Morphological evolution caused by many subtle-effect substitutions in regulatory DNA

Nicolás Frankel1*, Deniz F. Erezylmaz1*, Alistair P. McGregor2, Shu Wang1, François Payre3 & David L. Stern1

Morphology evolves often through changes in developmental genes, but the causal mutations, and their effects, remain largely unknown. The evolution of naked cuticle on larvae of Drosophila sechellia resulted from changes in five transcriptional enhancers of shavenbaby (svb), a transcript of the ovo locus that encodes a transcription factor that governs morphogenesis of microtrichiae, hereafter called ‘trichomes’. Here we show that the function of one of these enhancers evolved through multiple single-nucleotide substitutions that altered both the timing and level of svb expression. The consequences of these nucleotide substitutions on larval morphology were quantified with a novel functional assay. We found that each substitution had a relatively small phenotypic effect, and that many nucleotide changes account for this large morphological difference. In addition, we observed that the substitutions had non-additive effects. These data provide unprecedented resolution of the phenotypic effects of substitutions and show how individual nucleotide changes in a transcriptional enhancer have caused morphological evolution.

The genetic mechanisms underlying morphological evolution remain largely unknown1-5. Comparative studies indicate that changes in the timing (heterochrony), location (heterotopy), and level of gene expression have caused much of morphological evolution6-8. But, with a few exceptions9,10, we do not know the specific DNA changes responsible for altered expression, leaving several important questions unanswered. First, it is unclear how many genetic changes underlie new morphologies11. Second, we do not know whether multiple substitutions have independent effects or if they contribute instead to epistasis, where the effects of one change are dependent on other changes12-15. Third, it has been predicted that the changes that cause morphological evolution have minimal pleiotropic effects16-18. Last, we do not know how often transcriptional regulation evolves through deletion and de novo creation of enhancers as opposed to subtle modification of existing cis-regulatory modules19-21.

Here we identify the molecular changes in a transcriptional enhancer underlying a case of morphological evolution. To shed light on the interplay between gene expression divergence and morphological evolution, we evaluated the effects of these changes on timing and level of expression and also determined their effects on the resulting phenotype.

Modular enhancers regulate svb transcription

Drosophila melanogaster larvae are decorated with a complex pattern of trichomes that results from the differentiation of epidermal cells (Fig. 1a, b). We focus on the dorsolateral epidermis, which differentiates quaternary trichomes in D. melanogaster (Fig. 1b) and in most related species22. Evolution of cis-regulatory regions of the svb gene, which encodes a transcription factor that orchestrates trichome morphogenesis23,24, cause D. sechellia larvae to differentiate smooth cuticle, rather than quaternary trichomes25 (Fig. 1c). This derived phenotype resulted from the specific loss of svb expression in quaternary cells (Fig. 1d, e), while svb expression is conserved in other epidermal cells, such as those that produce the ventral stout trichomes, called denticles25.

Through systematic dissection of the ~110-kb D. melanogaster svb locus, we identified six embryonic enhancers of ~5 kb25-26 (Fig. 1f). In D. sechellia, five of these six enhancers have evolved reduced activity in quaternary cells27,28. One of these enhancers, E, drives strong expression in quaternary cells and in the ventral denticle cells of D. melanogaster embryos29. The orthologous E region from D. sechellia drives greatly diminished expression in quaternary cells, which directly contributed to trichome pattern evolution30, while expression driven by this enhancer in ventral cells is conserved30. The E cis-regulatory element thus represents an attractive target for identifying the individual genetic changes that have contributed to morphological evolution in D. sechellia.

We found that the ventral and dorsolateral expression driven by E are encoded in two distinct regions—each ~1 kb in length—that are separated by a 12-kb (Fig. 1g and Supplementary Fig. 1). The first region, E3, drives expression in ventral cells that differentiate denticles (Fig. 1h) and the second region, E6, drives mostly dorsolateral expression (Fig. 1i). No smaller constructs from the E6 region showed equivalent activity; E6 sub-fragments drove expression that was either strongly reduced, partial, or ectopic (Supplementary Fig. 1). The D. melanogaster E region thus comprises two cis-regulatory modules: E3, which drives expression in ventral cells, and E6, the minimal region that can drive a coherent pattern of expression in quaternary cells.

A svb enhancer evolved by level and timing changes

To assay the evolutionary modification of E activity between D. melanogaster and D. sechellia, for each species we generated E10 constructs, which included both the evolving E6 region and the conserved E3 region. The E3 region provided an internal control of conserved expression (Fig. 2e, f). The D. melanogaster E10 construct (mel_E10) drove expression in dorsal cells beginning at stage 12-13 (Fig. 2a, c).
Figure 1 | The pattern of trichomes has evolved between Drosophila species owing to changes in the enhancers of the svb gene. a, Lateral view drawing of a first instar larva of D. melanogaster. The rectangle indicates the region shown in b and c. b, c, The pattern of dorsolateral trichomes on the fourth abdominal segment of D. melanogaster (b) and D. sechellia (c). Some of the dorsolateral cells differentiate thin "quaternary" trichomes in D. melanogaster and naked cuticle in D. sechellia. d, e, Pattern of svb RNA expression in stage 14 embryos of D. melanogaster (d) and D. sechellia (e). f, Diagram illustrating the location of the six enhancers of svb (open boxes). The enhancers 7, E and A were referred to as proximal, medial and distal, respectively, in ref. 25. Genes in the region are indicated with grey boxes and only the first exon of svb is shown. g, Summary of the dissection of the E enhancer in D. melanogaster. Boxes indicate the enhancer constructs discussed in the text. h, The E5 region drives expression in ventral stripes. i, The E6 region drives expression in quaternary cells.

This pattern strengthened and spread to more lateral cells in later stages (Fig. 2e, g). In stage 16 embryos, mel_E10 expression persisted in many dorsal and lateral cells (Fig. 2l), while endogenous svb messenger RNA is not present at this stage (data not shown). These constructs therefore produce artificially high levels of mRNA in late-stage embryos. This experimental artefact allowed discovery of the surprising fact that, whereas the D. sechellia E10 (sec_E10) does not drive expression before stage 14 (Fig. 2h, d, f), it does drive expression in quaternary cells in late-stage embryos (Fig. 2h, j), albeit at a much lower level than does mel_E10. In a separate set of experiments, we confirmed that the D. sechellia E6 region indeed drives this late dorsal expression (data not shown), indicating that it retains some weak and heterochronic expression. In contrast, the ventral expression driven by sec_E10 matched the timing and levels driven by mel_E10. These data therefore show that conserved ventral expression and divergent dorsal expression of the E10 regions from D. melanogaster and D. sechellia is correlated with the patterns of trichomes produced by each species, further localizing evolutionary changes to within the E6 region.

The E6 enhancer evolved at an accelerated rate

Next we attempted to identify the DNA changes that caused the evolutionary shift in E6 function. We compared the sequences of the E6 region between D. sechellia and five closely related species, all of which, like D. melanogaster, produce a dense lawn of quaternary trichomes. Multiple sequence alignment allowed us to identify thirteen substitutions and one single-nucleotide deletion that are unique to D. sechellia (Fig. 3a, b and Supplementary Fig. 2). These D. sechellia-specific substitutions are located in a region of ~500 bp (the 'local region') of otherwise high sequence conservation, even in D. sechellia (Fig. 3a).

Given the functional importance of E6, we examined whether this apparent clustering of substitutions within a highly conserved block represented an unusual substitution rate. We sequenced the E6 focal region from eight additional isolates of D. sechellia. All nine D. sechellia sequences were identical (data not shown), which is consistent with the low levels of polymorphism detected in other regions of the D. sechellia genome.25,26. The absence of polymorphism in the E6 region in D. sechellia prevented us from using common tests of selection that rely on allele frequencies. Instead, we analysed substitution rates in the D. sechellia and D. simulans lineages, using D. melanogaster as an outgroup.29. We observed a significant increase in D. sechellia divergence, compared to D. simulans, in the focal region of E6 (Fig. 3c; Tajima's relative rate test, $\chi^2 = 6.25$, $P = 0.012$, 503 alignable bases). To determine whether this pattern of accelerated divergence reflects simply an accelerated evolutionary rate of substitution at this genomic locus in D. sechellia, we sequenced ~9,000 bp of DNA flanking the focal region, which does not include any of the other evolved enhancers, both from D. sechellia and from D. simulans. The ~9,000 bp region has not evolved at significantly different rates in the two lineages (Fig. 3d; Tajima's relative rate test, $\chi^2 = 0.56$, $P = 0.45$, 7,072 alignable
Figure 3 | Sequence conservation of the E6 region and location of the D. sechellia-specific substitutions. A, The aligned E6 sequences from D. melanogaster, D. simulans, D. mauritiana, D. sechellia, D. yakuba and D. erecta are represented as thick horizontal lines, with thin regions indicating gaps in the alignments. (Full alignment is provided in Supplementary Fig. 2.) Sequence conservation over a 10-bp sliding window is represented above by the height of the grey bars. The positions of D. sechellia-specific substitutions are indicated with vertical red lines, the seven clusters of substitutions are indicated below the red lines, and the 'focal region' is labelled. B, Sequences of the seven regions containing the D. sechellia-specific substitutions (enclosed in rectangles) with the aligned sequences from D. melanogaster (mel), D. simulans (sim) and D. sechellia (sec). C, D, Evolutionary trees of the E6 focal region (c) and 9 kb outside of the focal region (d), where branch lengths are proportional to the substitution rate. Numbers indicate number of substitutions per site on average.

In the D. sechellia lineage, the focal region experienced a significantly higher substitution rate (4.8 times higher) than did the flanking regions (Fisher's exact test, two-tailed P = 0.016). Therefore, when compared to neighbouring regions, the focal region of E6 evolved at a faster rate in the D. sechellia lineage, indicating that it has evolved under positive selection, or relaxed constraints, or both.

Substitutions in E6 altered enhancer function

To assess the effect of the D. sechellia-specific substitutions in E6 on enhancer activity, we introduced all of these substitutions into mel_E10. We also performed the reciprocal experiment by reversing the D. sechellia-specific substitutions to the D. melanogaster sequence in sec_E10. To enable trihome rescue experiments, the mutated E10 versions were placed upstream of a svb complementary DNA that contained a heterologous tag in the 3' untranslated region (UTR), which allowed us to differentiate expression driven by the transgene from expression driven by the endogenous svb gene.

In stage 14 embryos, the D. melanogaster E10 construct carrying all of the D. sechellia-specific substitutions (mel_mut_All) drove substantially weaker expression in the quaternary cells than did mel_E10 (Fig. 4a, c). Conversely, the D. sechellia E10 carrying all of the 'reverse' substitutions to the D. melanogaster state (sec_mut_All) drove substantially stronger dorsal expression than did sec_E10 (Fig. 4b, d). These manipulated enhancers did not perfectly reproduce the temporal and spatial differences between mel_E10 and sec_E10 (Fig. 4), indicating that at least one other substitution in E10 contributed to the functional divergence of these enhancers. All together, these results confirm that at least one of the D. sechellia-specific substitutions in the E6 region caused most of the species difference in E6 function.

Many substitutions caused morphological evolution

Next we asked which of the D. sechellia-specific substitutions caused the altered function of E6 in D. sechellia. As the D. sechellia-specific substitutions in the E6 enhancer appeared to be clustered in seven regions (Fig. 3a), we mutated separately these seven clusters of nucleotides (Fig. 3b) from the D. melanogaster to the D. sechellia sequence in mel_E10. We also performed the reverse experiment, separately mutating each of seven clusters from the D. sechellia to the D. melanogaster sequence in sec_E10. Some of the D. melanogaster constructs with individual mutated clusters showed weaker lateral expression in stage 14 embryos than mel_E10 did (data not shown). Quantification of the
onset of expression revealed further that five of seven of the D. melanogaster mutated enhancers drove significantly delayed expression when compared to mel_E10 (Fig. 4a and Supplementary Table 1). In the reciprocal experiments, some sec_E10 constructs with clusters of D. melanogaster substitutions drove slightly stronger dorsolateral expression in quaternary cells than did sec_E10 (data not shown). Some of these sec_mut constructs drove a significantly altered onset of expression than did sec_E10, but these differences were not of large magnitude (Fig. 4e and Supplementary Table 1). Most importantly, no single cluster of substitutions in either direction recapitulated the temporal onset of expression observed when all substitutions were introduced together (Fig. 4e).

These results indicate that at least five of the D. sechellia-specific substitutions in the E6 region contributed to the functional divergence of this enhancer. We therefore quantified the ability of these constructs to rescue trichomes in an embryo that lacked endogenous svb activity (Fig. 5). We tested first whether mel_E10 and sec_E10 could rescue the production of trichomes with normal morphology in the correct spatial domains (Fig. 5a–c, i). mel_E10 rescued many, but not all, of the quaternary trichomes (Fig. 5c, m, n) and recovered many ventral trichomes (Supplementary Fig. 3). The incomplete rescue of both dorsal and ventral trichomes was expected, because multiple svb enhancers together contribute to the complete pattern of svb expression16. sec_E10 rescued ventral trichomes as well as mel_E10 but (Supplementary Fig. 3), but recovered only a few dorsal trichomes (Fig. 5i, m), consistent with the conserved and evolved functions of E10. Therefore, this rescue assay provides a reliable readout of the normal function of svb enhancers.

Because the D. sechellia-specific substitutions in E6 are sufficient to almost completely recapitulate the differences in expression patterns between the species, we asked whether these changes were sufficient to modify trichome patterning. Introduction of all of the D. sechellia-specific substitutions from E6 into mel_E10, mel_mut_All, caused larvae to produce many fewer trichomes than did mel_E10, and thus to look more like D. sechellia (Fig. 5d, m, n). Conversely, larvae carrying the reversed substitutions in a D. sechellia background (sec_mut_All) looked more like D. melanogaster larvae (Fig. 5i, m).

To determine how many substitutions cause this species difference in enhancer activity, we tested whether each cluster of substitutions influenced trichome patterns. In mel_mut_2, mel_mut_3, mel_mut_4 and mel_mut_5, the D. melanogaster to D. sechellia substitutions reduced the number of trichomes produced by 4.6–33.5% (Fig. 5e–h, m, n and Supplementary Table 3). In contrast, in only sec_mut_2 and sec_mut_3 did the D. sechellia to D. melanogaster substitutions increase the number of trichomes by 9.9–14.6% (Fig. 5j, k, m, n and Supplementary Table 3).

Larvae carrying mel_mut_All differentiated significantly more trichomes than did larvae carrying sec_E10. The opposite is also true; sec_mut_All did not rescue as many trichomes as did mel_E10. Thus, additional substitutions within E10, other than those we tested, might also have contributed to the morphological difference between D. melanogaster and D. sechellia.

The functional rescue experiments show that at least four clusters of substitutions in E6 can alter trichome patterning on their own. Both the onset of expression data and the trichome rescue data indicate that the D. sechellia-specific substitutions show epistasis with respect to each other and with respect to the remaining E10 sequence. Indeed, the magnitude of the effect of mutating all seven clusters of substitutions together on trichome patterning is not recapitulated by summing up the effects of all clusters acting alone (Fig. 5m, n and Supplementary Table 3). The impact of each substitution on larval morphology is thus partly dependent on which other substitutions are already present.

Figure 5 | Effect of the engineered substitutions on trichome rescue in dorsal and lateral regions of the sixth abdominal segment of first instar larvae. a. Wild-type D. melanogaster. b. svb null. c. mel_E10 in a svb null background. The dorsal (D) and lateral (L) regions where trichomes were counted are delimited with a dashed line. d–h, mel_E10 constructs carrying all D. sechellia substitutions (d), or cluster 2 (e), 3 (f), 4 (g), or 5 (h) substitutions in a svb null background. i–l, mel_E10 in a svb null background. j–l, sec_E10 constructs carrying cluster 2 (e), 3 (f), or all D. melanogaster substitutions (l). m, n. Number of trichomes rescued by the mel (black) and sec (red) constructs in the dorsal (m) and lateral (n) regions. All larvae carrying sec_mut constructs differentiated zero trichomes in the lateral region, and for clarity these data are not shown in n. Open circles represent counts for each individual. Closed circles and lines indicate the means and standard deviations, respectively. Grey shading encompasses the constructs with trichome counts that were significantly different from the E10 construct of the respective species (P < 0.05, Dunnet’s test).
Note that there is not perfect congruence between the analysis of gene expression patterns and the functional readout of trichome number. For example, *mei_mat_6* altered expression timing, but not trichome number. This suggests that subtle expression differences may not always correctly predict the effects of genetic changes on morphological evolution.

**Discussion**

We have identified molecular changes in a *cis*-regulatory region that contributed to a morphological difference between closely related species. We found that, taken individually, each genetic change in a transcriptional enhancer had a relatively small effect on gene expression and on the final phenotype, but that when they were combined, they produced a large morphological difference. The substitutions that contributed to morphological evolution showed substantial epistasis, both with respect to the background *E10* construct and with respect to the other substitutions in *E6*. It is impossible to know the actual order in which these substitutions occurred nor whether all of the mutations went to fixation independently or whether some segregated. Thus we focused on the effects of individual clusters of substitutions in the background of the parental species.

Our results indicate that at least five substitutions in the *E10* region—at least four in the mutated clusters and at least one other site—contributed to altered function of the *E6* enhancer in *D. sechellia*. Similarly, a study of pigmentation differences among *D. melanogaster* populations showed that multiple polymorphisms of small effect in enhancers of the gene *ebony* account for large phenotypic differences. We propose that enhancer structure influences the patterns of genetic change. When the function of a *cis*-regulatory module relies on multiple transcription factor-binding sites, each with a small effect on expression, evolution may require changes of a large number of such sites to cause a significant phenotypic change.

Detecting the action of natural selection on specific non-coding genomic regions remains a major challenge for evolutionary genetics. The accelerated substitution rate that we observed in the *D. sechellia* *E6* focal region indicates that this region experienced either positive selection, relaxation of purifying selection, or both. In addition, none of the *D. melanogaster* to *D. sechellia* mutations led to a significant increase in trichome number, and none of the reciprocal mutations led to a decrease in trichome number. That is, along the lineage leading to *D. sechellia*, the *E6* enhancer seems to have accumulated only substitutions that decrease trichome number. These observations also are consistent with the action of directional selection, unless random mutations in this enhancer preferentially cause loss of expression.

When we reverted the *D. sechellia*-specific substitutions to the ancestral state, the *D. sechellia* *E10* construct regained most of the functionality present in the *D. melanogaster* *E10* construct. Thus, in principle, descendants of modern *D. sechellia* could re-evolve at least some trichomes through the accumulation of single-nucleotide substitutions in an existing enhancer. Our results contrast with other recent studies of cis-regulatory evolution that have discovered large deletions in transcriptional enhancers. For example, the wholesale deletion of an enhancer caused the loss of pelvic structures in some stickleback populations. Although this is a striking result, large deletions may contribute to morphological evolution only rarely. For example, enhancer deletions may have deleterious pleiotropic effects, because many single enhancer ‘modules’ in fact encode expression in multiple domains. In addition, new expression patterns may sometimes evolve through modification of existing enhancers. Widespread deletion of *cis*-regulatory DNA may thus reduce the evolutionary potential of existing enhancers. It is worth noting that the stickleback populations with different pelvic structures diverged less than 10,000 years ago. Our study focuses on morphological differences between species that diverged approximately 500,000 years ago. The markedly different genetic architecture discovered in these two cases may indicate that different kinds of mutations are selected over different evolutionary timescales.

Our results suggest an additional explanation for the predominance of single-nucleotide substitutions that have altered *E6* function. Some constructs carrying large deletions of the *E6* element generated ectopic expression (Supplementary Fig. 1). This may be a general feature of enhancers that require multiple activation and repressive activities to define a precise spatiotemporal pattern of expression. In such cases, large insertions or deletions may result in ectopic expression and, potentially, in dominant pleiotropic effects. In contrast, single-nucleotide substitutions within activator and repressor binding sites may result in subtle changes in expression with minimal pleiotropic effects. For example, substitutions that lead to heterochrony in enhancer activity can modify a transcriptional program without deleterious effects on development. Such a heterochronic shift in enhancer activity could result from either downregulation of enhancer activity or from a temporal delay in the initiation of enhancer activation. Either or both kinds of events may have occurred in the *D. sechellia* lineage.

**METHODS SUMMARY**

Embryos were collected and fixed using standard conditions and β-galactosidase expression was detected with immunohistochemistry using a rabbit anti-β-galactosidase antibody (Cappel) used at 1:2,000 and an anti-rabbit antibody coupled to HRP (Santa Cruz Biotech), also used at 1:2,000. Staining was developed with DAB/Nickel.

To detect the expression of transgenic *svb* transcripts, we made a RNA probe complementary to the go44 and go64 sequence in the 3' UTR of the *svb* CDNA using the Dig RNA labelling kit (Roche). We tested for heterochronic changes in the onset of transgene expression by comparing the proportion of embryos showing staining between constructs at a single stage. We then tested for differences in the proportions of stained embryos with the Barnard test using a sequential Bonferroni correction for multiple tests.

For trichome rescue experiments, we cloned *D. melanogaster* and *D. sechellia* *E10* into pRSQeBv. Mutant plasmids were generated using site-directed mutagenesis (Genescraft USA). Constructs were integrated into the ATP site of line M(3-3-3-3-RFP:ap5)Z1H- 86Fb; M(vas-int-Dm)Z1H-2A. Males homozygous for the transgene were crossed to svbv* /FMR correction/GFP females. Non-fluorescent first larval larvae from this cross were mounted on a microscope slide in a drop of Hoyer's liquid acid (1:1). Cleared cuticles were imaged with phase-contrast microscopy. Dorsal and lateral regions were defined using morphological landmarks and programmed as macro in ImageJ software. Trichomes were counted using the cell-counter option of ImageJ. We performed pairwise comparisons of trichome numbers between the wild-type construct and each mutant construct and the statistical significance of comparisons was determined with Dunn's test. Full details of methods used can be found in Supplementary Information.

Received 4 February; accepted 13 May 2011.


Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank G. Davis, P. Parish and P. Valenti for assistance with cloning and the Drosophila Species Stock Center for fly stocks. This work was supported by the Pew Charitable Trusts Latin American Fellows Program in the Biomedical Sciences Fellowship to N.F., a Ruth L. Kirschstein National Research Service Award to D.F.E. (F32 GM 83546-02), Agence Nationale de la Recherche (Blanc 2008, Netoshop) to F.P., and NIH (GM063622-06A1) and NSF (IOS-0640339) grants to D.L.S.

Author Contributions N.F., D.F.E., A.P.M. and D.L.S. designed the experiments and analysed the data. N.F., D.F.E., A.P.M., S.W. and F.P. performed the experimental work. N.F. and D.L.S. wrote the manuscript. D.F.E., A.P.M. and F.P. commented on the manuscript at all stages.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to D.L.S. (dstern@princeton.edu).