An RNA architectural locus control region involved in *Dscam* mutually exclusive splicing

Xuebin Wang[1,*], Guoli Li[1,*], Yun Yang[1], Wenfeng Wang[1], Wenjing Zhang[1], Huawei Pan[1], Peng Zhang[1], Yuan Yue[1], Hao Lin[1], Baoping Liu[1], Jingpei Bi[1], Feng Shi[1], Jinping Mao[1], Yijun Meng[1], Leilei Zhan[1] & Yongfeng Jin[1]

The most striking example of alternative splicing in a *Drosophila melanogaster* gene is observed in the *Down syndrome cell adhesion molecule*, which can generate 38,016 different isoforms. RNA secondary structures are thought to direct the mutually exclusive splicing of *Down syndrome cell adhesion molecule*, but the underlying mechanisms are poorly understood. Here we describe a locus control region that can activate the exon 6 cluster and specifically allow for the selection of only one exon variant in combination with docking site selector sequence interactions. Combining comparative genomic studies of 63 species with mutational analysis reveals that intricate, tandem multi-'subunit' RNA structures within the locus control region activate species-appropriate alternative variants. Importantly, strengthening the weak splice sites of the target exon can remove the locus control region dependence. Our findings not only provide a locus control region-dependent mechanism for mutually exclusive splicing, but also suggest a model for the evolution of increased complexity in a long-range RNA molecular machine.

[1] Institute of Biochemistry, College of Life Sciences, Zhejiang University (Zijingang Campus), Hangzhou, Zhejiang ZJ310058, China. * These authors contributed equally to this work. Correspondence and requests for materials should be addressed to Y.J. (email: jinyf@zju.edu.cn).

1

Alternative splicing is a widespread mechanism used by eukaryotes to expand protein diversity and to regulate gene expression[1–3]. Up to 95% of primary transcripts in humans have been estimated to undergo alternative splicing[4,5]. The most striking example of alternative splicing is in the insect gene *Down syndrome cell adhesion molecule* (*Dscam*), which can generate 38,016 different isoforms through mutually exclusive splicing of the four cassette exon clusters in *D. melanogaster*[6,7]. Mutually exclusive splicing is a strictly regulated form of alternative splicing in which the splicing machinery must choose one of two or more candidate exons to include in each messenger RNA (mRNA) isoform[8]. The most attractive model for mutually exclusive splicing involves competition among RNA secondary structures. This mechanism was initially discovered within the exon 6 cluster of *Dscam*[9,10]. We recently observed similar structural arrangements in several clusters of mutually exclusive exons, including the exon 4 and exon 9 clusters in *Dscam*[11].

Although the docking site-selector sequence interactions have a key role in ensuring mutually exclusive splicing[11,12], the role of these RNA structures remains poorly understood. A second component of this system involves the RNA-binding protein hrp36, which binds the exons throughout the exon 6 cluster and represses their inclusion in the mRNA isoform[13]. Nevertheless, hrp36 has no effect on the inclusion of exon variants from *Dscam* exon clusters 4 and 9 (ref. 13). In the present study, we identified a locus control region (LCR) that could activate the exon 6 cluster. Our findings not only provide an LCR-dependent mechanism for the selection of only one exon splice variant, but also suggest a model for the evolution of increased complexity in a long-range RNA molecular machine.

Results

An LCR essential for *Dscam* exon 6 splicing. Our initial analysis indicated that sequences upstream of the docking site had important roles in the activation of exon 6 through the generation of a deletion construct with exon 6.2 fused to exon 6.46 (Supplementary Fig. S1). To further analyse the elements involved in exon 6 activation, we created a series of constructs designed to mimic the approximation of sequences caused by RNA pairing between the docking site and selector sequences (Fig. 1a). These data revealed that only the most proximal exon outside the loop was activated, whereas the remaining exons were silenced (Fig. 1b,c), implying an approximation-activation mechanism. Importantly, these constructs could be used to identify the elements necessary for exon 6 activation. Consequently, the deletion of the ~200-bp sequences upstream of the docking site substantially decreased or even abrogated the inclusion of the proximal exon (data not shown). However, this sequence was not sufficient for the activation of the exon 6.47 variant by locus insertion, suggesting that other essential motif sequences are needed for efficient activation.

To identify other functional motifs, an ~700-bp region upstream of the docking site was scanned using by performing iterative deletions of 50–70-bp regions, and the effects of the deletions on variant inclusion were quantitatively determined (Fig. 1d). Unexpectedly, all of the deletions significantly decreased the activation of the exon 6.47 variant. Most strikingly, 7 of 11 deletions led to the almost exclusive omission of the 6.47 variant (Fig. 1e,f). This outcome indicated that functional motifs may be scattered throughout the intronic sequences upstream of the docking site. A similar trend was observed in exons 6.48 and 6.43, although the extent of the reduction of exons included in the mRNA isoforms from exons 6.48 and 6.43 differed from those with exon 6.47 (Supplementary Fig. S2). This striking conservation in the reduction pattern for the various activation constructs suggested a common activation pattern. Importantly, the alternative exon 6.47 could be activated by locus replacement of this region into the selector sequence IE47. Similarly, exon 6.48 was frequently activated when the selector sequence IE48 was replaced (Fig. 1g,h). Therefore, the large region upstream of the docking site could act as an activation element for efficient inclusion in exon 6. As this type of splicing of a cis-regulatory element can specifically activate the target exons through long-distance interactions between docking site-selector sequences, we refer to it as an LCR by analogy with the well-characterised LCR in transcriptional regulation[14–18].

RNA architecture and secondary structure of *Drosophila* LCR. The significant decrease observed in intron-wide scanning deletions led us to hypothesize that this intronic region may form a long-range structure. Moreover, combining comparative genomics with structural predictions revealed that this intronic sequence could potentially form a hexaleaf-shaped architecture consisting of six adjacent tandem stem–loop structures (I through VI, Fig. 2). Despite the relatively low sequence conservation, the predicted RNA architecture and secondary structures are highly conserved in the 22 *Drosophila* species analysed (Fig. 2 and Supplementary Fig. S3). Additionally, this double stranded RNA (dsRNA) region showed clear evidence of multiple covariations that maintain the structural integrity of the dsRNA (Supplementary Fig. S4). In each species, the 5′ portion of the first dsRNA (I) was located an average of 54 nt downstream of the 5′ splice site of exon 5, and the last dsRNA (VI) was located immediately upstream of the docking site (Fig. 2a,b). Importantly, as these six stem–loop structures were arranged in tandem, the 'effective' distance between the 5′ and 3′ ends was approximated to be <50 nt.

To determine whether these RNA structures are essential for exon 6 splicing, we tested the effects of disrupting and compensatory mutations on splicing in transfection experiments. Mutations M1 and M2 in stem I almost completely abolished the inclusion of exon 6.47 in the mRNA isoform (Fig. 2c). A structure-restoring double mutation (M21) restored the efficiency of exon 6.47 inclusion to the wild-type (WT) level, thereby validating predicted base–pair interactions. Likewise, the other five RNA stems (II through VI) were confirmed by disruptive and compensatory mutation analysis (Fig. 2b,c). Thus, the data obtained by disruptive and compensatory mutation analysis strongly suggest that these RNA structures are essential for efficient activation activity.

LCR architecture is evolutionarily conserved. Next, we explored whether the intricate architecture found in the *Drosophila* LCR is conserved throughout insect or arthropod evolution. We expanded this analysis to other arthropod species from seven orders (http://flybase.org/blast/). Together, these organisms encompass several major taxonomic groups of insects and crustaceans that last shared a common ancestor ~420 million years ago[19]. Previous studies[9,10,12,20], together with our current analysis, revealed that mutually exclusive splicing employed the docking/selector strategy in all of the species investigated. Remarkably, the sequence comparison revealed eight conserved intronic elements upstream of the docking site in exon cluster 6 among 15 hymenopteran species; the estimated divergence times ranged from 10 million to 150 million years ago (Fig. 3a). These Hymenoptera-specific intronic sequences form an architecture similar to that observed in *Drosophila*, albeit with four-dsRNA tandem arrays (Fig. 3b and Supplementary Fig. S5). Structure-restoring double mutations in four-dsRNA stems restored the
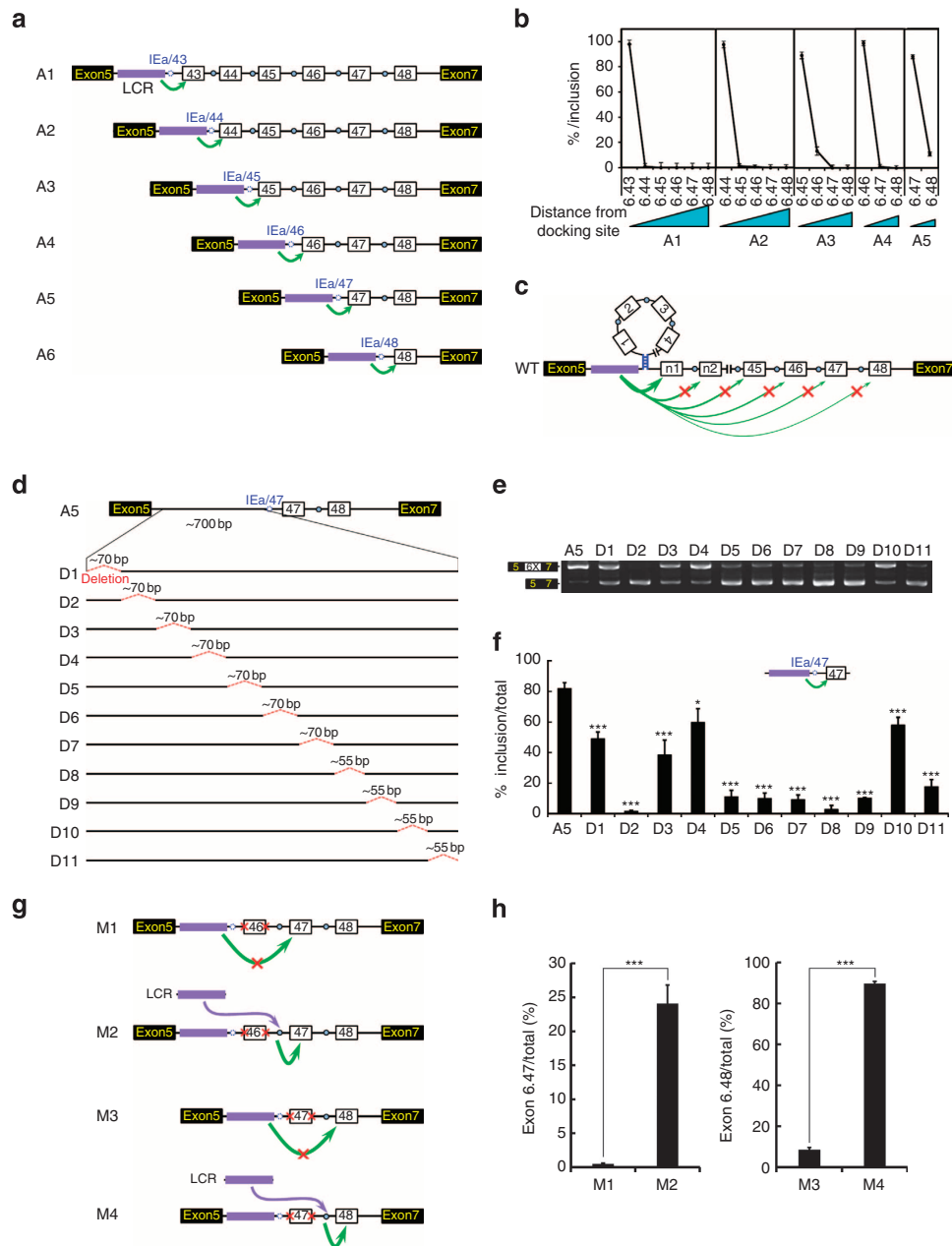
Figure 1 | A cis-acting LCR is essential for the efficient activation of *Dscam* exon 6 in *D. melanogaster*. (**a**) Schematic diagrams of a series of constructs that were designed to mimic the approximation of sequences caused by RNA pairing between the docking site and selector sequence. The dashed circle depicts the 'hardwired' site after deleting the loop–stem sequence. (**b**) Effects of deletions on the selection of alternative exons. (**c**) LCR elements act in a proximity-dependent mode. (**d**) Overview of scanning deletions of the ∼700-bp intronic sequence upstream of the docking site of *Dscam* exon cluster 6 in *D. melanogaster*. (**e,f**) Effect of deletion on exon 6 inclusion. (**g**) Overview of locus insertion to test the effects of the LCR on the inclusion of exon 6.47 and exon 6.48 is shown. A red cross in IE denotes the nucleotide mutations, which destroyed the splice sites of alternative exon 6; a green arrow depicts activating the inclusion of the alternative exon. (**h**) The alternative exon is activated by a locus-inserted LCR. Data are expressed as mean ± s.d. from three independent experiments (**b,f,h**). *$P < 0.05$; ***$P < 0.001$ (Student's *t*-test, two-tail). WT, wild-type.

efficiency of exon 6 inclusion to a degree similar to that observed with the WT minigene (Fig. 3b,c), thus validating the presence of the predicted four RNA stems. Importantly, clear evidence of compensatory structural evolution and evolutionary intermediates exist within each core region in all four dsRNAs (Fig. 3a and Supplementary Fig. S6).

Moreover, the architectures and secondary structures of the LCRs from the coleopteran and other species behaved similarly (Supplementary Figs. S7,8), although the LCR sequences were highly divergent. Thus, we concluded that these RNA architectures and secondary structures are an evolutionarily conserved component required for LCRs throughout insect and arthropod evolution. Notably, only one stem–loop structure was predicted at the corresponding position (Fig. 3d), whereas there are 26 variable exons in the exon 6 cluster of the waterflea (*Daphnia pulex* and *D. magna*) species[20], suggesting that one dsRNA structure might be sufficient for exon 6 activation in the Pancrustacea ancestor gene. Disruptive and compensatory mutation analysis strongly demonstrated that this RNA structure was essential for activation activity (Fig. 3d,e).
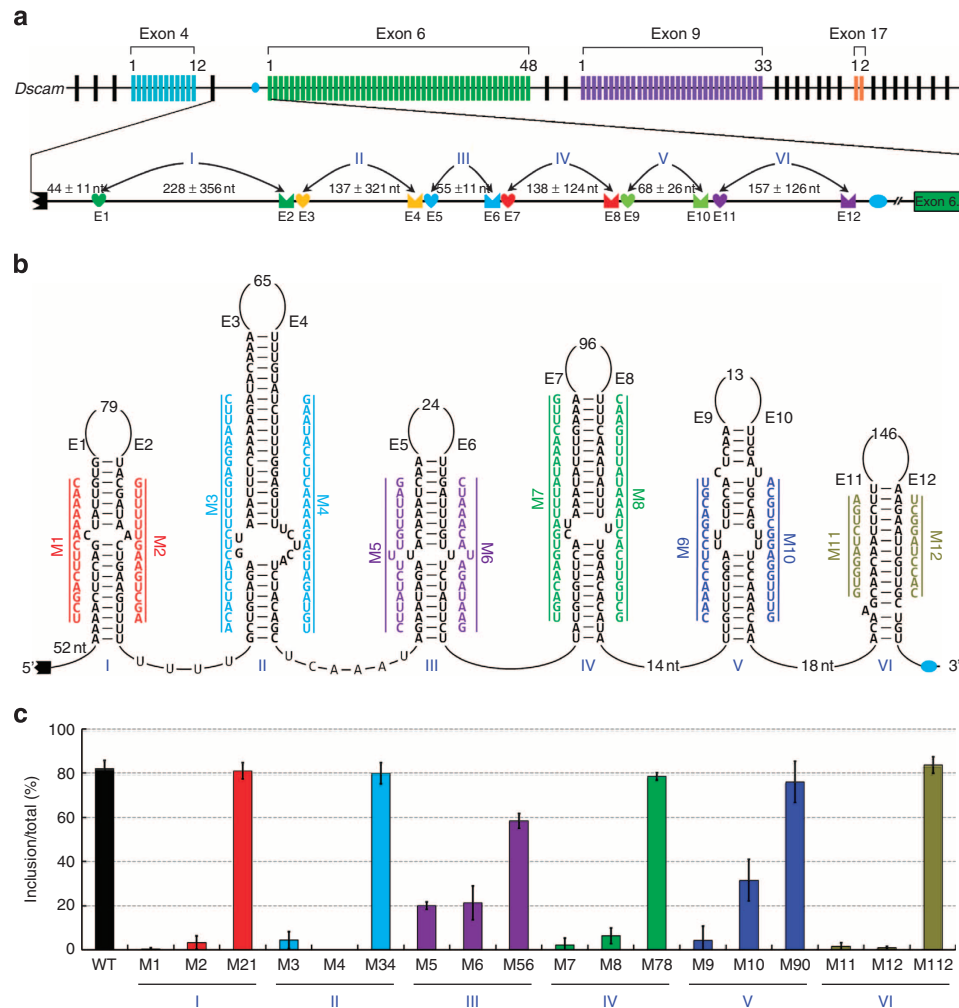
Figure 2 | RNA architecture and secondary structure of LCR in *Drosophila Dscam*. (a) Genomic organization of the *D. melanogaster Dscam* gene. Constitutive exons (in black boxes), alternative exons (in coloured boxes), docking sites (in blue oval) and introns (lines) are shown. Intronic elements (marked by saddle shapes) were reverse-complementary to upstream elements (marked by hearts) with stem–loop sizes as indicated (mean ± s.d.). **(b)** The predicted hexamer architecture of *Dscam* pre-mRNA. Mutations introduced into six individual dsRNAs (I–VI) are indicated on the left or right mutated sequences (M1–M12). **(c)** Effects of RNA architecture and secondary structures were validated by disruptive mutations (M1–M12) and compensatory double mutations (M21, M34, M56, M78, M90 and M112). WT, wild-type. Data are expressed as mean ± s.d. from three independent experiments.

Enhancer 'subunit' within the LCR. Next, we explored why RNA molecules such as LCRs fold into elegant and intricate shapes. First, we deleted the individual stem–loops and examined how each one affected the activation efficiency. Exon 6.47 inclusion was significantly decreased to 10–70% of the WT levels by individual deletions of structures I–VI (Fig. 4a,b), indicating that the I–VI region was essential for efficient activation of exon 6.47. Moreover, deletion mutations affected exon 6 inclusion less dramatically than disruptive mutations of individual stem–loops (Fig. 4a). The latter may disrupt the LCR architecture, which coincides with a change in their predicted structures. The pronounced discrepancies between disruptive and deletion mutations suggested that each stem–loop acted as an enhancer 'subunit'.

The deletion analysis of the LCR on exon 6.48 inclusion further supported this hypothesis. Individual deletion of the I–VI 'subunits' did not significantly decrease the inclusion of exon 6.48 if either 'subunit' was maintained (Fig. 4a,b). These data indicated that either of the I–VI subunits was sufficient to activate the alternative exon 6.48 in *D. melanogaster* similar to the single-subunit LCR in daphnian species. However, the disruptive mutations in the dsRNA significantly decreased exon 6.48

inclusion into the mRNA isoform (Fig. 4a). Similar trends have been observed in the exon 6.43 activation constructs. The striking conservation in the discrepancies of effects on the inclusion of different exon 6 variants between the disrupted and deletion mutations suggests that each stem–loop acts as an enhancer 'subunit'.

LCR activity is correlated with 'subunit' number. Next, we examined whether this elaborate structural LCR occurred in a highly regulated or a stochastic manner. By comparing the inclusion frequency in various deletion mutations of 'subunits', we examined how the architectural complexity of the LCR contributed to its activity. First, a series of deletion mutants was constructed to generate enhancer mutants, which contained different tandem 'subunits' (Fig. 4b). Consequently, the ability of the LCR to activate exon 6.43 diminished with the decreasing number of 'subunits' and was completely inhibited in ΔI–VI constructs; however, individual 'subunits' contributed unequally to activation activity (Fig. 4b). Similar trends have been observed in exon 6.47 activation constructs, which suggests that activation activity is
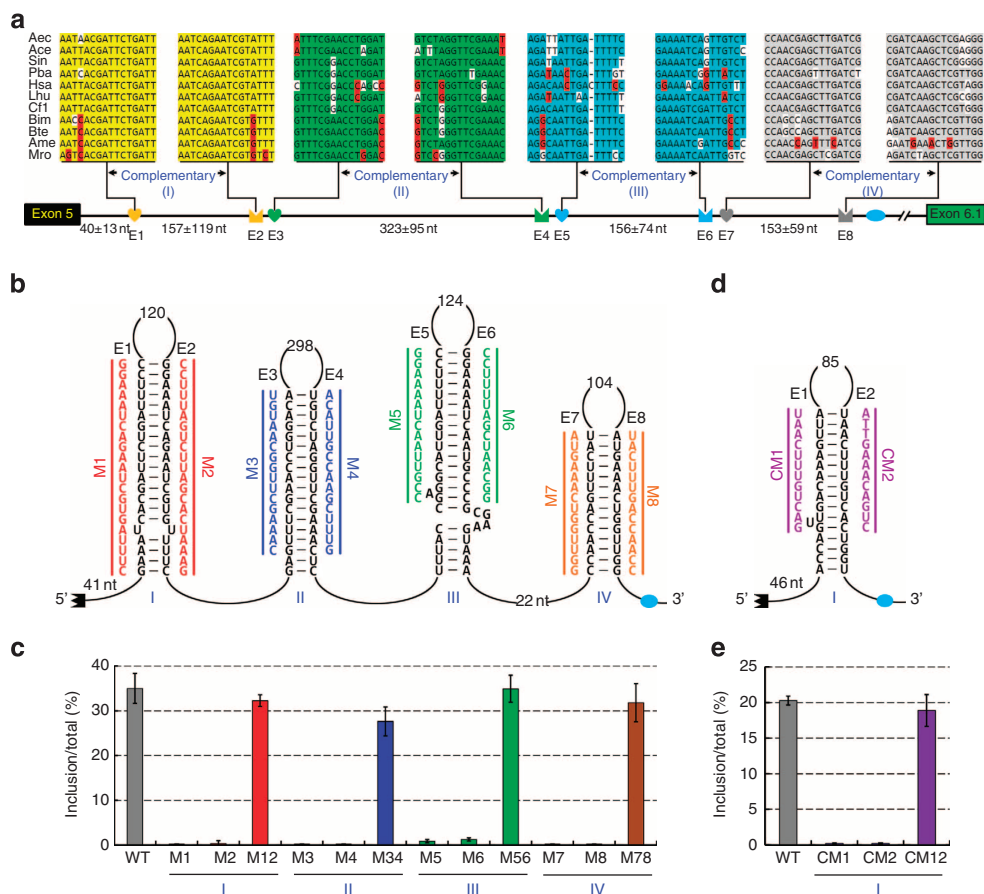
Figure 3 | The LCR architecture in _Hymenopteran_ and _Daphnian Dscam_. (**a**) A schematic diagram of the partial pre-mRNA in the _Hymenopteran_ species. Constitutive exons are depicted as black boxes, alternative exons as green boxes and introns as lines. Above are sequences of consensus intronic elements for different species taxa (see Supplementary Table S1 for abbreviations). The most identical nucleotides at each position are shown in different colours. Notably, most of these differences are structurally silent, whereas nucleotide covariations that maintain the structural integrity of the dsRNA were shaded in red. (**b**) The predicted tetramer architecture and secondary structures of the _A. mellifera_ pre-mRNA. Mutations were introduced into four dsRNAs (M1–M8). (**c**) Effects of RNA secondary structures were validated by disruptive mutations (M1–M8) and compensatory double mutations (M12, M34, M56 and M78). WT, wild-type. (**d**) The monomer architecture and secondary structure of _D. magna Dscam_ pre-mRNA. Mutations were introduced into dsRNA (CM1, CM2). (**e**) Effects of RNA secondary structure were validated by disruptive mutations (CM1, CM2) and compensatory double mutations (CM12). WT, wild-type. Data are expressed as mean ± s.d. from three independent experiments.

positively correlated with the number of 'subunits'. The deletion of more 'subunits' likely reduced the number of LCR functional motifs to cause a substantial reduction of its activity. Furthermore, these results support the correlation of the LCR architectural complexity and the number of alternative exon 6 variants in _Drosophila_, _Hymenoptera_ and the waterflea (Fig. 4c). Together, these findings suggest that the LCR architectural complexity contributed to the expansion of the exon 6 cluster during speciation.

Newly evolved subunits have a minimal contribution to LCR. Next, we examined the extent to which differences in individual I–VI subunits within LCRs contribute to exon 6 activation. When these constructs lacked either I–VI subunits, the inclusion of exon 6.47 was markedly decreased to 10–70% of the full construct's activity (Fig. 5a). The greatest decreases in exon inclusion (~10% of WT levels) were observed in the ΔVI constructs. The smallest decreases (~70% of WT levels) were observed in the ΔIII or ΔV constructs; interestingly, the III and V subunits were relatively small and poorly conserved (Fig. 5a). A detailed analysis revealed that deletion of the more conserved and relatively larger (older)

subunits had a greater effect on the activation activity than deletion of the newly evolved subunits. It appears that the size of the stem–loop reflects the evolutionary conservation of the RNA secondary structures. Furthermore, these results coincide with the correlation of the size of the direct activation efficiency of a single 'subunit'. When we replaced the LCR with individual I–VI subunits, exon 6.43 was activated in a different way (Fig. 5b). Statistical analysis indicated that the activation efficiency was positively correlated with the size of a single 'subunit' (Fig. 5b). Similar trends have been observed in exon 6.48 activation constructs (Fig. 5c). In this case, the newly evolved local dsRNA might act as an evolutionary intermediate. Collectively, these results suggest that conserved subunits, rather than the more recently evolved ones, increase activity to activate exon 6 more efficiently.

LCRs act in a taxon-specific manner. To determine whether LCRs were species-specific and how they evolved during insect evolution, the LCR in _D. melanogaster_ was replaced with that of _D. yakuba_, _D. elegans_, _D. ananassae_, _D. virilis_, _Bombyx. mori_ or _Apis mellifera_. These species represent different phylogenetic
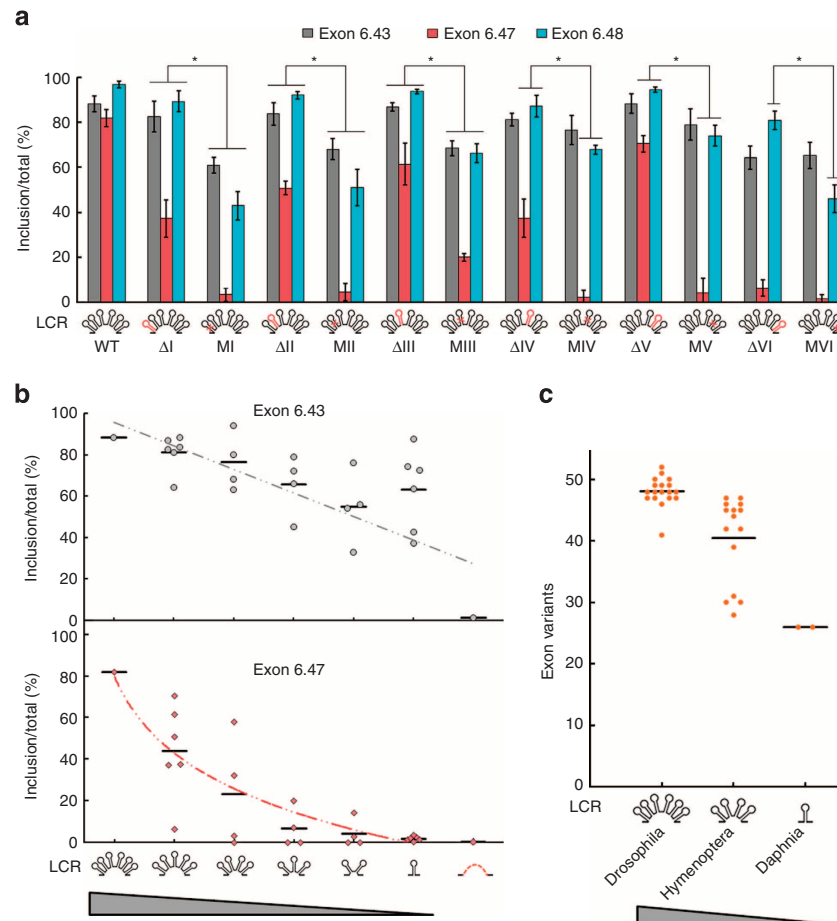
Figure 4 | The LCR 'subunits'. (a) The comparison of the effect on splicing between the disruptive and deletion mutations of dsRNAs. Deletion mutations had a much smaller effect on the inclusion of exon 6.43, exon 6.47 or exon 6.48 than disruptive mutations, suggesting that each dsRNA acts as enhancer 'subunit'. Data are expressed as mean ± s.d. from three independent experiments. * *P* < 0.05 (Student's *t*-test, two-tail). **(b)** Activation efficiency positively correlated with the number of enhancer 'subunits' in a variant-specific manner. **(c)** The LCR architectural complexity correlates with alternative exon 6 number. The number of species in each invertebrate lineage represented in the analysis is shown. WT, wild-type.

distances from 10 to 300 million years ago. We observed that the inclusion of exon 6.47 was reduced to ∼90% of the wild-type minigene level when the LCR from *D. yakuba* was present (Fig. 5d). In contrast, exon 6.47 was not robustly induced when the LCRs from the less closely related *D. elegans* and *D. ananassae* were present, whereas the *D. virilis* LCR still had some activity (Fig. 5d). This result suggested that other regulatory sites were essential for exon 6.47 activation, even though the LCR shape was conserved across *Drosophila*. Exon 6.47 was exclusively skipped when the LCRs from the distantly related *A. mellifera* were present. Conversely, the fly LCR could not efficiently activate the *A. mellifera* exons, indicating that the effect of heterologous LCRs on exon 6.47 inclusion decreased with increasing evolutionary distance from *D. melanogaster*. Inconsistent with exon 6.47, exon 6.48 was efficiently activated using LCRs from various *Drosophila* species, while it was markedly reduced when the *A. mellifera* LCRs were present (Fig. 5d). This result indicated that the LCR activated exon 6.48 in a *Drosophila*-specific manner. These data also indicated that the LCR activated the exon 6 variants in a taxon- or species-specific manner and suggested that adaptable coevolution occurred between the LCR and the activated exon.

Location effects of the LCR. As described above, deletion of the LCR markedly reduced or even abolished the inclusion of exon 6 variants in *cis* (Fig. 4b). When the selector sequence IE47 was

replaced by the LCR, exon 6.47 was partially activated, whereas the inclusion of exon 6.47 was exclusively inhibited in the anti-sense control construct (Supplementary Fig. S9a–c). Likewise, exon 6.48 was often activated when the selector sequence IE48 was replaced by the LCR (Supplementary Fig. S9d–f). However, comparison with the minigene constructs showed that the efficiency was greatly decreased by the use of a locus-inserted LCR. For example, exon 6.47 inclusion decreased to 20% of the WT minigene levels, whereas exon 6.48 inclusion decreased to 85% of the WT minigene levels. As the location of the LCR was constrained, the LCR activity was likely dependent on its location. To examine this hypothesis, a series of mutants were constructed to generate LCRs in different locations (Supplementary Fig. S9a,d). Whenever the distance between the 5′ splice site and the LCR, between the LCR and the 3′ splice site, or between LCRs increased, the efficiency greatly decreased (Supplementary Fig. S9d,f). Moreover, abundant reverse transcription–PCR product containing exon 6.47 and exon 6.48 spliced together was detected when the LCR was inserted downstream of the alternative exon 6.47 (M6, Supplementary Fig. S9e). Together, these results indicated that the LCR acted in *cis* depending on its location.

The LCR is required for Δhrp36 exon variants. We were particularly interested in how the LCR specifically and efficiently activated the most proximal exon. We envisioned two general
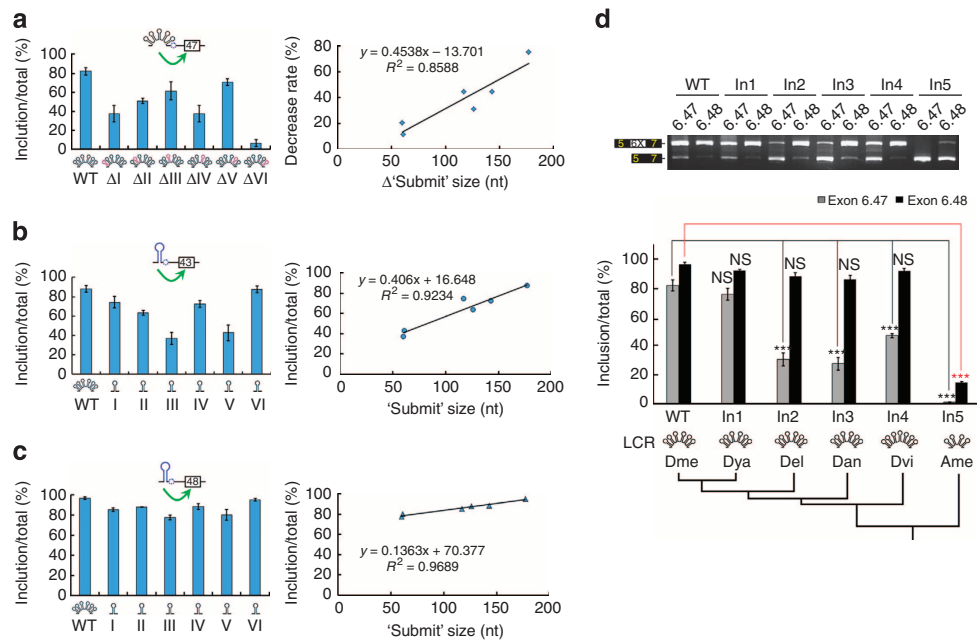
Figure 5 | Conserved subunits more efficiently activate exon 6 than newly evolved subunits. (**a**) The activation efficiency of exon 6.47 has been reduced by deleting a single 'subunit'. (**b,c**) The activation efficiency of a single 'subunit' on the inclusion of exon 6.43 (**b**) and 6.48 (**c**). These data revealed that the size of the 'subunit' correlated with activation activity. (**d**) The LCR acts in a taxon-specific manner. The phylogenetic relationship among *Drosophila* species was used[5]. This finding indicates that the effect of the heterologous LCR on exon 6 inclusion decreases with increasing evolutionary distance to *D. melanogaster*. Data are expressed as mean ± s.d. from three independent experiments. WT, wild-type; NS, nonsignificant. ***$P < 0.001$ (Student's *t*-test, two-tail).

scenarios that could explain this phenomenon. The first scenario is suggested by previous studies that indicated that hrp36 repressed splicing by binding throughout an exon, which prevents serine/arginine-rich proteins from binding to the exons and promotes their inclusion in the mRNA isoform[13]. These data led to the hypothesis that the LCR functions by overcoming the splicing inhibitor. In the second scenario, the LCR facilitates the recognition of the weak splice site. To distinguish between the above scenarios, we first relieved the suppression by mutating the hrp36 sites (Supplementary Fig. S10). We hypothesized that mutating the hrp36 sites reduces the LCR requirement because the inclusion of exon 6 in the mRNA isoform can be activated in the absence of LCRs. We found that exon 6.47 inclusion was not activated in the absence of LCRs when the hrp36 site was mutated within exon 6.47 (Supplementary Fig. S10). Likewise, the inclusion of exon 6.48 was not significantly enhanced in the absence of the LCR in the Δhrp36 construct. This result indicated that the LCR was required for efficient activation of the Δhrp36 exon variant. Although the LCR may have overcome other inhibitors, we believe that its main function is to promote the recognition of the splice site.

Moreover, evolutionary constraints on the distance between the LCR and both of the splice sites further support this concept. The LCR was always located ~50 bp downstream of the 5′ splice site of exon 5, and, importantly, its effective distance was <100 bp upstream of the 3′ splice site of targeted exon 6 only when the docking site was paired with its selector sequence. As the location of the LCR was conserved and because disrupting the LCR only interfered with the most proximal exon 6, the LCR likely functioned by modulating the activity of the splice site.

Strengthening weak splice sites relieves the LCR requirement.
To determine whether strengthening the weak splice sites relieves the LCR requirement, we examined how the LCR contributed to

the activation of the splice sites within exon cluster 6 of *Dscam*. Indeed, some exon variants in the exon 6 cluster of *Dscam* possessed a weak 5′ or 3′ splice site using the splice site predictor[21] (Fig. 6a). Intriguingly, the alternative splice sites contained so-called dual-specificity splice sites (Supplementary Fig. S11), which may 'confuse' the splicing machinery because they can be recognized as either a 5′ or 3′ splice site[22]. Importantly, such arrangements of weak splice sites within an exon cluster were highly conserved in the dipteran *D. melanogaster*, lepidopteran *B. mori*, coleopteran *Tribolium castaneum*, hymenopteran *A. mellifera*, and daphnian *D. pulex* (Supplementary Fig. S12). These observations suggested that the presence of weak or ambiguous splice sites renders the activation of splicing dependent on approximated LCRs for their efficient and correct recognition by the splicing machinery.

We hypothesized that the inclusion of exon 6 in the mRNA isoform could be activated in the absence of LCRs if weak splice sites were strengthened. Consequently, the inclusion of exon 6.47 could be activated in the absence of LCRs when exon 6.47 possessed either a stronger 5′ splice site, a stronger 3′ splice site, or both (Fig. 6b–d). Similar results have been achieved using the other constructs to activate exon 6.48 with a stronger 5′ splice site (Fig. 6e–g). These data indicated that the splicing efficiency of the mutant minigene was partially LCR-independent, implying that the LCR may activate exon 6 inclusion in the mRNA isoform by facilitating the recognition of the splice site.

This result was further supported by interfering with the interaction between the docking site and selector sequence. When a single splice site of exon 6.47 was converted to a strong splice site, exon 6 was partially or largely included even in the absence of IEa-IE47 RNA pairing (Supplementary Fig. S13). However, when both the 5′ and 3′ splice sites of exon 6 were modified to optimized sites, exon 6 was largely included even in the absence of IEa-IE47 RNA pairing (Supplementary Fig. S13b–d). These results confirmed that splicing of the mutant minigene was no
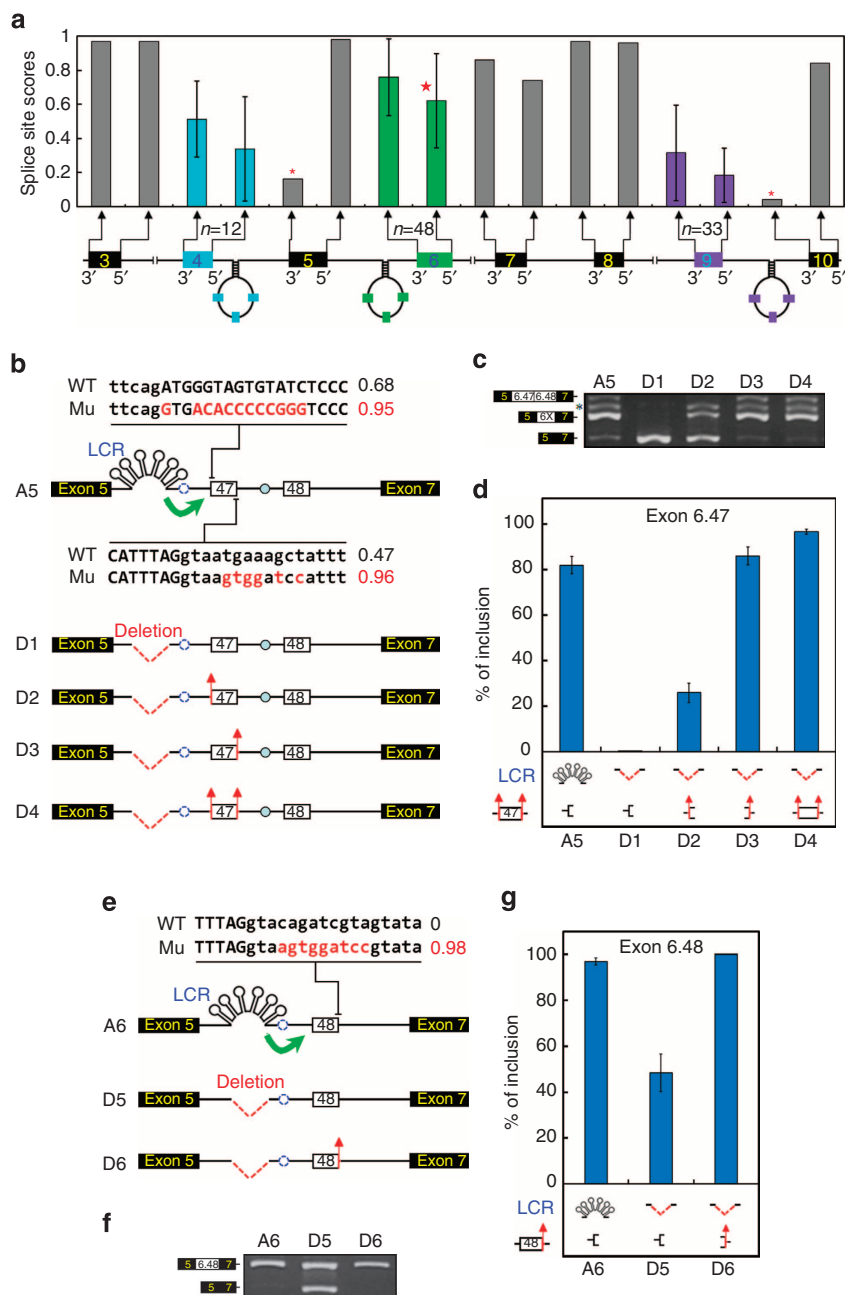
Figure 6 | Strengthening weak splice sites relieves the LCR requirement in *Dscam*. (**a**) Weak or suboptimal splice sites are built on sets of exon clusters in *Dscam*. The 5' and 3' ss motif scores were calculated using the splice site predictor[21] (range is 0 to 1, with higher values predicting stronger splice sites). '*' depicts the weak constitutive splice site downstream the docking site. '*' indicates splice sites that can be recognized as both 5' and 3' ss. (**b**) An overview of various mutant minigene constructs generated for the splicing assay for exon 6.47. A green arrow depicts activating the inclusion of the alternative exon; A red arrow depicts increasing the strength of the splice site. The wild-type and modified 5' or 3' splice site sequences are presented, with the scores on the right. Uppercase letters indicate exon sequences; lowercase letters indicate intron sequences. Mutated nucleotides are marked in red. (**c**) Effects of mutations on the inclusion of the alternative exon 6.47. Reverse transcription–PCR was performed to detect RNA splicing pattern. The band marked by '*' is a nonspecific RT–PCR product. (**d**) Effects of mutations on the inclusion of alternative exon 6.47. (**e**) Overview of various mutant minigene constructs generated for splicing assay exon 6.48. Mutated nucleotides are marked in red. (**f**) Effects of mutations on the inclusion of the alternative exon 6.48. Reverse transcription–PCR was performed to detect RNA splicing pattern. (**g**) Effects of mutations on the inclusion of the alternative exon 6.48. Data are expressed as mean ± s.d. from three independent experiments (**d,g**). WT, wild-type.

longer RNA pairing-dependent. Notably, in this model, exon 6.48 was mainly included even in the absence of IEa-IE48 RNA pairing when the 5' splice site was modified to optimized sites (Supplementary Fig. S13e). On the basis of these findings, we postulated that the LCR cooperated to activate alternative exons by promoting the recognition of the weak splice site.

Discussion

In this study, we identified a splicing LCR which activates the exon 6 cluster and specifically allows for only one selection of exon variants in combination with competing RNA structures between docking site-selector sequences. A recent deletion analysis demonstrated that the docking site and selector
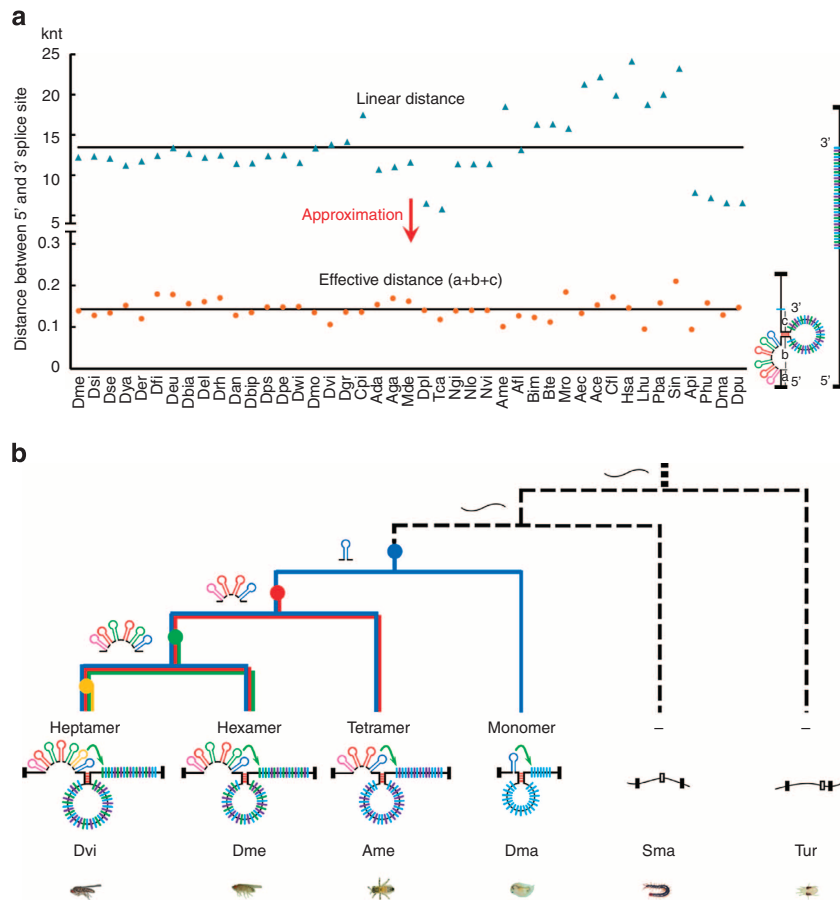
Figure 7 | The evolutionary constraints of the LCR architecture. (a) The architecture of the LCR led to physical distance constraints for only one activation. The last exon 6 variant is shown as an example. Although the distance between the constitutive 5′ ss and alternative 3′ ss is very large and differs among species (5–25 kb), the intricate architecture of the LCR combined with the interaction between the docking and selector sequence sites shorten their effective distance to only ∼200 nt. **(b)** The phylogeny of the architectural LCR in directing *Dscam* mutually exclusive splicing. The extant *Dscam* pre-mRNA structures and proposed ancestor molecules are shown associated with a cladogram of phylogenetic relationships in this study[19,27]. The LCR architecture and secondary structures in *D. virilis* (Dvi) were confirmed in *Drosophila S2* cells (Supplementary Fig. S17). The abbreviations that are used are shown in Supplementary Table S1. The nodes denoting the ancestral origins of particular enhancer 'subunits' are indicated by solid circles. The constitutive *Dscam* mRNAs are depicted as unstructured (black).

sequences were required for mutually exclusive splicing of exon 6 (ref. 12). In these deletions, the mutant phenotypes (that is, Δ*left*) should be mainly caused by the loss of interaction between the docking site and selector sequence because the deleted sequences have little effect on the LCR architecture. Mutational analyses demonstrated that such an intricate architecture has multiple roles in splicing regulation. First, the architecture of the LCR results in physical distance constraints for only one activation. Although this LCR (that is, *D. pulex*) may appear analogous to the iStem identified in the exon 4 cluster based on their locations[23], the iStem affects the inclusion of all 12 exon 4 variants equally. Importantly, when the docking site was paired with its selector sequence, the intricate architecture resulted in an effective distance of <200 bp between the constitutive 5′ splice site and the alternative 3′ splice site of the most proximal exon (Fig. 7a). Previous studies have demonstrated that weak splice site recognition across the intron ceases when the intron size reaches the threshold length of 200 nt (ref. 24). In this scenario, only the most proximal exon fits this activation index, whereas the other alternative exons were beyond this silencing threshold.

Additionally, because both the clean and partial deletion of LCR 'subunits' can significantly decrease the inclusion of exon 6 (Fig. 4), the LCR may have other important context-specific

functions. Moreover, the deletion and mutational analysis indicated that the context of the apical loop sequence contributed to the activation activity (data not shown); however, the destruction of the internal loop within the dsRNA stem had little effect on exon 6 splicing (Fig. 2). Therefore, the LCR requires additional specific motifs, possibly by binding multiple proteins. Owing to the evolutionarily conserved proximity of the RNA secondary structure to exon 6 when the selector sequence interacts with the docking site, it seemed most likely that the LCR modulated the activity of the 5′ and 3′ splice sites of intron 5.

Finally, the LCR acts as a 'transformer' in a regulated manner. The activity of the LCR is neither strong nor weak. Multi-subunits within the LCR RNA could be assembled to achieve combinatorial regulation and higher-order functions, which would guarantee that only the most proximal exon could be activated with 100% efficiency for all 48 context-specific exon 6 variants. Moreover, such multi-subunit LCRs would allow for the compensation of mutationally inactivated sites by intact, neighbouring sites.

To summarize, the present study extends previous models of mutually exclusive splicing mechanisms[9,11,13]. When a selector sequence interacts with the docking site, the LCR specifically activates the proximal alternative exon by promoting the

recognition of the splice site. Alternatively, approximated LCRs could partially contribute to the antagonism of the repressors for the proximal alternative exon. Our findings provide a comprehensive framework to guarantee the one-only choice through an intricate combination of competitive RNA secondary structures and the LCR.

The long-range activation in *Dscam* splicing is controlled by LCRs under intense purifying selection. By integrating the genetic and molecular data from 63 arthropod species, we propose a credible evolutionary model of increased complexity of the LCR for the activation of exon 6 variants (Fig. 7b). We suggest that one intronic element upstream of the docking site may have undergone purifying selection and could form the ancestral monomer structure. Such a structural subunit could act as an LCR to activate the inclusion of the proximal exon when the docking site interacted with a selection sequence. Thus, the simple monomer LCR was formed in the ancestral gene (Fig. 7b). Furthermore, the LCR complexity increased to adapt for the expanded exon cluster, changing from monomeric to tetrameric (*A. mellifera*), hexameric (*D. melanogaster*) and even higher-ordered structures in *D. virilis* (Fig. 7b and Supplementary Fig. S14). Thus, more complicated structures formed to adapt to exon cluster expansion and higher-order regulatory functions.

The extraordinary molecular diversity of Dscam could have biological significance[25]. We also found that the sequence immediately downstream of the docking site could possibly form a conserved dsRNA in *Drosophila* and *Hymenopteran Dscam* exon 4 (Supplementary Fig. S15,16). Analogous to long-range control in transcriptional regulation[14–18], LCR-guided long-range control may represent a novel mechanism of post-transcriptional RNA processing.

Methods
Materials. Insect and other species used in this study are presented in Supplementary Table S1. Fruitflies (*D. melanogaster*), silkworms (*B. mori*), red flour beetles (*T. castaneum*), honeybees (*A. mellifera*) and other insect species were obtained as previously reported[11]. Waterfleas (*D. magna*) were donated by the Institute of Pesticide and Environmental Toxicology, Zhejiang University.

Cloning and sequencing of LCR orthologues. The sequences of the *Dscam* genes from some insect species have been previously described (http://flybase.org/blast/, Supplementary Table S1). The *Dscam* sequences of *D. pulex* and *D. magna* were obtained from JGI genome web (http://genome.jgi-psf.org/, Supplementary Table S1). The interval sequences between exon 5 and the docking site of the *Dscam* genes for the other species were determined by PCR and sequencing. The 3′ primers were designed according to the highly conserved sequence of the docking sites, and 5′ degenerate primers were necessary to amplify products in most species. The primer sequences are listed in Supplementary Table S2.

Sequence alignments and RNA pairing predictions. The alignments of specific regions between species were performed using the ClustalW programme (http://www.ebi.ac.uk/Tools/msa/clustalw2/). The intronic RNA secondary structures were predicted using the Mfold programme[26].

Quantification of mRNA splice isoforms. We assayed the RNA splice isoform ratio using reverse transcription–PCR followed by exon-specific restriction digestion as previously described[11]. The error bars throughout this study were calculated from the average of three independent experiments. Statistical significance was evaluated with an unpaired two-tailed Student's *t*-test. A *P*-value <0.05 was deemed to indicate statistical significance.

Minigene construction, mutagenesis and transfection. Site mutagenesis was performed in both conserved intronic elements to disrupt this secondary structure on the schematic diagrams of minigene constructs. Compensatory mutagenesis was performed to restore RNA secondary structure, based on the schematic diagrams of minigene constructs (Fig. 2b, 3b and d; Supplementary Fig. S17). Other mutant constructs were analysed according to the schematic diagrams (Fig. 1, 4–6, Supplementary Fig. S1, S10 and S13). All constructs were confirmed by sequencing. *Drosophila* WT and mutant constructs were further cloned into the pMT/V5-His B vector (Invitrogen, Carlsbad, CA) under the metallothionein promoter. Transfections were performed as previously described[11].

References
1. Matlin, A. J., Clark, F. & Smith, C. W. Understanding alternative splicing: towards a cellular code. *Nat. Rev. Mol. Cell Biol.* **6**, 386–398 (2005).
2. Chen, M. & Manley, J. Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. *Nat. Rev. Mol. Cell Biol.* **10**, 741–754 (2009).
3. Nilsen, T. W. & Graveley, B. R. Expansion of the eukaryotic proteome by alternative splicing. *Nature* **463**, 457–463 (2010).
4. Pan, Q., Shai, O., Lee, L. J., Frey, B. J. & Blencowe, B. J. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat. Genet.* **40**, 1413–1415 (2008).
5. Sultan, M. *et al.* A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. *Science* **321**, 956–960 (2008).
6. Schmucker, D. *et al.* Drosophila Dscam is an axon guidance receptor exhibiting extraordinary molecular diversity. *Cell* **101**, 671–684 (2000).
7. Neves, G., Zucker, J., Daly, M. & Chess, A. Stochastic yet biased expression of multiple *Dscam* splice variants by individual cells. *Nat. Genet.* **36**, 240–246 (2004).
8. Smith, C. W. Alternative splicing--when two's a crowd. *Cell* **123**, 1–3 (2005).
9. Graveley, B. R. Mutually exclusive splicing of the insect *Dscam* pre-mRNA directed by competing intronic RNA secondary structures. *Cell* **123**, 65–73 (2005).
10. Anastassiou, D., Liu, H. & Varadan, V. Variable window binding for mutually exclusive alternative splicing. *Genome Biol.* **7**, R2 (2006).
11. Yang, Y. *et al.* RNA secondary structure in mutually exclusive splicing. *Nat. Struct. Mol. Biol.* **18**, 159–168 (2011).
12. May, G. E., Olson, S., McManus, C. J. & Graveley, B. R. Competing RNA secondary structures are required for mutually exclusive splicing of the Dscam exon 6 cluster. *RNA* **17**, 222–229 (2011).
13. Olson, S. *et al.* A regulator of *Dscam* mutually exclusive splicing fidelity. *Nat. Struct. Mol. Biol.* **14**, 1134–1140 (2007).
14. Kioussis, D., Vanin, E., deLange, T., Flavell, R. A. & Grosveld, F. G. β-globin gene inactivation by DNA translocation in γβ-thalassaemia. *Nature* **306**, 662–666 (1983).
15. Bulger, M. *et al.* Conservation of sequence and structure flanking the mouse and human β-globin loci: the β-globin genes are embedded within an array of odorant receptor genes. *Proc. Natl. Acad. Sci. USA* **96**, 5129–5134 (1999).
16. Tolhuis, B., Palstra, R. J., Splinter, E., Grosveld, F. & de Laat, W. Looping and interaction between hypersensitive sites in the active betaglobin locus. *Mol. Cell* **10**, 1453–1465 (2002).
17. Serizawa, S. *et al.* Negative feedback regulation ensures the one receptor-one olfactory neuron rule in mouse. *Science* **302**, 2088–2094 (2003).
18. Fuss, S. H., Omura, M. & Mombaerts, P. Local and *cis* effects of the H element on expression of odorant receptor genes in mouse. *Cell* **130**, 373–384 (2007).
19. Glenner, H., Thomsen, P. F., Hebsgaard, M. B., Sørensen, M. V. & Willerslev, E. The origin of insects. *Science* **314**, 1883–1884 (2006).
20. Brites, D. *et al.* The Dscam homologue of the crustacean Daphnia is diversified by alternative splicing like in insects. *Mol. Biol. Evol.* **25**, 1429–1439 (2008).
21. Reese, M. G., Eeckman, F. H., Kulp, D. & Haussler, D. Improved splice site detection in genie. *J. Comput. Biol.* **4**, 311–323 (1997).
22. Zhang, C., Hastings, M. L., Krainer, A. R. & Zhang, M. Q. Dual-specificity splice sites function alternatively as 5′ and 3′ splice sites. *Proc. Natl Acad. Sci. USA* **104**, 15028–15033 (2007).
23. Kreahling, J. M. & Graveley, B. R. The iStem, a long-range RNA secondary structure element required for efficient exon inclusion in the *Drosophila* Dscam pre-mRNA. *Mol. Cell. Biol.* **25**, 10251–10260 (2005).
24. Fox-Walsh, K. L. *et al.* The architecture of pre-mRNAs affects mechanisms of splice-site pairing. *Proc. Natl Acad. Sci. USA* **102**, 16176–16181 (2005).
25. Hattori, D. *et al.* Robust discrimination between self and non-self neurites requires thousands of Dscam1 isoforms. *Nature* **461**, 644–648 (2009).
26. Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31**, 3406–3415 (2003).
27. Regier, J. C. *et al.* Arthropod relationships revealed by phylogenomic analysis of nuclear protein-coding sequences. *Nature* **463**, 1079–1083 (2010).

Author contributions
X.W., G.L., W.Z., Y.Yang and W.W. conducted mutational studies and splicing analyses; B.L., L.Z. collected, cloned and analysed the nucleotide sequences; Y.Yue conducted mutational studies and splicing analyses in silkworm; H.P., P.Z. and J.B. made the DNA constructs; H.L. and J.M. analysed sequence data and RNA secondary structures. Y.J. conceived of this project, designed the experiments, analysed the data and wrote the manuscript; Y.Yang, Y.M. and F.S. analysed the data; all authors discussed the results and commented on the manuscript.

Additional information