. Li, W. Zhao, M. W. Veling, Y. Hu, L. Yang, B. Ye, isoTarget: A genetic method for analyzing the functional diversity of splicing isoforms in vivo. *Cell Rep.* **33**, 108361 (2020).
66. D. Dascenco, M. L. Erfurth, A. Izadifar, M. Song, S. Sachse, R. Bortnick, O. Urwyler, M. Petrovic, D. Ayaz, H. He, Y. Kise, F. Thomas, T. Kidd, D. Schmucker, Slit and receptor tyrosine phosphatase 69D confer spatial specificity to axon branching via Dscam1. *Cell* **162**, 1140–1154 (2015).
67. Z. Zhang, K. So, R. Peterson, M. Bauer, H. Ng, Y. Zhang, J. H. Kim, T. Kidd, P. Miura, Elav-mediated exon skipping and alternative polyadenylation of the dscam1 gene are required for axon outgrowth. *Cell Rep.* **27**, 3808–3817.e7 (2019).
68. A. B. Keeler, D. Schreiner, J. A. Weiner, Protein kinase C phosphorylation of a γ-Protocadherin C-terminal lipid binding domain regulates focal adhesion kinase inhibition and dendrite arborization. *J. Biol. Chem.* **290**, 20674–20686 (2015).
69. A. M. Garrett, D. Schreiner, M. A. Lobas, J. A. Weiner, γ-protocadherins control cortical dendrite arborization by regulating the activity of a FAK/PKC/MARCKS signaling pathway. *Neuron* **74**, 269–276 (2012).
70. R. O'Leary, J. E. Reilly, H. H. Hanson, S. Kang, N. Lou, G. R. Phillips, A variable cytoplasmic domain segment is necessary for γ-protocadherin trafficking and tubulation in the endosome/lysosome pathway. *Mol. Biol. Cell* **22**, 4362–4372 (2011).
71. A. Shonubi, C. Roman, G. R. Phillips, The clustered protocadherin endolysosomal trafficking motif mediates cytoplasmic association. *BMC Cell Biol.* **16**, 28 (2015).
72. G. R. Phillips, N. LaMassa, Y. M. Nie, Clustered protocadherin trafficking. *Semin. Cell Dev. Biol.* **69**, 131–139 (2017).
73. Y. Guo, Q. Xu, D. Canzio, J. Shou, J. Li, D. U. Gorkin, I. Jung, H. Wu, Y. Zhai, Y. Tang, Y. Lu, Y. Wu, Z. Jia, W. Li, M. Q. Zhang, B. Ren, A. R. Krainer, T. Maniatis, Q. Wu, CRISPR inversion of CTCF sites alters genome topology and enhancer/promoter function. *Cell* **162**, 900–910 (2015).
74. W. V. Chen, C. L. Nwakeze, C. A. Denny, S. O'Keeffe, M. A. Rieger, G. Mountoufaris, A. Kirner, J. D. Dougherty, R. Hen, Q. Wu, T. Maniatis, Pcdhαc2 is required for axonal tiling and assembly of serotonergic circuitries in mice. *Science* **356**, 406–411 (2017).
75. F. Yang, X. Y. Wang, Z. M. Zhang, J. Pu, Y. J. Fan, J. Zhou, C. C. Query, Y. Z. Xu, Splicing proofreading at 5′ splice sites by ATPase Prp28p. *Nucleic Acids Res.* **41**, 4660–4670 (2013).
76. G. C. Zondag, G. M. Koningstein, Y. P. Jiang, J. Sap, W. H. Moolenaar, M. F. Gebbink, Homophilic interactions mediated by receptor tyrosine phosphatases mu and kappa. A critical role for the novel extracellular MAM domain. *J. Biol. Chem.* **270**, 14247–14250 (1995).
77. C. Chen, H. Chen, Y. Zhang, H. R. Thomas, M. H. Frank, Y. He, R. Xia, TBtools: An integrative toolkit developed for interactive analyses of big biological data. *Mol. Plant* **13**, 1194–1202 (2020).
78. A. J. Bisogni, S. Ghazanfar, E. O. Williams, H. M. Marsh, J. Y. Yang, D. M. Lin, Tuning of delta-protocadherin adhesion through combinatorial diversity. *eLife* **7**, e41050 (2018).

Acknowledgments: We thank members of the Jin laboratory for suggestions and discussion during the course of this work. We thank X. Hong for *T. urticae*; J. Chen for Sf9 cells and pFastBac-Dual vector; Y. Xu for the modified pMT/V5-His B vector; N. Zhou for pEGFP N1 vector; and X. Wu for pFastBacHTB vector. **Funding:** This work was supported by research grants from the National Key Research and Development Program of China (2021YFE0114900), the National Natural Science Foundation of China (91940303, 91740104, and 31630089), the Natural Science Foundation of Zhejiang Province (LD21C050002), and the Starry Night Science Fund at Shanghai Institute for Advanced Study of Zhejiang University (SN-ZJU-SIAS-009).
Author contributions: Y.J. conceived this project. S.H., G.L., and B.X. designed and performed the experiments. S.H., G.L., B.X., and H.D. performed annotation and identification of *Dscam* genes. S.H. analyzed phylogeny of *T. urticae sDscam*β and heatmap of *sDscam*β variable region. Y.M., S.H., and G.L. analyzed the RNA-seq data. S.H., G.L., B.X., J.F., and H.D. performed RNA secondary structure prediction and evolutionary conserved elements analysis. S.H., G.L., S.Z., Y.F., J.S., and L.L. carried out vector construction. S.H., and G.L. performed the quantitative analysis of mRNA splice isoforms. S.H., H.D., Y.F., S.Z., L.L., J.S., and J.F. were responsible for cell aggregation assay and quantification. S.H. and G.L. carried out immunofluorescence assay. S.H., G.L., B.X., H.D., S.Z., Y.F., J.S., and L.L. performed the binding specificity assay and calculation of coaggregation index. B.X., S.H., and G.L. predicted the homology modeling and protein-protein docking. G.L. and S.H. conducted coimmunoprecipitation and Western blot. S.H., G.L., B.X., H.D., F.S., Y.M., and J.F. analyzed the data. Y.J., S.H., G.L., and B.X. wrote the manuscript. All authors discussed the results and commented on the manuscript. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials.

Submitted 3 January 2022
Accepted 21 May 2022
Published 6 July 2022
10.1126/sciadv.abn9458