CRISPR/Cas9 deletions in a conserved exon of Distal-less generates gains and losses in a recently acquired morphological novelty in flies
CRISPR/Cas9 deletions in a conserved exon of Distal-less generates gains and losses in a recently acquired morphological novelty in flies

Gowri Rajaratnam, Ahiraa Supeinthiran, Rudolf Meier, and Kathy F.Y. Su

SUMMARY
Distal-less has been repeatedly co-opted for the development of many novel traits. Here, we document its curious role in the development of a novel abdominal appendage ("sternite brushes") in sepsid flies. CRISPR/Cas9 deletions in the homeodomain result in losses of sternite brushes, demonstrating that Distal-less is necessary for their development. However, deletions in the upstream coding exon (Exon 2) produce losses or gains of brushes. A dissection of Exon 2 reveals that the likely mechanism for gains involves a deletion in an exon-splicing enhancer site that leads to exon skipping. Such contradictory phenotypes are also observed in butterflies, suggesting that mutations in the conserved upstream regions have the potential to generate phenotypic variability in insects that diverged 300 million years ago. Our results demonstrate the importance of Distal-less for the development of a novel abdominal appendage in insects and highlight how site-specific mutations in the same exon can produce contradictory phenotypes.

INTRODUCTION
Insects display a remarkable amount of morphological diversity. Such diversity can be generated through the co-option of existing gene modules in novel environments. One such repeatedly co-opted gene is Distal-less (Dll), which codes for a transcription factor that is essential for insect appendage patterning. Dll has been shown to be involved in the development of many morphological novelties that are essentially distal cuticular projections, like the grasping structures on antenna in male water striders (Khila et al., 2012), the nasus in termites (Toga et al., 2012), and the thoracic horns in scarab beetles (Moczek and Rose, 2009). Dll has even been co-opted for the development of novel pigmentation patterns in the wings of flies (Arnoult et al., 2013) and eyespots in butterflies (Zhang and Reed, 2016). Here, we look at its role in the development and evolution of another morphological novelty, an abdominal sternite brush in sepsids.

Males of some sepsid species (Sepsidae: Diptera) have evolved a moveable appendage on their fourth abdominal segment, the fourth sternite brush (Figure 1A). This novel appendage in sepsids is a modification of a sternite plate on the ventral fourth abdominal segment. In some species, the fourth sternite has been dramatically modified into a large, sclerotized leg-like structure that is complete with underlying musculature, articulation, and long distal brushes (Ang, 2013). Male sepsids use these elaborate appendages to stimulate females during copulation (Figure 1B, https://www.youtube.com/watch?v=BL9wffTKO50&feature=youtu.be). The fourth sternite brushes range from simple flat plates with few bristles to highly elaborate appendages (Figure 1C). A phylogenetic analysis reveals that the fourth sternite brush has a complex evolutionary history. It was acquired early in the radiation of this family, then lost multiple times, and then reacquired at least once (Bowsher et al., 2013).

These structures also have an interesting developmental origin. Unlike typical dipteran appendages (e.g., legs and wings) that develop from imaginal discs, the novel sternite brush appendage in male sepsids develops from a cluster of ventral histoblast cells (Bowsher and Nijhout, 2007). In flies, histoblasts are clumps of imaginal cells that proliferate to form the adult abdominal epidermis, including the sternites and tergites. An earlier study characterized the expression pattern of Dll in late in sepsid pupal development; they found that Dll was expressed only in the bristle cell nuclei of late-stage pupal tissues that give rise to the fourth sternite brush (Bowsher and Nijhout, 2009). Here, we develop and use CRISPR/Cas9 to show that Dll is involved in the development of these novel abdominal appendages in a non-model dipteran species, Themira biloba (Figure 1A). This is the first functionally characterized gene in a gene regulatory network involved in the development and evolution of a novel abdominal appendage. In the process, we discover
an interesting phenomenon wherein deletions in a conserved Dll-coding region generated via CRISPR/Cas 9 produced not only the expected losses but also gains of novel appendages on different abdominal segments.

RESULTS
Genome Editing at the Dll Locus Generates Loss-of-Function Mosaic Phenotypes

We designed three guides targeting two different exons within the Dll locus: the homeodomain (Exon 3) and the coding exon upstream of the homeodomain (Exon 2) (Figure 2A, Table S6). We screened for potential off-targets by conducting a local BLASTN search comparing the transcriptome of T. biloba (Melicher et al., 2014) against the three guide sequences.

We injected 4-hr-old T. biloba embryos with Cas9 protein and guides targeting these coding regions. We observe that Dll loss-of-function mutations are highly lethal; this result is similar to what has been shown in Drosophila melanogaster (Cohen and Ju¨ rgens, 1989). After optimization, we find that 20%–30% of embryos survive microinjection with 22%–37% of larvae developing into adults. In total, we obtained 80 mosaic mutants (Table 1). A total of 58 mutants were observed with malformed or missing sternites; monomorphic sternite malformations were observed in both males and females. However, sternite losses were only observed in males. We observed a higher proportion of abdominal or posterior phenotypes; this is likely due to the injection of CRISPR/Cas9 and single guide RNA (sgRNA) complex into the posterior end of the embryo. Even so, one in eight of the recovered mutants exhibited leg or wing phenotypes consistent with known Dll mutants in D. melanogaster: mosaic mutants with wing margin deformities, hindleg deformities, and loss of tarsal structures were observed (Figures 2B and 2C). All mutants were successfully verified with next-generation sequencing (NGS) on an Illumina MiSeq 2 x 300-bp platform with tagged amplicon sequencing (7-bp tags). The read counts and three most abundant mutant haplotypes for each mutant specimen can be found in Table S3. Only the most abundant mutant haplotype for each specimen was used for downstream alignments and analyses.

As a control, we injected approximately 1,000 embryos with Cas9 alone without obtaining mutant phenotypes (Table 1). The surviving adults from the control injections were genotyped through sequencing with NGS and found to have no mutations in the Dll-coding region. We also sequenced injected individuals displaying a wild-type phenotype and found no mutations in the Dll-coding region.
A Functional Dll Homeodomain Is Necessary for Fourth Sternite Brush Development

Targeted genome editing at the Dll homeodomain produced mosaic mutant males with loss of the fourth sternite appendage. The fourth sternite appendage in T. biloba consists of a bifurcated sternite with a pair of cuticular protrusions or lobes that each terminate in a tuft of bristles. Of the 21 mosaic mutants with sternite malformations, five exhibited malformed or missing fourth sternite appendages (Figure 3C): mutants E3 M1 and E3 M3 display a complete loss of the sternite, lobe, and bristles in one-half of the fourth sternite appendage, whereas in the others (E3 M2, E3 M4, and E3 M5) the lobe and bristles of one-half of the fourth sternite appendage were lost, leaving behind small misshapen remnants of the sternite. Characterization of these mutants with NGS showed deletions in the target region, which disrupted the reading frame (Figure 3B). These mosaic mutant phenotypes confirm that a functional Dll gene product is necessary for the development of this appendage in T. biloba males.

Genome Editing at the Coding Exon Upstream of the Homeodomain Produces Ectopic Structures

Targeting Exon 2, using sgRNA-2B, which is immediately upstream of the homeodomain, produced mosaic mutants with sternite malformations as expected (Figure S1). However, we also recover unexpected mutant males with small ectopic sternite brushes on the third ventral segment (Figure 4A) and the fourth dorsal segment (Figures 4B and 4C). In mutant E2B M1, we observed a small cuticular protrusion with a tuft of bristles on the top right margin of the third ventral sternite (Figure 4A). A similar ectopic cuticular protrusion with bristles was also observed in E2B M2 and E2B M3, wherein the protrusion disrupts the right margin of

| Exon        | Injected Eggs | Larvae | Adults | Mutants | Leg/Wing | Sternite Malformations | Ectopic Structures | Others (e.g., Clasper Malformations) |
|-------------|---------------|--------|--------|---------|----------|------------------------|--------------------|-------------------------------------|
| Exon 2 (sgRNA 2A) | 1,715         | 605    | 151    | 1       | 8        | 1                      | 0                  | 0                                   |
| Exon 2 (sgRNA 2B) | 4,631         | 957    | 363    | 8       | 29       | 3                      | 6                  | 2                                   |
| Exon 3 (sgRNA3)  | 4,003         | 1,187  | 261    | 1       | 21       | 0                      | 2                  |                                     |
| Control       | 1,004         | 153    | 109    | 0       | 0        | 0                      | 0                  | 0                                   |

Table 1. Summary of CRISPR/Cas9 Injection Results

Control injections were carried out using only Cas9 protein without any sgRNA.
the fourth tergite (Figures 4B and 4C). To rule out the possibility that the ectopic phenotypes were a result of off-target gene editing, we screened for potential off-target effects (see Methods and Table S2 for additional details) and found no good matches (>65% identity).

Such exon-specific gain and loss phenotypes have similarly been observed in Lepidoptera wherein deletions in the Dll homeodomain result in losses of eyespots in *Bicyclus anynana* (Connahs et al., 2017), whereas deletions in the region upstream of the homeodomain can produce ectopic eyespots in *B. anynana* (Connahs et al., 2017) and *Vanessa cardui* (Zhang and Reed, 2016) as well as enlarged eyespots in *Junonia coenia* (Zhang and Reed, 2016). Comparative sequence analysis between Lepidoptera and Diptera reveals that this region of *Dll* is almost as highly conserved as the *Dll* functional homeodomain. Further investigation shows that it is conserved even across Holometabola (Figure S2).

To understand the genetic mechanism underlying the development of the ectopic brush phenotype, we sequenced all mosaic mutants using NGS. We find that a single point mutation underlies all ectopic mutant phenotypes (Figure 4D). A bioinformatics analysis of Exon 2 using ESEfinder 3.0 (Cartegni et al., 2003) revealed that this single point mutation lies within a predicted exonic splicing enhancer (ESE) site for the SR protein, splicing factor 2 (SF2) (Table S1). SF2 is an RNA-binding, sequence-specific splice factor that binds to ESE sites to promote inclusion of the exon during alternative splicing at the pre-mRNA level. In contrast, all other mutant phenotypes observed were due to mutations that lie outside of this predicted ESE site (Figure 4D). Previous studies of naturally occurring and artificially induced mutations in ESEs have shown that such mutations can lead to exon skipping (Hong-Xiang et al., 2001; Wang et al., 2002). This naturally occurring form of cellular RNA splicing occurs when defective portions of exons are “skipped” over to restore the reading frame. This produces an altered but sometimes still functional protein, which
is in some instances over-expressed (Chang et al., 2007). It has also been recently shown in adenocarcinoma cells that CRISPR/Cas9-induced mutations lead to exon skipping (Mou et al., 2017).

We hypothesized that the ectopic phenotypes in sepsids may be gain-of-function mutants that resulted from an altered \( Dll \) protein that retains partial or full function. To test this, we investigated whether mutations in Exon 2 (sgRNA-2B) generate exon-skipped \( Dll \) transcripts in embryos. We injected two replicates of 160 embryos with Cas9 protein, and sgRNA-2B. RNA was extracted from the two mutant replicates (C1 and C2) as well as two wild-type replicates. RT-PCR was performed on 12-hr-old mutant and wild-type embryos. Primers were anchored in Exon 1 and Exon 3 of \( Dll \) (see Table S4 for primers). We found an alternatively spliced transcript lacking Exon 2 in both mutant embryo replicates, but not in the wild-type replicates (Figure 5).

Sanger sequencing of the shorter PCR product obtained from C1 and C2 confirms that Exon 2 was skipped. Based on sequence predictions, we find that with leaky scanning this exon-skipped transcript could produce an altered \( Dll \) protein with an intact homeodomain (Figure 6A). If translation of the exon-skipped transcript is initiated at the original start codon, the downstream homeodomain is disrupted, yielding a non-functional protein. However, examination of the \( Dll \) translation initiation site identified a putative alternative start site 1 bp immediately downstream of the Kozak sequence. Translation initiation at this alternative start codon would produce an exon-skipped protein with a functional homeodomain (Figure 6A). To show the feasibility of this mechanism, we performed an in vitro protein synthesis assay using a plasmid containing the \( Dll \)-coding sequence lacking Exon 2 (see Methods for details). We expressed a protein (>30 kDa) in vitro using the PURExpress kit (New England Biolabs) (Figure 6B). An analysis with the TripleTOF 5600 detected the presence of the bioinformatically predicted protein that is obtained only if the alternative initiation codon is utilized. The TripleTOF obtained peptides that correspond to Exon 1 of the predicted leaky-scanned protein, the intact \( Dll \) homeodomain, as well as Exon 5 (Figures 6A and S3).

These experiments suggest that exon skipping occurs and that the ectopic sternite brush phenotypes could arise from a change in expression of an altered \( Dll \) transcript with full or partial function. However, \( Dll \) expression was previously observed only in the developing fourth sternite brush during the late pupal stage (48 hr after puparium formation) (Bowsher and Nijhout, 2009), which could not explain the presence of ectopic brushes on the third abdominal segment as observed in mutant E2B1. To discern whether the underlying predicted change in expression levels was a spatial gain of expression or a mis-regulation of an already expressed gene, we investigated \( Dll \) expression in the developing histoblast clusters in the late larval stage in \( T. biloba \) and \( D. melanogaster \). RT-PCR was performed on dissected epidermal larval segments; the epidermis includes the histoblast clusters that eventually develop into the adult sternites and tergites. We found that \( Dll \) expression in \( D. melanogaster \) was detected in all abdominal segments as well as in the thoracic segment (Figure S5). To rule out any artifacts from imprecise dissections, another gene, Abdominal-B (Abd-B), was tested. As expected in \( D. melanogaster \), Abd-B was only detected in the fifth to eighth abdominal segments. Similar to \( D. melanogaster \), in \( T. biloba \) the RT-PCR results indicate that \( Dll \) is present in all abdominal segments in the larval stages (Figure S4 and Table S5).

### Genome Editing at Exon 2 (sgRNA-2b) Results in Exon Skipping and the Production of an Altered Dll Protein with an Intact Homeodomain

To test the hypothesis that specific mutations within ESE sites produce mutants with ectopic phenotypes, we designed a guide targeting another predicted ESE site within Exon 2 (sgRNA-2A, Table S1) and found genome editing of this site produced both gain and loss phenotypes. (B) and (C) Ectopic mutants with ectopic brushes on the fourth dorsal abdominal segment (E2B1, E2B2, E2B3) lie within a predicted exonic splicing enhancer (ESE), whereas mutations underlying sternite malformation phenotypes (see Table S1) lie outside of the predicted ESE. The alignment is reverse complemented to display the putative ESE site sequence as predicted by ESEfinder 3.0 (Cartegni et al., 2003).
comparable results. We observed nine mutants with sternite malformations with one mutant exhibiting a complete loss of one-half of the fourth sternite appendage (E2A M2). Interestingly, we also observed one mutant male (E2A M1) with an ectopic fourth sternite brush: on the right half of the fourth sternite appendage, the third sternite is expanded, whereas the fourth sternite is misshapen and ends in two bristled lobes instead of the expected one (Figure 7A). Upon characterization of all the mutants using NGS, we observed that only the mutations underlying the ectopic brush mutant disrupted the targeted ESE site. The most abundant mutant haplotype was a single point deletion within the ESE site, whereas the second most abundant mutant haplotype was a 57-bp deletion that disrupts the targeted ESE site as well as an additional downstream putative ESE site. As predicted, the mutations underlying the loss and other sternite malformations lay outside of the ESE site. These results lend further support to the hypothesis that specific CRISPR/Cas9-induced mutations that disrupt ESE sites have the potential to yield ectopic phenotypes, either through an over-expression or a prolonged expression of an altered but functional Dll protein.

To discern if this exon-skipped transcript was exclusively the result of targeted genome editing at Exon 2 (sgRNA-2A), we performed targeted long-read isoform sequencing on Pacbio Sequel to qualitatively characterize the alternative splice forms of Dll present in both exon-skipped mutants and wild-type individuals. Using a guide targeting Exon 2, we injected 405 embryos and screened the surviving third instar larvae. RNA was extracted from 12 injected and 2 wild-type larvae and used to synthesize cDNA. Using an in vitro cleavage assay, 2 of the 12 larvae were identified as exon-skipped mutants (M4 and M6). We then designed Dll-specific tagged primers anchored in Exon 1 and Exon 7 (see Table S4 for tagged primer sequences) to amplify Dll splice forms, which were then sent in for Pacbio Sequel targeted isoform circular consensus sequencing. This generates a consensus sequence from multiple passes of a single circular molecule. The reads were analyzed using the Pacbio SMRT Link 5.1.0 analysis software (see Methods for parameters).

From the sequencing data, we identified three splice forms alongside the wild-type splice form (Figure 8). Splice forms B and C were detected in M6 and both wild-type larvae. However, splice form A, which excluded Exon 2, was only detected in the exon-skipped mutants, M4 and M6. These results suggest that the exon-skipped transcript (splice form A) does not naturally occur and is instead produced only when Exon 2 (sgRNA-2A) is disrupted with CRISPR/Cas9.

To test if changed expression levels of Dll can be detected in ectopic mutants, we carried out a quantitative PCR assay on injected mosaic mutant larvae despite the high risk of such an experiment yielding a false-negative result because (1) mosaic mutant larvae have a mixture of wild-type, mutant, and exon-skipped transcripts; (2) only mosaic mutants with a high proportion of wild-type cells are predicted to be viable;
and (3) the natural expression levels ofDllat this stage are low. The results of the qPCR assay were inconclusive (see Methods for details).

DISCUSSION

The overall body plan of winged insects has remained remarkably conserved with regard to the position of appendages, which are restricted to the head, thorax, and posterior-most segments of the 11-segmented abdomen (genitalia). Occasional exceptions to this body plan are losses or reductions of structures, such as the reduction of maxillary and mandibular structures in D. melanogaster (Angelini and Kaufman, 2006). However, gains of articulated appendages, especially in the pregenital abdominal segments, are rare. Sepsids are one of the few exceptions (Hoch et al., 2014), having very recently evolved a novel abdominal appendage. This makes sepsids an attractive model system for understanding how a gene regulatory network is assembled for the development of novel appendages.

Morphological novelties are often hypothesized to arise from either existing or de novo genetic machinery (True and Carroll, 2002; Wagner and Lynch, 2010). Genes can either be co-opted along with their existing network or be assembled differently into a de novo network. In addition, novel phenotypes may also arise from the evolution of de novo or orphan genes. Recent studies have shown that the genetic architecture underlying morphological novelties in insects is more complex than expected; in some instances, a combination of both de novo genes and the co-option of an existing gene regulatory network underly a novel phenotype (Santos et al., 2017; Hilgers et al., 2018), whereas in others, existing gene regulatory networks are modified and/or partially co-opted (Hu et al., 2018; Moczek, 2009; Moczek and Rose, 2009; Glassford et al., 2015). We illustrate in this study another instance whereby an important gene in the appendage-patterning gene regulatory network,Dll, is co-opted in the development of a novel morphological structure: the fourth sternite brush.

Here, we show conclusively thatDllis necessary for the development of a novel abdominal appendage in T. biloba; disruptions in the homeodomain and in the coding region upstream lead to losses of the fourth sternite brush. We also observed malformed monomorphic sternites in both males and females, raising the question of howDllfits into a gene regulatory network that patterns a sexually dimorphic trait. To obtain a better understanding of the underlying gene architecture responsible for building this morphological novelty, functional characterization of more appendage and sex-patterning genes (e.g., doublesex) would have to be carried out.

Moreover, we also reveal that specific deletions in Exon 2 can lead not only to the expected losses but also to unexpected gains of the novel fourth sternite brush. Through a detailed dissection of the upstream coding exon, we show that deletions in ESE sites can generate the ectopic phenotypes observed and go further to prove that the ensuing exon-skipped transcripts can be translated in vitro into an altered Dll protein that retains an intact homeodomain. These findings highlight how small modifications ofDllhave the potential to generate very different phenotypes. This would be a single observation in a fly species if it were not
for the fact that a similarly diverse set of phenotypes can be generated by mutations in the same coding region in butterflies, which diverged from flies nearly 300 million years ago (Misof et al., 2014). In both sepsid flies and several species of butterflies (V. cardui and B. anynana), disruption of the region upstream of the Dll homeodomain produces ectopic structures (Zhang and Reed, 2016; Connahs et al., 2017) and exon-skipped Dll transcripts (Connahs et al., 2017). Moreover, a comparative analysis across Holometabola reveals that the region of Dll immediately upstream of the homeodomain is highly conserved. The conserved nature of this protein region as well as the appearance of ectopic traits in two divergent lineages suggests that mutations at this conserved region may have had the potential to generate morphological variation for at least 300 million years. Although we find in T. biloba that the exon-skipped transcript does not occur naturally, further investigation into B. anynana and other holometabolan lineages might provide more insight into the possible role of this conserved region in the evolution and development of structural novelties.

Based on our results, we propose that Dll genome editing studies should target multiple exons and include the exon upstream of the homeodomain. Such screening may yield particularly interesting results for those species that have novel traits whose development involves Dll. Our study suggests that this gene may not only be important for the origin of novel traits but also has the potential for generating morphological diversity via splicing regulation. These findings also have implications beyond Dll as CRISPR/Cas9 is now extensively adopted for genome editing purposes from single cells to whole organisms. Future CRISPR/Cas9 studies should target multiple exons, particularly functional domains, to identify phenotypes generated by exon skipping. Bioinformatics tools should be also used to predict

---

**Figure 7. Mutant Phenotypes Obtained with Genome Editing of Second Putative ESE Site in Exon 2 (sgRNA-2A)**

(A–C) (A) Mutant with ectopic brush on fourth ventral abdominal segment (E2A M1). (B) Mutant with loss of half of the fourth sternite brush (E2A M2). (C) Sequences confirm that mutations underlying the ectopic brush phenotype (E2A M1) lie within a predicted exonic splicing enhancer (ESE), whereas mutations underlying sternite malformation phenotypes lie outside of the predicted ESE. The two most abundant mutant haplotypes for E2A M1 are shown in this alignment (E2A M1a is a single-point deletion, whereas E2A M1b is a 57-bp deletion that disrupts this target ESE site as well as another downstream ESE site).
ESE sites in potential target regions, which should either be avoided to reduce the chances of generating conflicting results or targeted with CRISPR/Cas9 to test whether gain-of-function mutations can be produced.

Limitations of Study
Our study suggests that deletions in ESE sites result in exon skipping and the development of ectopic structures. However, the process by which exon-skipped transcripts produce the ectopic structures remains unclear because quantifying gene expression of mutated cells within a mosaic mutant is difficult, i.e., we were not able to isolate the signal from mutant cells alone. We predict that the exon-skipped transcript is functional because the homeodomain is intact. However, in vivo tests for protein folding and functionality would be desirable.

METHODS
All methods can be found in the accompanying Transparent Methods supplemental file.

DATA AND SOFTWARE AVAILABILITY
Most of the processed sequencing data files are available on a Mendeley database https://doi.org/10.17632/ps3p7jnb5t.1.

SUPPLEMENTAL INFORMATION
Supplemental Information includes Transparent Methods, five figures, and six tables and can be found with this article online at https://doi.org/10.1016/j.isci.2018.11.036.

ACKNOWLEDGMENTS
This work was supported by a Lee Kuan Yew Postdoctoral Fellowship grant (R-154-000-646-112) and an MOE AcRF Tier 2 grant (R-154-000-A62-112). The authors would like to thank Amrita Srivathsan and Ang...
Yuchen for their assistance with bioinformatics and imaging, respectively. We also thank Zhang Yizhong and Lee Zheng Fen for their assistance with microinjections. We extend our gratitude to Mindy Tuan for providing us with the mating behavior of Themira superba, Dacotah Melicher for providing us with Dll scaffolds, Fong Siao Wan for her assistance with the RNA extractions and RT-PCR experiments, Antonia Monteiro and Heidi Connahs for the stimulating discourse, and Sunita Subramaniam, Lim Teck Kwang, Anjali Gupta, and Thorsten Wohland for assistance with the protein experiments. We would also like to thank Liu Yih-Cherng for feedback on the protein experiments. For help with our qPCR experiment, we would like to thank Hiong Kum Chew and Ip Yuen Kwong. We also thank Julia H. Bowsher for her assistance with the design of the RT-PCR experiments and her insightful feedback on the manuscript.

AUTHOR CONTRIBUTIONS
K.F.Y.S. and G.R. planned the study, performed most of the experiments, and wrote the paper with feedback from the other authors. K.F.Y.S., G.R., and A.S. performed the microinjection experiments, and G.R. conducted the imaging, characterization, and analysis of CRISPR/Cas9 mutants. K.F.Y.S. designed the guides for CRISPR experiments and performed the in vitro protein synthesis assay, reverse-transcriptase PCR, qPCR experiments, and alignments of Dll across Holometabola. G.R. performed the SDS protein gels and the ProteinPilot analysis. K.F.Y.S. designed the Dll isoform experiment, and K.F.Y.S. and G.R. processed and analyzed the results. R.M. supervised and supported the research; all authors discussed the results and provided comments for the manuscript.

DECLARATION OF INTERESTS
The authors declare no competing financial interests.

Received: August 8, 2018
Revised: November 26, 2018
Accepted: November 27, 2018
Published: December 21, 2018

REFERENCES
Ang, Y. (2013). Modern Morphological Techniques and the Evolutionary Biology and Taxonomy of Sepsidae (Diptera), PhD thesis (National University of Singapore).

Angelini, D.R., and Kaufman, T.C. (2006). Insect development and comparative ontogenetics. Evolution: cell diversity in development and evolution: a case study of the sepsid fly Themira superba. Evolution 60, 1423–1426.

Bowsher, J.H., and Nijhout, H.F. (2007). Evolution of novel abdominal appendages in a sepsid fly from histoblasts, not imaginal discs. Evol. Dev. 9, 347–354.

Bowsher, J.H., and Nijhout, H.F. (2009). Partial co-option of the appendage patterning pathway in the development of abdominal appendages in the sepsid fly Themira biloba. Dev. Genes Evol. 219, 577–587.

Cartegni, L., Wang, J., Zhu, Z., Zhang, M.Q., and Kraainer, A.R. (2003). ESEfinder: a web resource to identify exonic splicing enhancers. Nucleic Acids Res. 31, 3568–3571.

Chang, Y.-F., Chan, W.-K., Imam, J.S., and Wilkinson, M.F. (2007). Alternatively spliced T-cell receptor transcripts are up-regulated in response to disruption of either splicing elements or reading frame. J. Biol. Chem. 282, 29738–29747.

Cohen, S.M., and Jürgens, G. (1989). Proximal—distal pattern formation in Drosophila: cell autonomous requirement for Distal-less gene activity in limb development. EMBO J. 8, 2055–2055.

Connahs, H., Tili, S., van Creij, J., Loo, T.Y.J., Banerjee, T., Saunders, T.E., and Monteiro, A. (2017). Disrupting different Distal-less exons leads to ectopic and missing eyespots accurately modeled by reaction-diffusion mechanisms. bioRxiv.

Glassford, W.J., Johnson, W.C., Dall, N.R., Smith, S.J., Liu, Y., Boll, W., Noll, M., and Rebeiz, M. (2015). Co-option of an ancestral hox-regulated network underlies a recently evolved morphological novelty. Dev. Cell. 34, 520–531.

Hilgers, L., Hartmann, S., Hofreiter, M., and von Rintelen, T. (2018). Novel genes, ancient genes, and gene co-option contributed to the genetic basis of the radula, a molluscan innovation. Mol. Biol. Evol. 35, 1638–1652.

Hoch, H., Wessel, A., Asche, M., Baun, D., Beckmann, F., Braunig, P., Ehrig, K., Mühlthaler, R., Riesemeier, H., Stauda, A., et al. (2014). Non-sexual abdominal appendages in adult insects challenge a 300 million year old bauplan. Curr. Biol. 24, R16–R17.

Hong-Xiang, L., Cartegni, L., Zhang, M.Q., and Krainer, A.R. (2001). A mechanism for exon skipping caused by nonsense or missense mutations in BRCA1 and other genes. Nat. Genet. 27, 55–58.

Hu, Y., Schmitt-Engel, C., Schwarz, J., Stroehlein, N., Richter, T., Majumdar, U., and Bucher, G. (2018). A morphological novelty evolved by co-option of a reduced gene regulatory network and gene recruitment in a beetle. Proc. Biol. Sci. 285.

Khila, A., Abouheif, E., and Rowe, L. (2012). Function, developmental genetics, and fitness consequences of a sexually antagonistic trait. Science 336, 585–589.

Melicher, D., Torson, A.S., Dworkin, I., and Bowsher, J.H. (2014). A pipeline for the de novo assembly of the Themira biloba (Sepsidae: Diptera) transcriptome using a multiple k-mer length approach. BMC Genomics 15, 188.

Misof, B., Liu, S., Meusemann, K., Peters, R.S., Donath, A., Mayer, C., Frandsen, P.B., Ware, J., Flouri, T., Beutel, R.G., et al. (2014). Phylogenomics resolves the timing and pattern of insect evolution. Science 346, 763–767.

Moczek, A.P. (2009). On the origins of novelty and diversity in development and evolution: a case study of the sepsid fly Themira superba. Cell Press

iScience 10, 222–233, December 21, 2018
study on beetle horns. Cold Spring Harb. Symp. Quant. Biol. 74, 289–296.

Moczek, A.P., and Rose, D.J. (2009). Differential recruitment of limb patterning genes during development and diversification of beetle horns. Proc. Natl. Acad. Sci. U S A 106, 8992–8997.

Mou, H., Smith, J.L., Peng, L., Yin, H., Moore, J., Zhang, X.-O., Song, C.-Q., Sheel, A., Wu, Q., Ozata, D.M., et al. (2017). CRISPR/Cas9-mediated genome editing induces exon skipping by alternative splicing or exon deletion. Genome Biol. 18, 108.

Santos, M.E., Le Bouquin, A., Crumière, A.J.J., and Khila, A. (2017). Taxon-restricted genes at the origin of a novel trait allowing access to a new environment. Science 358, 386–390.

Toga, K., Hojo, M., Miura, T., and Maekawa, K. (2012). Expression and function of a limb-patterning gene Distal-less in the soldier-specific morphogenesis in the nasute termite Nasutitermes takasagoensis. Evol. Dev. 14, 286–295.

True, J.R., and Carroll, S.B. (2002). Gene co-option in physiological and morphological evolution. Annu. Rev. Cell Dev. Biol. 18, 53–80.

Wagner, G.P., and Lynch, V.J. (2010). Evolutionary novelties. Curr. Biol. 20, R48–R52.

Wang, J., Hamilton, J.I., Carter, M.S., Li, S., and Wilkinson, M.F. (2002). Alternatively spliced TCR mRNA induced by disruption of reading frame. Science 297, 108–110.

Zhang, L., and Reed, R.D. (2016). Genome editing in butterflies reveals that spalt promotes and Distal-less represses eyespot colour patterns. Nat. Commun. 7, 11769.
Supplemental Information

CRISPR/Cas9 deletions in a conserved exon of Distal-less generates gains and losses in a recently acquired morphological novelty in flies

Gowri Rajaratnam, Ahiraa Supeinthiran, Rudolf Meier, and Kathy F.Y. Su
Supplemental Information

CRISPR/Cas9 deletions in a conserved exon of Distal-less generates gains and losses in a recently acquired morphological novelty in flies.

Authors:
Gowri Rajaratnam¹, Ahiraa Supeinthiran³, Rudolf Meier¹,²* and Kathy F.Y. Su¹*

¹Department of Biological Sciences, National University of Singapore, Singapore
²Lee Kong Chian Natural History Museum, Singapore
³University of Toronto Scarborough, Canada
*corresponding author
**Transparent Methods**

**Fly cultures**

*Themira biloba* fly cultures were kept at a constant temperature of 23 °C with a 12-hour light and dark cycle and maintained on a diet of duck and bovine dung.

**Preparation of guide RNAs**

Three CRISPR guide RNAs targeting Exon 2 and Exon 3 (homeodomain) were designed for *T. biloba*. *T. biloba* assembled transcriptome was downloaded from NCBI, Accession number PRJNA218740 (Melicher et al., 2014). Using CLC Genomics Workbench we searched for Dll transcript sequences within the *T. biloba* transcriptome using BLASTn against *D. melanogaster* Dll. Using CLC Genomics Workbench we then designed guides targeting Exon 2 and Exon 3 of *T. biloba* Dll. We searched for regions that contained GGN_{19}NGG. The guides were screened for off-target effects using a blastn query against the assembled *T. biloba* transcriptome (Melicher et al., 2014). Specifically, for sgRNA-2B, this search did not find any 100% off-target matches. However, it did identify three hits with a hit-length of 13 bp and E-values around 1E-3. Further investigation of these hits reveal that they do not flank a PAM site and so are unlikely to result in off-target effects (Supplementary table 2).

The single guide RNA (sgRNA) templates were synthesised artificially using gBlocks® Gene Fragments (Integrated DNA Technologies). Following *in vitro* transcription of the gblocks DNA template with a T7 polymerase (New England Biolabs), the sgRNAs were purified using the TURBO DNase kit followed by the Qiagen RNeasy kit. *In vitro* Cas9 cleavage assays were then carried out to test the guide’s ability to cut the target DNA fragment in the presence of Cas9 protein (PNAbio).

**Microinjection**

Eggs were extracted from the egg-laying substrate after 4 hours, rinsed gently, aligned onto a coverslip, and covered with a thin layer of oil to prevent desiccation. 1ul of 1ug/ul Cas9 protein (PNAbio) was mixed with 1ul of 1ug/ul of sgRNA and injected into the posterior of the eggs using a 76mm needle. Needles were pulled fresh with a Sutter P-97 Flaming/Brown type micropipette puller. To overcome the lethality of a *Dll* knockdown in early embryogenesis (Cohen and Jürgens, 1989), the protocol was optimized to reduce the number
of mosaic mutant cells generated by injecting 4-hour old embryos with a reduced volume of
the sgRNA/Cas9 complex.

The injected eggs were then rinsed thoroughly and kept in a moist chamber for 24 hours. The
emerging larvae were then picked out and placed into petri dishes of bovine dung to develop.
After 8-10 days, the pupae were recovered and left to develop in a moist chamber. The
emerging adults were allowed to completely sclerotize for 3 days before they were screened
for mutations. The mutant individuals were then preserved in 70% ethanol for imaging and
dNA extraction. Images were obtained using a Visionary Digital Imaging system.

Genotyping of Mutations

QuickExtract (Bioline) solution was used to extract gDNA from mutant tissue dissected from
the vicinity of affected structure. Extracted DNA was used to amplify the gene region of
interest (~200 bp fragment). The tarsi of a wildtype T. biloba male were dissected and used as
a control for this experiment.

PCR products were cleaned up with SureClean and sent for Illumina Miseq (2x301bp)
sequencing at 10,000x coverage (the read count and coverage for each sample is listed in
Supplementary table 3). These mutations were mostly small deletions within the guide
sequence. To ensure that the mutations observed were not an artefact of PCR or sequencing
error, 3 PCR replicates (with tagged primers, Supplementary table 4) were carried out for all
mutants. The reads for all three replicates were processed separately and combined after
ascertaining that the three most abundant reads for each were identical. This combined
dataset was then used for the quantification of the read count and proportions of mutant
haplotypes for each specimen.

The DNA reads were then recovered using an in-house pipeline: the sequences were merged
using PEAR(Zhang et al., 2014) and demultiplexed using an in-house script (Meier et al.,
2016) to generate a fasta file for each PCR product/specimen. The reads were aligned using
MAFFT v7.0 (Katoh et al., 2002). Once aligned, the sequences were submitted to the ‘DNA
to haplotype collapsor and converter’ Fabox tool (Villesen, 2007) This generates a table of
haplotypes and read counts as well as a summary of all the haplotypes observed. The three
most abundant haplotypes were recorded (see Supplementary table 3). The most abundant
mutant haplotype for each specimen was used for downstream alignments and comparisons.
The results were also confirmed with CRISPresso (Pinello et al., 2016) (window size set to
30 bp and sequence homology for an HDR occurrence set to 98%). Bioinformatics analysis
for ESE sites were performed using ESEfinder (v 3.0)(Cartegni et al., 2003). We used the weighted matrix values for SRSF1, the human homolog of the Drosophila SF2/ASF at the threshold of 1.956 (Smith et al., 2006).

**Control Injections**

To rule out off-target effects and injection artefacts, 1004 embryos were injected with 500ng of Cas9 alone. Note that the control injections were used to train students in the microinjection technique and as such, the mortality for the control injections is high overall due to mechanical damage. No mutants of any type were observed for the surviving adults. Adults from control injections were genotyped and no mutations were observed.

**Exon-skipping**

In order to detect if exon skipping was occurring when Exon 2 was targeted, RNA was extracted from two independent replicates of 160 injected embryos (C1 and C2) as well as two wildtype replicates using TRIZol (Invitrogen). The RNA was transcribed into cDNA using the ProtoScript II First Strand cDNA synthesis kit (New England Biolabs). The gene region of interest was then amplified. The same volume of product was loaded onto a 1% agarose gel for both the wildtype and mutant replicates. A shorter band (~ 400bp) was observed for only C1 and C2 and not the wildtype replicates (Fig.5A). The shorter band was then sent for Sanger sequencing, where a transcript lacking Exon 2 was found.

**Protein analysis**

PCR was used to generate the template needed for in vitro protein synthesis. Primers specific to the start and stop codons of Dll coding sequence were used to add specific adaptor sequences (see PURExpress manual and supplementary table 4) to the Dll coding sequence lacking Exon 2. A T7 promoter and ribosome binding site were added to the upstream of the start of Dll translation and a 35-mer loop structure was added to the 3’ of Dll coding sequence (See PURExpress manual and supplementary table 4 for primer sequences). This PCR product was purified using the Qiagen PCR purification kit and used as a template for in vitro protein synthesis with the PURExpress kit (New England Biolabs). Both the Dll template as well as a control reaction with no template were set-up and processed in the same way. 8ul of this synthesised protein product was run out on an SDS-PAGE gel (10% Mini-PROTEAN® TGX™ Precast Gel) at 120V for 90 minutes, along with the control. The >30 kDa band, representing the synthesized Dll protein, was excised from the gel and sent for
analysis on the Triple TOF 5600 (Ab SCIEX). Unfortunately, we noticed that although the same amount of the reaction was loaded onto the gel, the control reaction appeared fainter than the reaction with the Dll template.

As an additional measure to rule out the possibility that the >30kDa band was simply not observed in the fainter control, we excised out the corresponding area in the control gel lane and sent it in for mass spectrometry analysis as well.

The Mass Spectra raw data was searched against the NCBI D. melanogaster protein database and the E. coli database supplemented with the Dll predicted protein sequence for Themira biloba with ProteinPilot® v4.5 (Revision: 1656; Paragon Algorithm: 4.5.0.0, 1654). Searches were run as thorough identification searches, specifying tryptic digestion and cys-alkylation (Iodoacetamide). For the Dll protein sample, peptides were found with strong matches to an alternative initiation codon as well as to the homeodomain and Exon 4 (Supplementary Figure 3). No confident peptide matches were found for the analogous control sample (supplementary file in Mendeley resource).

Distal-less isoform characterisation

In order to qualitatively characterise the alternative splice forms of Dll present in both mutant and wildtype individuals, targeted long-read isoform circular consensus sequencing (Gonzalez-Garay, 2016) was conducted for 4 individuals. 405 embryos were injected with sgRNA-2A (targeting the second ESE site within Exon 2). After 7 days, total RNA and DNA was isolated from 12 injected 3rd instar larvae using TRIzol® Reagent (Invitrogen). cDNA was synthesised using the ProtoScript II First Strand cDNA synthesis kit (New England Biolabs) for each of these individuals and an in-vitro cleavage assay was performed to identify the mutant specimens. Two individuals (M4 and M6) were observed with mutations and exon skipping.

Similarly, RNA was extracted and used for cDNA synthesis for two wildtype 3rd instar larvae (WT1 and WT2). Primers were designed within the Dll locus to amplify and capture Dll specific splice forms (Supplementary Table 4). 5’ phosphorylated forward primers were designed in Dll exon 1 (5’ UTR) and reverse 16bp-tagged primers were designed for Dll exon 7 (3’ UTR). A 16bp tag unique to each specimen was attached to the 3’ end of the reverse primer so as to identify the two wildtype and two mutant sequences (see Supplementary table 4 for primer and tag sequences). The PCR products were purified with Sureclean and sent in for Pacbio Sequel circular consensus sequencing with a library insert size of 2kb.
Pacbio IsoSeq files were error-corrected and converted into CCS reads; We performed 10 minimum passes with a minimum predicted accuracy of 0.9 using the PacBio SMRT analysis software (v5.1.0). The analysis generated 225,740 CCS reads with a median CCS read length between 1200 – 1249 bp (see Supplementary Figure 5 for results statistics).

The reads were then demultiplexed using Geneious with a 100% stringency match to primer and tag sequences. The demultiplexed sequences were then filtered for low read length (< 500bp) yielding 18194 reads for M4, 124385 reads for M6, 13422 reads for WT1 and 18817 reads for WT2. For each individual, the reads were collapsed into haplotypes using the ‘DNA to Haplotype converter’ tool in FaBox (Villesen, 2007). Haplotypes with read counts below 0.5% of the total count were discarded. The remaining haplotypes were aligned using MAFFT v7.0 and then analysed in Aliview to identify alternative splice forms (alignment files in Mendeley resource).

**Distal-less expression in developing histoblast clusters**

To determine if Dll was naturally expressed in the 3\textsuperscript{rd} abdominal segment where an ectopic sternite brush was observed, RNA was extracted from dissected epidermal tissues of 3\textsuperscript{rd} instar larvae. The epidermis of the thoracic segment and the abdominal segments (8 abdominal segments for D. melanogaster and 7 abdominal segments for T. biloba) were dissected for 5 wildtype 3\textsuperscript{rd} instar larvae for both T. biloba and D. melanogaster. RNA was extracted for each segment and cDNA was synthesised. RT PCR for Dll was carried out for each segment to check for Dll expression in the late larval stages.

The same was done for D. melanogaster, which served as a control for the RT-PCR experiment. RT- PCR for an additional gene, Abdominal-B, was also carried out in D. melanogaster as a control to rule out the presence of any artefacts from the epidermal dissections. All PCR products were then PCR purified and sent in for sanger sequencing to confirm the correct products were amplified (results in Supplementary Table 5, Supplementary Figure 4).

**Distal-less alignment across Holometabola**

A protein search performed on NCBI using the following search terms “Distal-less[All Fields] AND ("Mandibulata"[Organism] OR Mandibulata[All Fields])” yielded 467 protein sequences. This dataset was filtered for sequences that: were not Distal-less, had no homeodomain, were incomplete or of poor quality. Sequences belonging to Coleoptera,
Lepidoptera, Diptera and Hymenoptera were extracted from the filtered dataset and aligned with MAFFT v7.0 (Katoh et al., 2002) and visualised in Aliview (Larsson, 2014).

Quantitative PCR

We injected embryos with both Cas9 and the guide targeting Exon 2 and compared expression levels to control embryos that were injected with Cas9 alone. We let embryos develop into first instar larvae. All control larvae were individually extracted for RNA using Trizol. All treatment larvae were first genotyped to confirm presence of mutations in Exon 2 using the T7 endonuclease kit (New England Biolabs). RNA from mutant larvae was then used for cDNA synthesis to generate template for qPCR. The Forkhead transcription factor (Mnf) gene was used as a housekeeping gene. We utilised a customised TaqMan gene expression assay with primers nested within the homeodomain and with a probe complementary to the homeodomain. The expression levels were not significantly different between wildtype and mutant larvae. However, note that CRISPR/Cas9 generates mosaic mutants that consist mostly of wildtype cells and mutants that survive are likely to only have small amounts of mutant cells. This makes it difficult to disentangle expression levels between wildtype mRNA from wildtype cells and exon-skipped transcripts of mutant cells. With naturally low Dll expression (Ct value > 30), a lack of significant expression difference is inconclusive.

Data availability

Most of the processed sequencing data files are available on a Mendeley database (DOI: 10.17632/ps3p7jnb5t.1). However, for pre-filtered/processed files and other data, please contact Rajaratnam, G. For the in-house bioinformatics script, please contact Meier, R.

References

CARTEGNI, L., WANG, J., ZHU, Z., ZHANG, M. Q. & KRAINER, A. R. 2003. ESEfinder: a web resource to identify exonic splicing enhancers. Nucleic Acids Research, 31, 3568-3571.

COHEN, S. M. & JÜRGENS, G. 1989. Proximal—distal pattern formation in Drosophila: cell autonomous requirement for Distal-less gene activity in limb development. The EMBO Journal, 8, 2045-2055.

GONZALEZ-GARAY, M. L. 2016. Introduction to Isoform Sequencing Using Pacific Biosciences Technology (Iso-Seq). In: WU, J. (ed.) Transcriptomics and Gene Regulation. Dordrecht: Springer Netherlands.

KATOH, K., MISAWA, K., KUMA, K.-I. & MIYATA, T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Research, 30, 3059-3066.
LARSSON, A. 2014. AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics*, 30, 3276-3278.

MEIER, R., WONG, W., SRIVATHSAN, A. & FOO, M. 2016. $1 DNA barcodes for reconstructing complex phenomes and finding rare species in specimen-rich samples. *Cladistics*, 32, 100-110.

MELICHER, D., TORSON, A. S., DWORKIN, I. & BOWSHER, J. H. 2014. A pipeline for the de novo assembly of the Themira biloba(Sepsidae: Diptera) transcriptome using a multiple k-mer length approach. *BMC Genomics*, 15, 188.

PINELLO, L., CANVER, M. C., HOBAN, M. D., ORKIN, S. H., KOHN, D. B., BAUER, D. E. & YUAN, G.-C. 2016. Analyzing CRISPR genome-editing experiments with CRISPRe. *Nat Biotech*, 34, 695-697.

SMITH, P. J., ZHANG, C., WANG, J., CHEW, S. L., ZHANG, M. Q. & KRAINER, A. R. 2006. An increased specificity score matrix for the prediction of SF2/ASF-specific exonic splicing enhancers. *Hum Mol Genet*, 15, 2490-508.

VILLESEN, P. 2007. FaBox: an online toolbox for fasta sequences. *Molecular Ecology Notes*, 7, 965-968.

ZHANG, J., KOBERT, K., FLOURI, T. & STAMATAKIS, A. 2014. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics*, 30, 614-620.
Figures

Figure 1

Figure 2
Figure 3.
Figure 4
Figure 5

Figure 6
Figure 7
### Tables

#### Table 1.

|                | Injected eggs | Larvae | Adults | Mutants |
|----------------|---------------|--------|--------|---------|
|                |                |        |        | Leg/Wing | Sternite malformations | Ectopic structures | Others (e.g. clasper malformations) |
| Exon 2 (sgRNA 2A) | 1715           | 605    | 151    | 1        | 8                    | 1                    | 0                  |
| Exon 2 (sgRNA 2B) | 4631           | 957    | 363    | 8        | 29                   | 3                    | 6                  |
| Exon 3 (sgRNA3)  | 4003           | 1187   | 261    | 1        | 21                   | 0                    | 2                  |
| Control         | 1004           | 153    | 109    | 0        | 0                    | 0                    | 0                  |
Supplementary information

Supplementary Table 1. Summary of ESEfinder search with Matrix values for SRSF1 (human homolog for SF2/ASF) with threshold of 1.956. The ESE motif present in Exon 2 sgRNA-2B is highlighted in bold while the ESE motif used to design sgRNA-2A is italicized. Related to Figure 4.

| Sequence ID | Motif | Position on Exon 2 | Motif | Score  |
|-------------|-------|---------------------|-------|--------|
| gi|654231031|gb|GBGG01003309.1| SRSF1 | 100 | cgcccat | 2.21734 |
| gi|654231031|gb|GBGG01003309.1| SRSF1 | 17  | cagcgg | 4.39195 |
| gi|654231031|gb|GBGG01003309.1| SRSF1 | 49  | gtcagga | 3.28514 |
| gi|654231031|gb|GBGG01003309.1| SRSF1 | 75  | cgccgg | 2.69094 |
| gi|654231031|gb|GBGG01003309.1| SRSF1 | 15  | cgcagtg | 2.23323 |
| gi|654231031|gb|GBGG01003309.1| SRSF1 | 169 | cagcagc | 3.67496 |

Supplementary Table 2. Summary of blastn results. Related to Figure 4.

| Query | Number of HSPs | Lowest E-value | Accession (E-value) | Greatest identity % | Greatest HSP length | Greatest bit score | Sequence overlap with guide % |
|-------|----------------|----------------|---------------------|---------------------|---------------------|-------------------|-----------------------------|
| gi|654231031|gb|GBGG01003309.1| (Themira biloba Dll) | 1 | 1.38661E-06 | sgRNA-2B | 100 | 20 | 32.7626 | 100 |
| gi|654228759|gb|GBGG01004492.1| | 1 | 0.0066504 | sgRNA-2B | 100 | 13 | 21.9569 | 65 |
| gi|654228161|gb|GBGG01004814.1| | 1 | 0.00242708 | sgRNA-2B | 100 | 13 | 21.9569 | 65 |
| gi|654223893|gb|GBGG01007092.1| | 1 | 0.00259143 | sgRNA-2B | 100 | 13 | 21.9569 | 65 |

Supplementary Table 3. Summary of read counts for mutant haplotypes. The 3 most dominant mutant haplotypes per individual are recorded here. The dominant haplotype (in bold) per individual is used for downstream analyses. See separate excel document. Related to Figure 2, 3, 4 and 7.
### Supplementary Table 4. Primer sequences. Tags are in lower case. Related to Figure 2, 3, 4, 5, 7 and 8.

| Purpose                                      | Name            | Primer sequence (5' → 3')                                                                 |
|----------------------------------------------|-----------------|------------------------------------------------------------------------------------------|
| Characterisation of homeodomain mutants      |Dll_sgrna12_F1   | cagtcctgGACAAATGCGAAGATTCTGG                                                            |
| Characterisation of homeodomain mutants      |Dll_sgrna12_F2   | catgggaGACAAATGCGAAGATTCTGG                                                             |
| Characterisation of homeodomain mutants      |Dll_sgrna12_F3   | tcacgtaGACAAATGCGAAGATTCTGG                                                             |
| Characterisation of homeodomain mutants      |Dll_sgrna12_F4   | tggtccaGACAAATGCGAAGATTCTGG                                                             |
| Characterisation of homeodomain mutants      |Dll_sgrna12_F5   | aacctgtGACAAATGCGAAGATTCTGG                                                             |
| Characterisation of homeodomain mutants      |Dll_sgrna12_F6   | cttggttGACAAATGCGAAGATTCTGG                                                             |
| Characterisation of homeodomain mutants      |Dll_sgrna12_R1   | cagtcctgCTGCGTTTGTGTAGGCCCCA                                                            |
| Characterisation of homeodomain mutants      |Dll_sgrna12_R2   | catgggaCTGCGTTTGTGTAGGCCCCA                                                            |
| Characterisation of homeodomain mutants      |Dll_sgrna12_R3   | tcacgtaCTGCGTTTGTGTAGGCCCCA                                                            |
| Characterisation of homeodomain mutants      |Dll_sgrna12_R4   | tggtccaCTGCGTTTGTGTAGGCCCCA                                                            |
| Characterisation of homeodomain mutants      |Dll_sgrna12_R5   | aacctgtCTGCGTTTGTGTAGGCCCCA                                                            |
| Characterisation of homeodomain mutants      |Dll_sgrna12_R6   | cttggttCTGCGTTTGTGTAGGCCCCA                                                            |
| Characterisation of Exon 2 mutants           |Dll_sgrna4_F1    | aacctgtTACGCGTGCGCATTCGCAGCGGT TATC                                                     |
| Characterisation of Exon 2 mutants           |Dll_sgrna4_F2    | agaagtgTACGCGTGCGCATTCGCAGCGGT TATC                                                     |
| Characterisation of Exon 2 mutants           |Dll_sgrna4_F3    | cgggtatTACGCGTGCGCATTCGCAGCGGT TATC                                                     |
| Characterisation of Exon 2 mutants           |Dll_sgrna4_F4    | gtacactTACGCGTGCGCATTCGCAGCGGT TATC                                                     |
| Characterisation of Exon 2 mutants           |Dll_sgrna4_F5    | gtgatgaTACGCGTGCGCATTCGCAGCGGT TATC                                                     |
| Characterisation of Exon 2 mutants           |Dll_sgrna4_F6    | cttggttTACGCGTGCGCATTCGCAGCGGT TATC                                                     |
| Characterisation of Exon 2 mutants           |Dll_sgrna4_R1    | aacctgtGGCGCTCGACATGGCGGG                                                               |
| Characterisation of Exon 2 mutants           |Dll_sgrna4_R2    | agaagtgGGCGCTCGACATGGCGGG                                                               |
| Characterisation of Exon 2 mutants           |Dll_sgrna4_R3    | cgggtatGGCGCTCGACATGGCGGG                                                               |
| Characterisation of Exon 2 mutants           |Dll_sgrna4_R4    | gtacactGGCGCTCGACATGGCGGG                                                               |
| Characterisation of Exon 2 mutants | Dll_sgrna4_R5 | tgtgtGCGCTCGACATGGGCGGG |
|-----------------------------------|---------------|--------------------------|
| Characterisation of Exon 2 mutants | Dll_sgrna4_R6 | cttgtGCGCTCGACATGGGCGGG |
| Characterisation of ESE mutants | Dll_intron1_F1 | aacctgtGCGCTCGACATGGGCGGG |
| Characterisation of ESE mutants | Dll_intron1_F2 | aagatgtGCGCTCGACATGGGCGGG |
| Characterisation of ESE mutants | Dll_intron1_F3 | cggatatGCGCTCGACATGGGCGGG |
| Characterisation of ESE mutants | Dll_intron1_F4 | gtacactGCGCTCGACATGGGCGGG |
| Characterisation of ESE mutants | Dll_intron1_F5 | tgtgtGCGCTCGACATGGGCGGG |
| Characterisation of ESE mutants | Dll_intron1_F6 | cttgtGCGCTCGACATGGGCGGG |
| Characterisation of ESE mutants | Dll_sgrna4_R1 | aacctgtGCGCTCGACATGGGCGGG |
| Characterisation of ESE mutants | Dll_sgrna4_R2 | aagatgtGCGCTCGACATGGGCGGG |
| Characterisation of ESE mutants | Dll_sgrna4_R3 | cggatatGCGCTCGACATGGGCGGG |
| Characterisation of ESE mutants | Dll_sgrna4_R4 | gtacactGCGCTCGACATGGGCGGG |
| Characterisation of ESE mutants | Dll_sgrna4_R5 | tgtgtGCGCTCGACATGGGCGGG |
| Characterisation of ESE mutants | Dll_sgrna4_R6 | cttgtGCGCTCGACATGGGCGGG |

| Amplification of Dll in Drosophila melanogaster | Dmel_Dll_F2 | CCGATAAGTGCGAGGACTCCGG |
|-----------------------------------------------|-------------|------------------------|
| Amplification of Dll in Drosophila melanogaster | Dmel_Dll_R | CTGCCTTTGCGAGGACTCCGG |
| Amplification of Abd-B in Drosophila melanogaster | Dmel_ABDB_442_F | CCCACCTACTCCTCGCCAGGCGGG |
| Amplification of Abd-B in Drosophila melanogaster | Dmel_ABDB_571_R | TCCACTCGTCAGTCGAGGATTG |
| Shortened reverse primer with lower Tm for cycle sequencing of Drosophila melanogaster Abd-B amplicon | Dmel_ABDB_571_SEQ_R | TCCACTCGTCAGTCGAGGATT |
| Amplification of Dll in Themira biloba | DLL Exon 2 sgRNA4 F | TAC GGT GGC ATT CGC AGC GGT TAT C |
| Amplification of Dll in Themira biloba | DLL Exon 2 sgRNA4 R | CAC ATG GCG GGG CAT AAC TGC CTA AAT G |
Amplification of Dll for isoform sequencing
DLL 5'UTR Pacbio F
/5Phos/CAC GCT TAA TTT CAC AGA TTT AGG GAG CCC C

Amplification of Dll for isoform sequencing
DLL exon 7 Pacbio R1
/5Phos/tca gac gat ggc tca tAG GTG GAG GTG GTA ATT GCG GCG AAT GG

Amplification of Dll for isoform sequencing
DLL exon 7 Pacbio R2
/5Phos/tct att agt act ctg cAG GTG GAG GTG GTA ATT GCG GCG AAT GG

Amplification of Dll for isoform sequencing
DLL exon 7 Pacbio R3
/5Phos/tct att atg act ctg cAG GTG GAG GTG GTA ATT GCG GCG AAT GG

Amplification of Dll for isoform sequencing
DLL exon 7 Pacbio R4
/5Phos/tgt gta tca gta cat gAG GTG GAG GTG GTA ATT GCG GCG AAT GG

Generating exon-skipped template for protein in vitro synthesis
Tbil Dll start codon F
GGCTTAAGTATAAGGAGGAAAAAAT ATGGATGCGCCCGATGACCGCATA

Generating exon-skipped template for protein in vitro synthesis
Tbil Dll stop codon R
AAACCCCTCCGTTTAGAGAGGGTT ATGCTAGTTATTAGTGTTGTTGTTGTGGTG GTGTGAACCTGCCGACACCGTTT AATAGCGATGG

Supplementary Table 5. Sanger sequencing results for the RT-PCR to determine Distal-less expression in developing histoblast clusters. See separate excel document. Related to Figure 7.

Supplementary Table 6. Target region sequences for single guide RNA for CRISPR/Cas9 genome editing. Related to Figure 2, 3, 4 and 7.

| sgRNA   | Target sequence (5' → 3')          |
|---------|-----------------------------------|
| sgRNA-2A| GGTACCAGTGGCATTCCAGCCG            |
| sgRNA-2B| GGAATAGGAAATTCTAGTCAGGG           |
| sgRNA-3 | GGTTTCAACGCACCCAGTTT             |
Supplementary Figure 1. Deletions in Dll Exon 2 produce mosaic mutant (E2B M8) with loss of the 4th sternite brush. Related to Figure 4.
Supplementary Figure 2. Alignment of Distal-less protein sequences across holometabola. Alignment in fasta format is provided in the Mendeley resource. Related to Figure 4.
Supplementary Figure 3. Fragmentation evidence for Dll peptides recovered from ProteinPilot. (a) Fragment ion mass values and peptide peak intensity for ‘QSTWMAAQR’, a peptide matching to an alternative initiation codon (b), (c), (d) and (e) Fragment ion mass values and peptide peak intensity for peptides matching to Dll homeodomain. Related to Figure 6.
Supplementary Figure 4. RT-PCR results to determine Dll expression in late larval epidermal tissues. *Drosophila melanogaster* was used as a control. Primers for this experiment are found in Supplementary Table 5. Identity of amplified products were confirmed with sanger sequencing (results in Supplementary Table 6). A1-A8: Tissue from the respective abdominal epidermal segments. T: Tissue from the thoracic epidermal segment. (a) In the control, *D.melanogaster*, Abd-B is present and amplified in the 5th – 8th abdominal segments as expected. (b) Dll is present and amplified in all 8 abdominal segments in *D.melanogaster* (c) Dll is present and amplified in all 7 abdominal segments in *Themira biloba*. Related to Figure 7.
Supplementary Figure 5. Read length distribution of Pacbio Isoseq error-corrected reads (10 minimum passes with a minimum predicted accuracy of 0.9) using the PacBio SMRT analysis software (v5.1.0). The analysis generated 225,740 CCS reads with a median CCS read length between 1200 – 1249 bp. Related to Figure 8.
