PLoS GENETICS

Perspective

A Closer Look at the *eve* Stripe 2 Enhancers of *Drosophila* and *Themira*

Justin Crocker, Albert Erives*

Department of Biological Sciences, Dartmouth College, Hanover, New Hampshire, United States of America

Introduction

Gene regulatory sequences have been investigated and/or proposed to be important targets of natural selection during animal evolution [1–14]. However, much controversy has been generated by the contention that they are not likely to be as important as functional protein-coding evolution given the low number of such examples established to date [15,16]. However, an important obstacle in identifying such sequences is our lack of understanding of the organizational basis for such sequences. Such an understanding could allow the rapid identification and annotation of gene regulatory functions in sequenced genomes.

Gene regulatory sequences function by displaying clusters of sites for DNA sequence-specific binding factors. Such clusters are called *cis*-regulatory modules (CRMs), of which the transcriptional enhancers constitute a large and important class. The degree to which the constituent binding elements of enhancers are necessarily organized by position, orientation, and relative spacing in order to function will dictate the constraints governing enhancer evolution. Thus, the internal functional organization of enhancers is important for understanding the mode and tempo of gene regulatory evolution as well as for deciphering and annotating genomic sequences.

Arguably, no other metazoan *cis*-regulatory module has yet been as genetically and biochemically defined as the *even-skipped* (*eve*) stripe 2 enhancer module of *Drosophila melanogaster* [17–23]. For this reason, this module has been intensely studied from a phylogenetic perspective amongst drosophilids [24–27]. These phylogenetic studies of the *eve* stripe 2 enhancer have now been extended into *Themira*, a sepsid fly [28]. This latest study is used to make a central claim that a lack of sequence conservation between the *eve* stripe 2 enhancers of *Drosophila melanogaster* and *Themira putris* suggests that "complex animal regulatory sequences can tolerate nearly complete rearrangement of their transcription factor binding sites". Thus, this study is being interpreted to reach

conclusions addressing an important ongoing debate on the degree of functional organization of enhancers [29]. The results of this debate therefore impact the much larger discussion on the genetic loci of evolution [15,16].

Both *Drosophila* and *Themira* are acalyptrate flies and last shared a common ancestor at least ~110 Mya, and so it is suggested that this distance is long enough for the sequences to be completely scrambled in divergent organisms still sharing a similar embryonic patterning system. While the sepsid study presents an informative taxonomic collection of an evolving enhancer, this study falls short in critically testing the claim that enhancer organization is not important. Moreover, here we report that we find extensive homology in what is claimed to be an exemplar of scrambled enhancer sequences. As explained below, these ordered blocks of homology encompass well-known activator and repressor binding sites. Thus, the organization of Acalyptratae *eve* stripe 2 enhancers has not diverged enough to rule out organized assembly of higher-order enhancesome complexes at these sequences.

Extensive Homology in the *eve* Stripe 2 Enhancers of *Drosophila* and *Themira*

We first began by graphing the *Themira* and *Drosophila* stripe 2 enhancer sequences on two-dimensional sequence alignment plots (Figure 1). Such a dot plot or graphic matrix shows all regions of similarity between two sequences [30]. Such an alignment is helpful for visualizing possible insertions, deletions, rearrangements, inversions, repeats, and overall homology, without being constrained by global align-

ments. We also computed the same dot plot using the reverse complement of one of the sequences (Figure 1B and 1E). In addition to showing similar sequences that happen to occur in the opposite orientation, graphing the reverse complement serves as an internal negative control for conservation of serial blocks of sequence. Here, we report that when we graph the *eve* stripe 2 enhancers in parallel orientations, we see large blocks of alignment spanning ~600 bp, almost the entire length of the enhancer (Figure 1A). These blocks are larger and more numerous compared to the number and types of alignable blocks achieved when we align them in anti-parallel orientation, i.e., when we plot against the reverse complement of one of the sequences (compare Figure 1A and 1D versus 1B and 1E, or see score distributions in 1C and 1F, respectively). We made such plots for two different thresholds that correspond to an ~14 bp length of alignment that would encompass most binding sites (Figure 1A–1C) as well as a more extensive ~20 bp length of alignment (Figure 1D–1F). At the more stringent level, most of the alignments in the anti-parallel direction are lost (Figure 1E). However, a clear identity line of ordered blocks of conservation is visible in the parallel alignment (Figure 1D). Thus, there exists ordered blocks of highly conserved sequence of a length consistent with multiple binding sites spanning the length of the enhancer.

The *Drosophila/Themira* study of an embryonic enhancer of the anterior posterior (A/P) axis could have been better informed by considering the *Drosophila/Anopheles* study of an embryonic enhancer of the dorsal/ventral axis (D/V) [31]. This

* E-mail: Albert.Erives@Dartmouth.edu
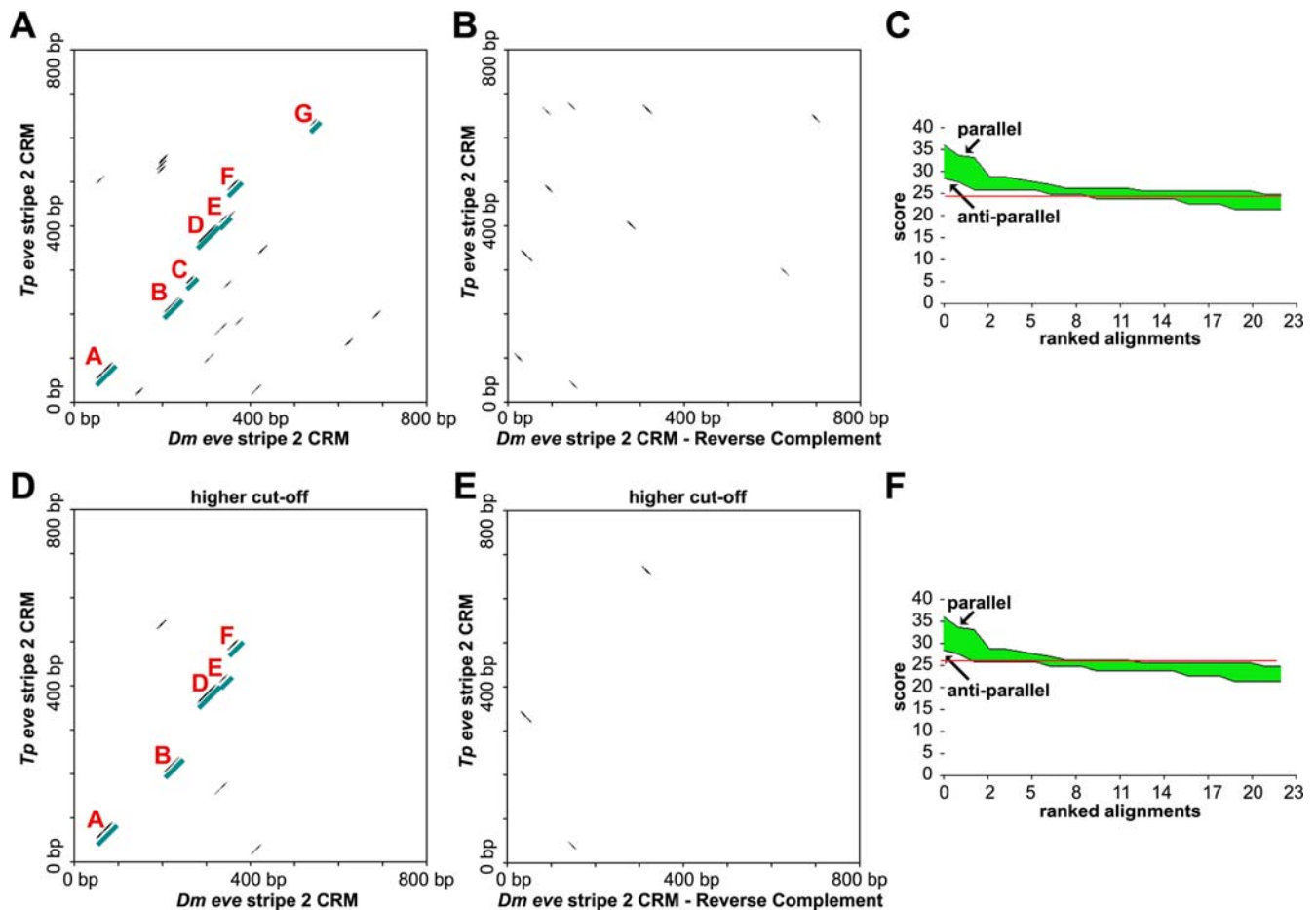
Figure 1. Two-dimensional dot plots of the *eve* stripe 2 enhancers of *Drosophila* and *Themira*. (A) A two-dimensional dot plot for parallel orientations of a pair of *Drosophila* and *Themira eve* stripe 2 enhancers, which have been diverging for at least ~110 My, shows extensive blocks of conservation that are maintained in the same serial order in each species (see blocks labeled A–G in red). (B) Shown is the anti-parallel orientations of the same sequences as in (A) except that the reverse complement of one of the sequences is used in order to find additional, possibly compensatory, sites that may have changed in their orientation. Plotting the anti-parallel alignments also serves as an internal negative control. (C) A plot of the ranked alignment scores is shown for both the parallel and anti-parallel enhancer pair orientations. The score corresponds to the number of nucleotides of perfect identity within the un-gapped block of alignment. This plot shows that there are more extensive blocks of alignment in the parallel orientation than in the anti-parallel orientation. This is consistent with basic conservation of the entire enhancer. The red line indicates the threshold used for plotting points in (A) and (B). (D–F). Same as (A–C) except a higher or more stringent threshold is used. Note that a broken identity line of highly conserved blocks is easily seen in the parallel enhancer orientations (D), while most of the blocks of alignment seen in (B) disappear in the anti-parallel orientation at this stringency (E).
doi:10.1371/journal.pgen.1000276.g001

study analyzed homologous *vnd* neuroectodermal enhancers from both *Drosophila* and the mosquito *Anopheles*, which last shared a common ancestor at least ~250 Mya (Figure 2A). This study shows that core *cis*-elements are organized in a similar architectural plan (Figure 2B). Moreover, this conserved organization was present in *non-homologous* neuroectodermal enhancers that had evolved in parallel at other loci [31,32]. However, the *Drosophila* and *Anopheles vnd* enhancers are

so scrambled that it is difficult to find any evidence of serial sequence homology unlike the *Drosophila/Themira* pair (Figure 2C and 2D). This is consistent with the additional ~140 My of divergence between Acalyptratae and mosquitoes on top of the ~110 My of divergence between the *Drosophila* and *Themira* (Figure 2A).

The lesson in the mosquito example that should have informed the sepsid *eve* stripe 2 study is that the absence of

extensive sequence homology is not indicative of the absence of conserved organization of binding sites. Therefore, a simple claim that an enhancer is scrambled is insufficient grounds to rule out functional organization of sites. However, in this particular case, the sepsid enhancer is actually more conserved than the *Anopheles* enhancer relative to each of their *Drosophila* orthologs (compare graphs and score in Figures 1 and 2). Below we show that these blocks of alignment in Acalyptratae se-
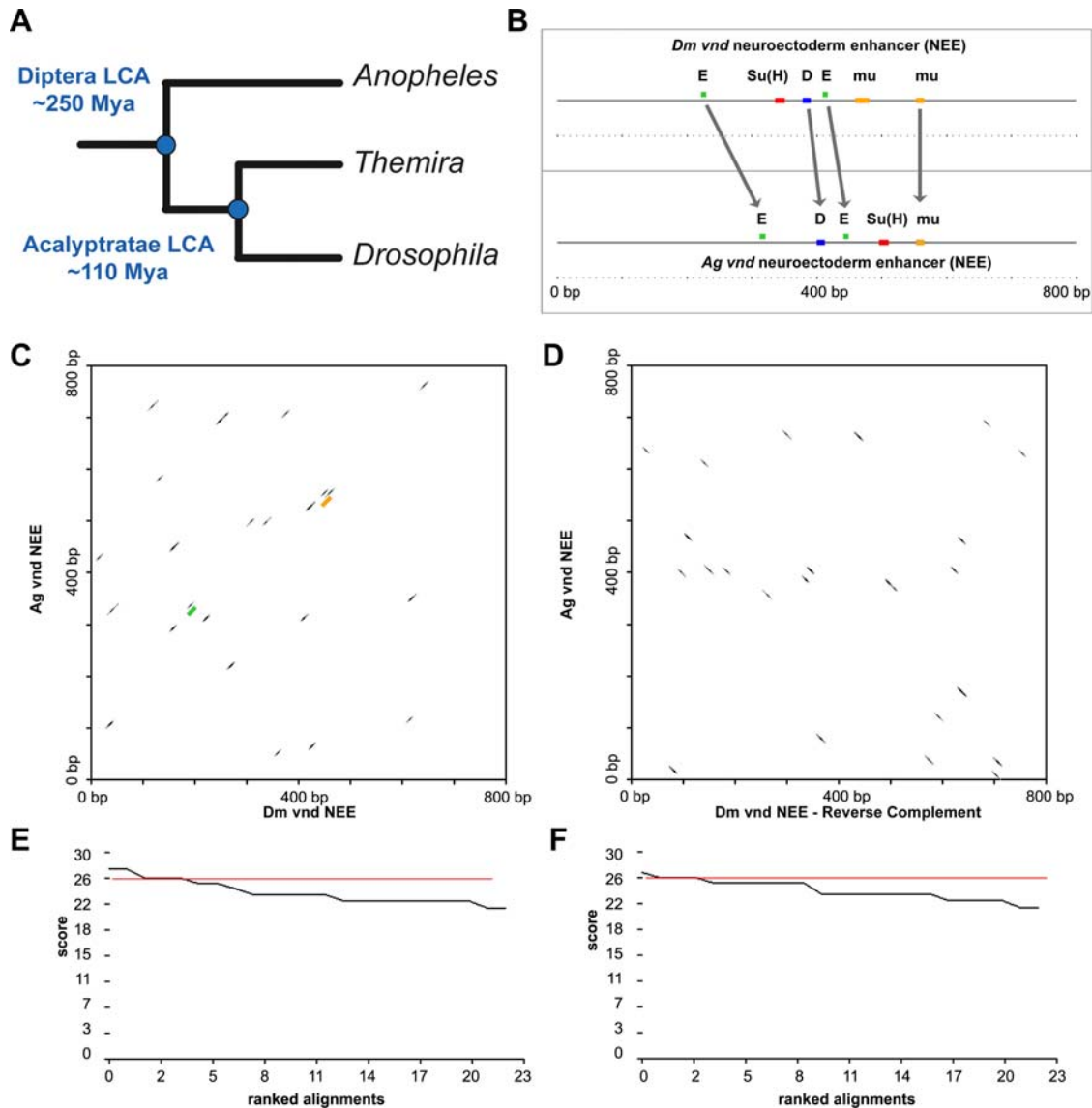
Figure 2. Evolutionary scrambling at the *Drosophila* and *Anopheles vnd* neuroectoderm enhancers (NEEs). (A) Shown is a phylogenetic tree of the three dipteran species discussed in the study: the sepsid fly *Themira putris*, the fruitfly *Drosophila melanogaster*, and the mosquito *Anopheles gambiae*. The amount of divergence from their latest common ancestors (LCA) is depicted in Millions of years ago (Mya). (B) The *Drosophila* and *Anopheles vnd* enhancers still share a common organization of functional binding sites as previously reported [30]. The colored boxes represent matches to the Dorsal (blue), Twist (green), mu (orange), and Su(H) (red) motifs. Two-dimensional homology plots for the *Drosophila* and *Anopheles vnd* enhancers in parallel (C) and anti-parallel (D) orientations reveal spurious blocks of alignment, as would be seen between two DNA sequences chosen randomly. Only two of the motifs shown in (B) (highlighted in green and orange) appear in the plot in (C) as indicated. The anti-parallel two-dimensional plot of these enhancers does not differ qualitatively or quantitatively from the parallel plot. (E–F) Score distributions for (C) and (D), respectively, are quite similar as well. Therefore, it is difficult to rule out organized enhancer elements without extensive sequence inspection.
doi:10.1371/journal.pgen.1000276.g002

quences correspond to known transcription factor binding sites.

Activator and Repressor Binding Sites in the Highly Ordered Blocks of Conservation

There are seven large blocks of alignment between the *Drosophila*/*Themira eve* stripe 2 enhancers, and these span the entire length of the enhancer (Figures 1, 3, and 4). A priori, such blocks of alignment are typical of evolution at insect regulatory modules that preserve binding sites while experiencing relatively greater amounts of turnover, insertions, and deletions within intervening sequences. We began by locating in the conserved blocks of the

Drosophila/*Themira eve* stripe 2 enhancers all of the well-known sites as indexed in the original biochemical and phylogenetic studies [18–25]. We use position-weighted matrices (PWMs) only when they accurately call the experimentally confirmed sites in *D. melanogaster* with high specificity. We note that this conservative technique may result in under-calling of *Themira* sites,

Figure 3. Sequence for identity blocks between *Drosophila* and *Themira eve* stripe 2 enhancers. The seven blocks of conservation shown in Figure 1, blocks (A–G), correspond to sequences encompassing well-known binding sites for Bicoid, Kruppel, and Giant in the *Drosophila eve* enhancer. The percent identity is given for each block. Abbreviations: GT, Giant; KR, Kruppel; BCD, Bicoid; KRW, weak Kruppel, i.e., low-affinity Kruppel binding; PWM, position-weighted matrix; FP, biochemical footprint.
doi:10.1371/journal.pgen.1000276.g003

including organized sites, because the position-weighted matrices were developed to *Drosophila* sequences, and because the *Themira* binding preferences may have diverged since their latest common ancestor, resulting in an artifactual phylogenetic decay of detection. Nonetheless, here we report that these seven large blocks of alignments, which are present in a conserved order or serial arrangement in both species, correspond to well-known binding sites for both activators and repressors (Figures 3 and 4).

Specifically, two high-affinity Kruppel repressor binding sites, KR-6 and KR-5, occur in conserved blocks A and B, respectively, while one and two low-affinity Kruppel binding sites (KRW sites) are present in conserved blocks E and F, respectively (Figures 3 and 4). Thus, this organized array of conserved Kruppel repressor binding sites spans ~300 bp. Both low and high affinity sites are likely to be important in precisely reading out gradients of repressor activity [19–22]. Additionally, Bicoid activator binding

BCD-5 and BCD-4 are present in conserved block B. Last, known Giant repressor binding sites are present in blocks B and D. Block D, the largest block of alignment at 41 bp, also corresponds quite well to the large biochemical footprint for *Drosophila* Giant at this site [20]. Two other conserved blocks, blocks C and G, are conserved and present in the same order in both species, but do not match any known biochemical activities. Thus, five of the seven blocks of alignment, each corresponding to a length wider than the
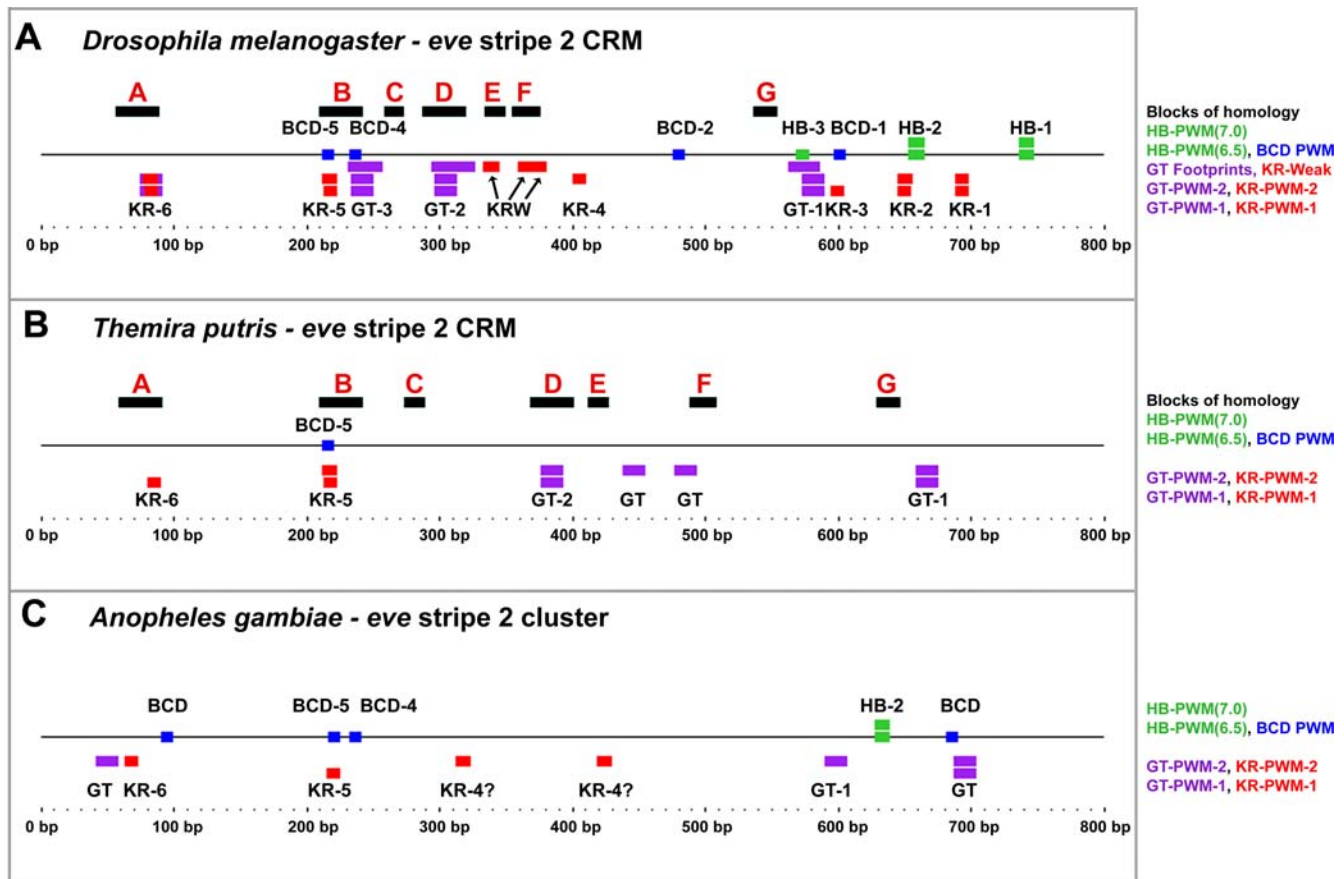
Figure 4. Organization of dipteran *eve* stripe 2 enhancers. The seven blocks of conservation whose serial order is conserved over a stretch of 500–600 bp across the *eve* stripe 2 enhancers of *Drosophila* and *Themira* are depicted. The blocks are shown in the order depicted in Figure 1A, blocks A–G. The colored boxes represent matches to Hunchback PWMs at two levels of stringency (lime green), a Bicoid PWM (blue), a Giant PWM (purple), and Kruppel PWMs at two levels of stringency (red). DNA binding activities for Giant and Kruppel as determined by biochemical assays are also depicted. Numbering system follows previous studies [24,25].
doi:10.1371/journal.pgen.1000276.g004

typical binding motif, encompass well-known activator and repressor binding sites conserved in a basic order spanning the length of the enhancer for each species. This organization is of a much longer range than the conservation of adjacent binding sites noted in the study.

Similar analyses at other *even-skipped* enhancers for A/P modulated stripes reveals a similar conserved organization of binding sites (Figure 5). For example, the *eve* stripe 4+6 enhancer contains ordered blocks corresponding to known Hunchback, Tailless, and Knirps binding sites (e.g., see Figure 5C). Additionally, there are locally rearranged blocks of sequence that destroy homology, but nonetheless preserve the presence of

specific sites in the same higher-order organization (e.g., Motif block * in Figure 5D). Such sequence signatures are consistent with selection for compensatory mutations preserving binding sites in equivalent micro-neighborhoods within the enhancer [26,32,33]. Such a process can preserve functional organization while destroying alignment homology at specific sites.

Conclusions

The conclusion of the sepsid study is premature because the basic premise of scrambled enhancers is doubly flawed: 1) these enhancers are not scrambled, and 2) even if they were scrambled, this would be

insufficient grounds to rule out the importance of enhancer-wide functional organization of motifs as demonstrated by evolution at the dipteran *vnd* enhancer. A good test of the importance of this order of functional elements would be to rearrange these sites by mutagenesis and verify whether an "imperturbable core" is or is not present in *eve* stripe 2 enhancers. In conclusion, even though we can now easily generate panoramic views of entire genomes, we should still focus on the finer details of DNA sequence and functionally test their properties before making claims on the internal fine-structural organization of individual enhancers.
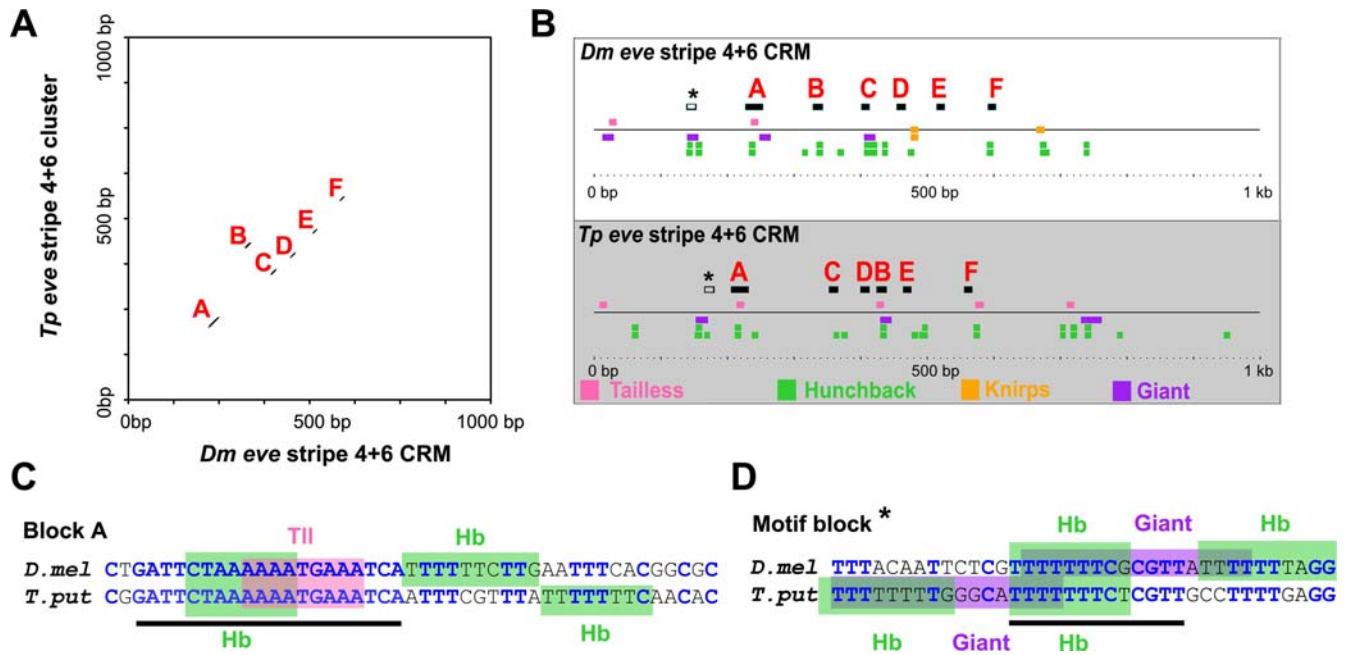
Figure 5. Similar amounts of organization at the *eve* stripe 4+6 enhancers of *Drosophila* and *Themira*. (A) A dot plot alignment of the *eve* stripe 4+6 enhancers of *Drosophila* and *Themira* at high stringency reveals highly conserved blocks of sequence, most of which are conserved in their basic order. (B) The blocks of alignment correspond to binding motifs for factors known to work at this enhancer: Tailless (pink), Hunchback (green), Knirps (orange), and Giant (purple). The block marked with an asterisk corresponds to sites that have been recreated in the same location, see also (D). Consequently, these sites do not appear in the parallel homology plot in (A). (C) An example of one of the blocks of conserved sequence, block A, is depicted. (D) An example of a block of sequence (asterisk) that has apparently experienced compensatory mutations that have shifted binding sites enough to obliterate homology. Despite such sequence turnover, such signatures preserve the higher-order organization of the enhancer.
doi:10.1371/journal.pgen.1000276.g005

References

1. Britten RJ, Davidson EH (1971) Repetitive and non-repetitive DNA sequences and a speculation on the origins of evolutionary novelty. Q Rev Biol 46: 111–138.
2. Ono S (1972) Gene duplication, mutation load, and mammalian genetic regulatory systems. J Med Genet 9: 254–263.
3. Wilson AC, Maxson LR, Sarich VM (1974) Two types of molecular evolution. Evidence from studies of interspecific hybridization. Proc Natl Acad Sci U S A 71: 2843–2847.
4. King MC, Wilson AC (1975) Evolution at two levels in humans and chimpanzees. Science 188: 107–116.
5. McDonald JF, Chambers GK, David J, Ayala FJ (1977) Adaptive response due to changes in gene regulation: A study with Drosophila. Proc Natl Acad Sci U S A 74: 4562–4566.
6. Levine M, Tjian R (2003) Transcription regulation and animal diversity. Nature 424: 147–151.
7. Wittkopp PJ, Haerum BK, Clark AG (2004) Evolutionary changes in cis and trans gene regulation. Nature 430: 85–88.
8. Gompel N, Prud'homme B, Wittkopp PJ, Kassner VA, Carroll SB (2005) Chance caught on the wing: cis-regulatory evolution and the origin of pigment patterns in Drosophila. Nature 433: 481–487.
9. Carroll SB (2005) Evolution at two levels: On genes and form. PLoS Biol 3: e245. doi:10.1371/journal.pbio.0030245.
10. Davidson EH, Erwin DH (2006) Gene regulatory networks and the evolution of animal body plans. Science 311: 796–800.
11. Marcellini S, Simpson P (2006) Two or four bristles: Functional evolution of an enhancer of scute in Drosophilidae. PLoS Biol 4: e386. doi:10.1371/journal.pbio.0040386.

12. Prud'homme B, Gompel N, Rokas A, Kassner VA, Williams TM, et al. (2006) Repeated morphological evolution through cis-regulatory changes in a pleiotropic gene. Nature 440: 1050–1053.
13. Prud'homme B, Gompel N, Carroll SB (2007) Emerging principles of regulatory evolution. Proc Natl Acad Sci U S A 104 Suppl 1: 8605–8612.
14. Wittkopp PJ, Haerum BK, Clark AG (2008) Regulatory changes underlying expression differences within and between Drosophila species. Nat Genet 40: 346–350.
15. Hoekstra HE, Coyne JA (2007) The locus of evolution: evo devo and the genetics of adaptation. Evolution Int J Org Evolution 61: 995–1016.
16. Stern DL, Orgogozo V (2008) The Loci of evolution: How predictable is genetic evolution? Evolution 62: 2155–2177.
17. Goto T, Macdonald P, Maniatis T (1989) Early and late periodic patterns of even skipped expression are controlled by distinct regulatory elements that respond to different spatial cues. Cell 57: 413–422.
18. Harding K, Hoey T, Warrior R, Levine M (1989) Autoregulatory and gap gene response elements of the even-skipped promoter of Drosophila. Embo J 8: 1205–1212.
19. Stanojevic D, Hoey T, Levine M (1989) Sequence-specific DNA-binding activities of the gap proteins encoded by hunchback and Kruppel in Drosophila. Nature 341: 331–335.
20. Warrior R, Levine M (1990) Dose-dependent regulation of pair-rule stripes by gap proteins and the initiation of segment polarity. Development 110: 759–767.
21. Small S, Kraut R, Hoey T, Warrior R, Levine M (1991) Transcriptional regulation of a pair-rule stripe in Drosophila. Genes Dev 5: 827–839.

22. Small S, Blair A, Levine M (1992) Regulation of even-skipped stripe 2 in the Drosophila embryo. Embo J 11: 4047–4057.
23. Arnosti DN, Barolo S, Levine M, Small S (1996) The eve stripe 2 enhancer employs multiple modes of transcriptional synergy. Development 122: 205–214.
24. Ludwig MZ, Kreitman M (1995) Evolutionary dynamics of the enhancer region of even-skipped in Drosophila. Mol Biol Evol 12: 1002–1011.
25. Ludwig MZ, Patel NH, Kreitman M (1998) Functional analysis of eve stripe 2 enhancer evolution in Drosophila: Rules governing conservation and change. Development 125: 949–958.
26. Ludwig MZ, Bergman C, Patel NH, Kreitman M (2000) Evidence for stabilizing selection in a eukaryotic enhancer element. Nature 403: 564–567.
27. Ludwig MZ, Palsson A, Alekseeva E, Bergman CM, Nathan J, et al. (2005) Functional evolution of a cis-regulatory module. PLoS Biol 3: e93. doi:10.1371/journal.pbio.0030093.
28. Hare EE, Peterson BK, Iyer VN, Meier R, Eisen MB (2008) Sepsid even-skipped enhancers are functionally conserved in Drosophila despite lack of sequence conservation. PLoS Genet 4: e1000106. doi:10.1371/journal.pgen.1000106.
29. Arnosti DN, Kulkarni MM (2005) Transcriptional enhancers: Intelligent enhanceosomes or flexible billboards? J Cell Biochem 94: 890–898.
30. Maizel JV Jr, Lenk RP (1981) Enhanced graphic matrix analysis of nucleic acid and protein sequences. Proc Natl Acad Sci U S A 78: 7665–7669.
31. Erives A, Levine M (2004) Coordinate enhancers share common organizational features in the Drosophila genome. Proc Natl Acad Sci U S A 101: 3851–3856.

32. Crocker J, Tamori Y, Erives A (2008) Evolution acts on enhancer organization to fine-tune gradient threshold readouts. PLoS Biol 6: e263. doi:10.1371/journal.pbio.0060263.

33. Moses AM, Pollard DA, Nix DA, Iyer VN, Li XY, et al. (2006) Large-scale turnover of functional transcription factor binding sites in Drosophila. PLoS Comput Biol 2: e130. doi:10.1371/journal.pcbi.0020130.