Developmental Biology 362 (2012) 282–294

Contents lists available at SciVerse ScienceDirect

Developmental Biology

journal homepage: www.elsevier.com/developmentalbiology

Evolution of Developmental Control Mechanisms

Ancestral and conserved cis-regulatory architectures in developmental control genes

Mark Rebeiz [a,1], Brian Castro [a], Feng Liu [a], Feng Yue [b], James W. Posakony [a,*]

[a] Division of Biological Sciences/CDB, University of California San Diego, La Jolla, CA 92093, USA
[b] Ludwig Institute for Cancer Research, La Jolla, CA 92093, USA

ARTICLE INFO

Article history:
Received for publication 15 July 2011
Revised 1 December 2011
Accepted 6 December 2011
Available online 13 December 2011

Keywords:
Cis-regulatory evolution
Conserved motifs
Notch pathway
Dpp signaling
Hairy/Enhancer of split repressor genes
Bearded family genes

ABSTRACT

Among developmental control genes, transcription factor-target gene "linkages" – the direct connections between target genes and the factors that control their patterns of expression – can show remarkable evolutionary stability. However, the specific binding sites that mediate and define these regulatory connections are themselves often subject to rapid turnover. Here we describe several instances in which particular transcription factor binding motif combinations have evidently been conserved upstream of orthologous target genes for extraordinarily long evolutionary periods. This occurs against a backdrop in which other binding sites for the same factors are coming and going rapidly. Our examples include a particular Dpp Silencer Element upstream of insect *brinker* genes, in combination with a novel motif we refer to as the Downstream Element; combinations of a Suppressor of Hairless Paired Site (SPS) and a specific proneural protein binding site associated with arthropod Notch pathway target genes; and a three-motif combination, also including an SPS, upstream of deuterostome *Hes* repressor genes, which are also Notch targets. We propose that these stable motif architectures have been conserved intact from a deep ancestor, in part because they mediate a special mode of regulation that cannot be supplied by the other, unstable motif instances.

Introduction

It is now well recognized that changes in transcriptional cis-regulatory elements, particularly those that direct the expression of developmental control genes, represent a fundamental mechanism underlying animal evolution (Davidson, 2006; Wray, 2007). Such cis-regulatory novelties have been shown to confer both loss (Chan et al., 2010; Jeong et al., 2008; Prud'homme et al., 2006) and gain (Gompel et al., 2005; Prud'homme et al., 2006; Rebeiz et al., 2011) on a gene's repertoire of expression specificities. But cis-regulatory evolution is not restricted to the generation of major alterations in gene activity. Even orthologous enhancer modules that drive very similar patterns of expression in two species can differ enormously in their cis-regulatory architecture — the number, order, spacing, and orientation of their component transcription factor binding sites (Hare et al., 2008; Ludwig et al., 2000; Markstein et al., 2004; Romano and Wray, 2003; Swanson et al., 2011).

In this context, it is important to distinguish between a transcription factor-target gene "linkage" – the direct regulatory connection between factor and target – and the specific binding site instances that mediate and define this connection. A transcriptional regulatory linkage might be quite stable evolutionarily even as the relevant binding sites are turning over.

We have previously suggested that transcriptional linkages that confer abstract or generic developmental regulatory capabilities, of general utility to all metazoans, might be expected to be retained for especially long evolutionary periods (Rebeiz et al., 2005). We described one such example, the direct transcriptional repression of genes encoding proneural basic helix-loop-helix (bHLH) activator proteins by bHLH repressor factors of the Hairy/Enhancer of split (Hes) class. We found that bilaterian proneural genes belonging to both the *achaete-scute* and *atonal* classes (representing an ancient division that predates the cnidarian–bilaterian divergence) are consistently associated with a high-affinity binding site for a Hes repressor, suggesting that this linkage might be more than 500 million years (My) old. The generic ability to shape spatial patterns of proneural gene expression by direct repression would in principle be valuable regardless of the specific nature of a given species' nervous system, and we suggested that this might be the basis for the long-term maintenance of this regulatory linkage. To our surprise, we also saw evidence in this phylogenetic study that not only was the Hes repressor-proneural gene linkage being retained in evolution, but that in some cases the specific binding site itself was also conserved over very long periods (Rebeiz et al., 2005).

Here we investigate the evolutionary history of two other transcriptional regulatory linkages involving developmental control genes. Insect genomes include a single gene encoding the transcriptional repressor protein Brinker, which plays an important role in regulating other genes that are targets of the Decapentaplegic (Dpp) signaling pathway

* Corresponding author at: Division of Biological Sciences/CDB, UC San Diego, 9500 Gilman Drive, Mail Code 0349, La Jolla, CA 92093, USA. Fax: +1 858 822 3021.
E-mail address: jposakony@ucsd.edu (J.W. Posakony).
[1] Present address: Department of Biological Sciences, University of Pittsburgh, Pittsburgh, PA 15260, USA.

0012-1606/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.ydbio.2011.12.011

(Affolter and Basler, 2007). Transcription of the *brinker* (*brk*) gene is itself subject to repression in response to Dpp signaling (Muller et al., 2003). This is mediated by cis-regulatory motifs upstream of *brk* known as Dpp Silencer Elements (SEs) (Pyrowolakis et al., 2004), which bind a tetrameric complex that includes the transcription factors Mothers against dpp (Mad), Medea (Med), and Schnurri (Shn) (Gao et al., 2005). Remarkably, the *brk* gene in some species is associated with multiple SEs; the fruit fly *Drosophila melanogaster* has 11, while the mosquito *Anopheles gambiae* has 12, leading to the suggestion that this architecture has been evolutionarily conserved (Yao et al., 2008). We show here that other species have only a single SE upstream of their *brk* ortholog. Moreover, we have identified in nine species representing five insect orders a unique SE upstream of *brk* that is not only unusually related between species but is also uniquely associated with a novel motif we refer to as the Downstream Element (DE). We propose that this SE + DE motif combination has been conserved from a common insect ancestor, even as the number of other SEs upstream of *brk* has been changing rapidly in evolution.

The second regulatory linkage we have analyzed involves target genes of the Notch cell–cell signaling pathway (Bray, 2006; Fiuza and Arias, 2007). Suppressor of Hairless [Su(H); CBF1 in vertebrates] functions as the transducing transcription factor for this pathway. In the absence of signaling through the Notch receptor, Su(H) acts to repress Notch target genes. Activation of the receptor leads to the cleavage of its intracellular domain (NICD), which enters the nucleus and forms a trimeric complex with Su(H) and the co-activator protein Mastermind (Mam); this complex now transcriptionally activates the formerly repressed targets.

The known repertoire of Notch pathway targets in both protostomes and deuterostomes includes genes encoding members of the Hes family of bHLH transcriptional repressor (bHLHR) proteins (Bailey and Posakony, 1995; Jarriault et al., 1995; Lecourtois and Schweisguth, 1995). These factors function to inhibit the expression of genes associated with cell fates that are antagonized by Notch signaling. Arthropods also have a second class of Notch targets, the *Bearded* (*Brd*) family genes (*BFMs*) (Bailey and Posakony, 1995; Fontana and Posakony, 2009; Lai et al., 2000; Nellesen et al., 1999). The Notch ligands Delta and Serrate require mono-ubiquitination of their ICDs by the E3 ligase Neuralized (Neur) in order to be active in signaling (Le Bras et al., 2011). In cells that receive and respond to Notch signals, Brd proteins act as competitive inhibitors of the Neur-ligand binding interaction, thereby preventing ligand activation (Bardin and Schweisguth, 2006; Fontana and Posakony, 2009). This helps keep Notch responder cells from themselves becoming effective signalers, thus ensuring the desired directionality of the signaling event.

Su(H) typically binds to its targets via one or more occurrences of an eight-nucleotide motif (Tun et al., 1994), but a small subset of target genes are associated with a special regulatory element known as the Su(H) Paired Site (SPS) (Bailey and Posakony, 1995; Nellesen et al., 1999). This consists of two high-affinity binding sites in opposite orientations, typically separated by 15–17 base pairs. By mediating the cooperative binding of two Su(H)/NICD/Mam trimers, the SPS drives an especially sensitized response to low levels of Notch signaling (Arnett et al., 2010; Nam et al., 2007; Ong et al., 2006). We show that an SPS motif is associated with certain orthologous *Hes* repressor and *Brd* family genes that last had a common ancestor hundreds of millions of years ago. In each instance, the SPS is consistently accompanied by specific binding sites for one or more other key regulatory factors. As with the insect *brk* SE + DE motif combination, we propose that these SPS-containing motif ensembles are ancestral and have been conserved for extraordinarily long evolutionary periods.

Why might a subset of the binding motifs that constitute the cis-regulatory architecture of a developmental control gene be conserved from a deep ancestor, while other motifs are changing freely in evolution? We extend our earlier proposal to include not only the principle of the linkage's utility to a diverse range of organisms, but also the concept that these ancestral and conserved motifs mediate special modes of transcriptional regulation that are not conferred by other binding sites, even for the same factor or factors.

Materials and methods

Genome sequences

The following genome sequences were utilized in this study: *D. melanogaster* (Adams et al., 2000); *A. gambiae* (Holt et al., 2002); *Aedes aegypti* (Nene et al., 2007); *Apis mellifera* (Honeybee Genome Sequencing Consortium, 2006); *Bombyx mori* (International Silkworm Genome Consortium, 2008; Mita et al., 2004); *Tribolium castaneum* (Tribolium Genome Sequencing Consortium, 2008); *Nasonia vitripennis* (Werren et al., 2010); *Rhodnius prolixus* (http://genome.wustl.edu/genomes/view/rhodnius_prolixus/); *Acyrthosiphon pisum* (International Aphid Genomics Consortium, 2010); *Mayetiola destructor* (http://www.hgsc.bcm.tmc.edu/); *Pediculus humanus corporis* (http://www.vectorbase.org/); *Daphnia pulex* (Colbourne et al., 2011); *Homo sapiens* (Venter et al., 2001); *Xenopus tropicalis* (Hellsten et al., 2010); *Danio rerio* (http://www.sanger.ac.uk/); *Gallus gallus* (International Chicken Genome Sequencing Consortium, 2004); *Strongylocentrotus purpuratus* (Sea Urchin Genome Sequencing Consortium, 2006); *Saccoglossus kowalevskii* (http://www.hgsc.bcm.tmc.edu/); *Branchiostoma floridae* (Putnam et al., 2008); *Amphimedon queenslandica* (Srivastava et al., 2010); *Trichoplax adhaerens* (Srivastava et al., 2008); *Nematostella vectensis* (Putnam et al., 2007); and *Acropora digitifera* (Shinzato et al., 2011).

Gene annotation and figure preparation

Gene structure annotation, detection of transcription factor binding motifs, and gene diagram figure preparation was carried out using the GenePalette software tool (Rebeiz and Posakony, 2004) (www.genepalette.org).

Hierarchical clustering analysis

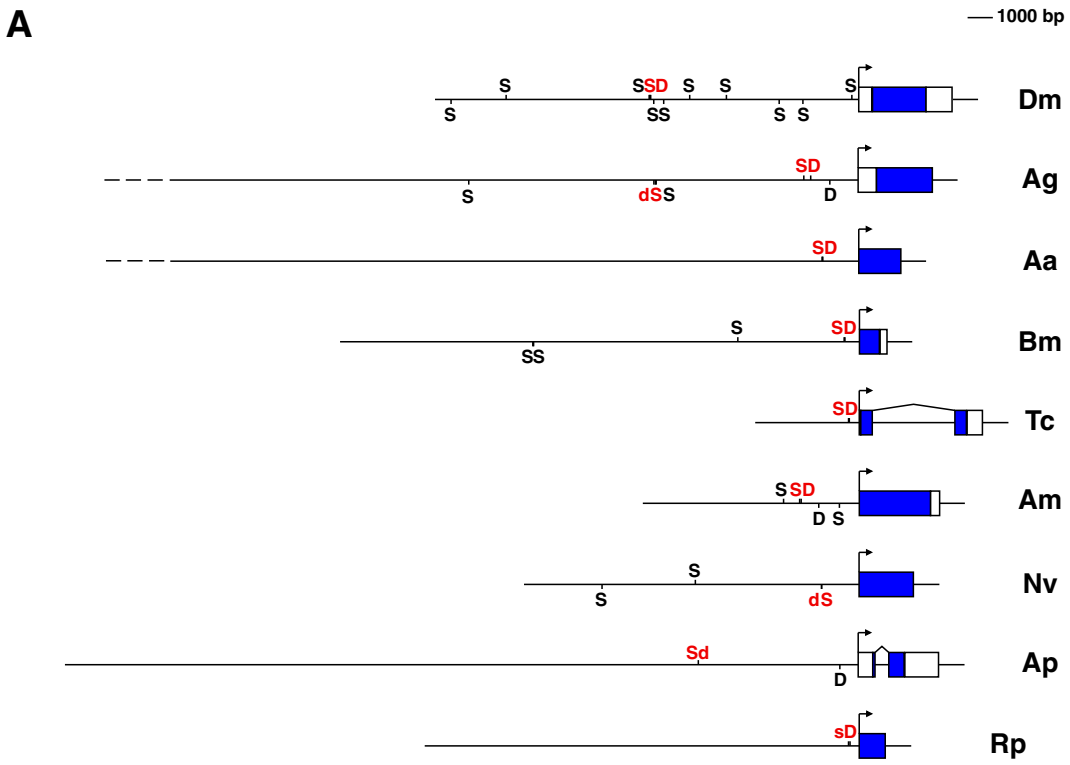
A total of 32 Dpp Silencer Element (SE) motifs conforming to the original GRCGNCN$_5$GTCTG definition (Pyrowolakis et al., 2004) were detected upstream of *brk* in the nine insect species shown in Fig. 1A. Sequence relationships between the motifs were investigated by multiple alignment using ClustalX version 2.1 (Larkin et al., 2007); the resulting phylogenetic tree was displayed using NJplot version 2.3 (Perriere and Gouy, 1996).

Logo plots

Sequence logo plots were generated using WebLogo version 2.8.2 (Crooks et al., 2004).

Comparison of sequence information content

A total of 42 SE motifs conforming to the modified GNCKNCN$_5$GTCTG definition suggested by Yao et al. (2008) were detected upstream of *brk* in the nine species shown in Fig. 1A. Ten of these belong to the "special" subset. To compare the information content of the "special" subset to that of the remaining 32 motif instances ("other"), we inventoried the *D. melanogaster* genome sequence (version R5/dm3) for all occurrences of the modified SE motif in noncoding regions; 1251 were found. The sequence information content (SIC, uncorrected for small sample size) of 1000 randomly chosen sets of 10 and 1000 randomly chosen sets of 32 (all drawn from the 1251 genomic instances) were computed, and the mean SIC values of these sets were then compared to those of the "special" (10) and "other" (32) subsets, respectively.



C

	SE	DE
Dm	CAAGACT**GGCGACATTCTGTCTG**GTGGCGATCGCC----------	**GCAACGTTGCCAC**TTTTCAT
Ag1	TGCCGGT**GGCGACAATTTGTCTG**GCCCG	GCGGC**GtTTTGTTGCCGC**TTGTTGC
Ag2	GCGCCGT**GGCGACAATTTGTCTG**GTGCG	TTGAT**GCTGCGTTGCCAC**GTTTGCC
Aa	AAGTCGT**GGCGACAATTTGTCTG**GTCGGTATTATATA----------	CTTTGC**GCGACGTTGCCGT**TGTGTGC
Bm	GCGAGCC**GGCGACACCCTGTCTG**CCACGGTCG----------	**GCGGTGTTGCCGT**TTCGCAC
Tc	CTGGCTT**GGCGACAGCTTGTCTG**GCGTGACCAGTT----------	**GCGCAGTTGCCAC**TTCCCCG
Am	CCGAGCT**GGCGACACCCTGTCTG**GCAAGACCGGCCGACGGACAAAGGAGGGGGAGCCTGGCC-	**GCTGCGTTGCCAC**GTCCTCG
Nv	CAGTCCC**GGCGCCATGTTGTCTG**CTC----------	**GaTCGGTTGCCGC**CGCCGCA
Ap	GGCGTTC**GGCGTCACCCTGTCTG**GCCGCTGCGGACCGAAC----------	**GCCACGaTGCCAC**GTCGTGC
Rp	CCGAGCT**GGCGTCGCCCTGTCTa**GCAGTATCGGCTCCCACTACAACACTAAGCATGTGGGCGA	**GCAATGTTGCCAC**ATTCATT

Identification of shared sequence motifs

To confirm that our identification of conserved sequence motifs upstream of *Hes* and *BFM* genes is substantially complete, we made use of the San Diego Supercomputer Center (SDSC) implementation of MEME (Bailey and Elkan, 1994) (http://meme.sdsc.edu/).

Transgene construction

A wild-type GFP reporter transgene bearing the "E + SPS + P" enhancer module from upstream of the *A. gambiae Brd* family gene (*Ag BFM-GFP*) was constructed as follows. A 1.0-kb genomic DNA fragment covering −37 to −1037 upstream of *Ag BFM* was PCR-amplified using the following primers: Fwd: 5′-gaattcCTCTGAATAGCGCAAAACACAACACAATCGCAGGAC-3′ and Rev: 5′-ctcgagCCGACCCCGACCCCGACCCTTTCCACG-3′ (lowercase letters represent added EcoRI and XhoI restriction sites, respectively). The fragment was inserted into the multiple cloning site of the insulated P-element transformation vector pH-Stinger (Barolo et al., 2004). Mutant versions of this reporter transgene were prepared by changing the single proneural protein binding site (P) from GCAGGTG to GAAGCTT (Van Doren et al., 1992) (*Ag BFM-Pm-GFP* and *Ag BFM-UmPm-GFP*) and/or by changing the two Su(H) binding sites (U) in the SPS (of the form YGTGGGAA) to YGTGGCAA (Bailey and Posakony, 1995) (*Ag BFM-Um-GFP* and *Ag BFM-UmPm-GFP*).

The *brk* enhancer module "7/8/9" described by Yao et al. (2008) was amplified by PCR from genomic DNA using the following primers: Fwd: 5′-gcatgcTATATACATGGGGTGGCATGAGCATGTGCG-3′ and Rev: 5′-ggcgcgccCCACAAGGCGCTAGAACGAGATGGCGCACA-3′ (lowercase letters represent added SphI and AscI restriction sites, respectively). Following sequence verification, the fragment was introduced into the multiple cloning site of the attB-H-Stinger GFP reporter vector (S.W. Miller, UC San Diego, unpublished; further information available upon request). The single Downstream Element (DE) in the 7/8/9 module was mutated by changing the wild-type sequence GCAACGTTGCCACTT to TCAACTTGGACCCTG.

Generation of transgenic fly lines

Wild-type and mutant versions of the *Ag BFM-GFP* enhancer-reporter construct were introduced into the w^{1118} recipient strain via P element-mediated germline transformation (Rubin and Spradling, 1982).

Wild-type and mutant versions of the *brk* 7/8/9 enhancer-reporter construct were introduced into the genome using the ϕC31-integrase system (Bischof et al., 2007). The recipient strain carries an attP docking site on the third chromosome (attP2 site at 68A4) and the nanos-ϕC31 integrase gene on the X chromosome (Stock #25710, Bloomington Drosophila Stock Center). Multiple independent insertion lines were obtained for both the wild-type and DEm versions. Integration events were confirmed by PCR according to Venken et al. (2006).

Tissue preparation, antibody staining, and confocal microscopy

Wing imaginal discs from late third-instar larvae and nota from pupae at 14 h after puparium formation (APF) were dissected in PBT (1X PBS, 0.1% Triton-X 100) and fixed in 4% paraformaldehyde (in PBT) for 30 min at room temperature. Discs were washed in PBT, mounted, and imaged using a Leica TCS SP2 confocal microscope. Nota were first stained with anti-Hnt (monoclonal, Developmental Studies Hybridoma Bank) primary antibody and Alexa 555 (Molecular Probes) secondary antibody before mounting. Confocal average projections were generated with Z-axis sections at 2-μm (discs) or 1-μm (nota) intervals using Leica Confocal Software version 2.5 (Leica Microsystems). To compare fluorescence between wild-type and DEm *brk* 7/8/9 enhancer-reporter constructs, wing discs of the two genotypes were processed side-by-side and imaged using the same gain.

Results

Conservation of a unique Dpp Silencer Element upstream of insect brinker genes

The 16.6-kb intergenic region upstream of the *D. melanogaster brk* gene (Fig. 1A) includes no fewer than nine instances of the original GRCGNCN$_5$GTCTG definition of the SE motif (Pyrowolakis et al., 2004) and 11 occurrences of the modified motif GNCKNCN$_5$GTCTG suggested by Yao et al. (2008). Based on their thorough analyses using reporter transgenes in vivo, Yao et al. (2008) concluded that these 11 SE motifs probably correspond to nine or ten distinct cis-regulatory modules, each of which, by integrating activator and SE inputs, contributes independently to the *brk* expression pattern.

Their identification of 12, 11, and 11 SE sites upstream of *brk* in *Drosophila pseudoobscura* (23 kb), *Drosophila virilis* (24 kb), and the mosquito *A. gambiae* (95 kb), respectively, prompted Yao et al. to suggest that this unusual cis-regulatory organization is evolutionarily conserved. However, a different picture emerges when the *brk* upstream regions of other insects are examined (Table 1; Fig. 1A). The 20 kb upstream of *brk* in *B. mori* (silk moth) and the 8.5 kb of upstream sequence in *A mellifera* (honeybee) each include three original SE motifs (*Bombyx* has four matches to the Yao et al. site definition). In both *T. castaneum* (red flour beetle) and *A. pisum* (pea aphid), the *brk* upstream region (4.1 kb and 31 kb, respectively) contains only a single occurrence of either SE motif. And while another mosquito (*A. aegypti*) has 311 kb of intergenic sequence upstream of *brk*, this very large region includes only three and six instances of the original and the Yao et al. SE motifs, respectively, and only one and two instances, respectively, in the first 200 kb upstream. Thus, even if we assume that all of these motif occurrences represent functional SEs in vivo, the presence of large numbers of them is not a shared characteristic of insect *brk* genes.

Fig. 1. A Silencer Element (SE) + Downstream Element (DE) cis-regulatory motif combination is a shared feature of insect *brk* genes. A: Annotated scale diagrams of the *brk* gene and its upstream non-coding sequences in representatives of various insect orders. Blue boxes represent *brk* protein coding sequences; white boxes represent untranslated regions. Arrows denote direction of transcription and are positioned at either the transcription start site or the start codon of the gene. Except in the case of *Ag* and *Aa*, the entire intergenic region between *brk* and the next upstream gene is shown. Dpp Silencer Elements (Yao et al., 2008) are indicated by "S"; Downstream Elements are denoted by "D". Shared S + D motif combinations are shown in red. Lower-case "s" upstream of *Rp brk* denotes single-base mismatch to the S motif definition; lower-case "d" upstream of *Ag*, *Nv*, and *Ap brk* denotes single-base mismatch to the D motif definition (see C for alignment). Other single-base mismatches to either motif are omitted. Sequence scale indicated in upper right corner. *Dm, Drosophila melanogaster* (Diptera); *Ag, Anopheles gambiae* (Diptera); *Aa, Aedes aegypti* (Diptera); *Bm, Bombyx mori* (Lepidoptera); *Tc, Tribolium castaneum* (Coleoptera); *Am, Apis mellifera* (Hymenoptera); *Nv, Nasonia vitripennis* (Hymenoptera); *Ap, Acyrthosiphon pisum* (Hemiptera); *Rp, Rhodnius prolixus* (Hemiptera). B: Logo plots of the "special" SE subset (upper, 10 occurrences) and all "other" SE motifs (lower, 32 occurrences) upstream of the insect *brk* genes shown in A. Note the additional sequence constraint in the "special" subset in both unconstrained (5, 8, 10, 12–14) and partially degenerate (7) positions within the motif, as well as in flanking positions (3, 20, 21), compared to that in the "other" set. The "special" subset has an uncorrected sequence information content of 30.35 bits, far higher than the mean of 21.01 bits for 1000 randomly chosen sets of 10 SE motif instances drawn from the non-coding portion of the fly genome (see Materials and methods). By contrast, the "other" subset has an uncorrected sequence information content of 20.43 bits, compared to a mean of 18.62 bits for 1000 randomly chosen sets of 32 SE motif instances. C: Sequence alignment of the SE + DE motif combinations shown in red in A. Except for the two occurrences in *Ag*, actual distances between SEs and the corresponding DEs are shown. D: Logo plot of the DE motifs associated with the 10 "special" SEs (see A, C).

Table 1
SE and DE motif occurrences upstream of insect *brinker* genes.

Species[a]	Upstream (kb)[b]	#oSE[c]	#mSE[d]	#DE[e]	SE + DE motif combination(s)[f]
Dm	17	9	11	1	2xSE[g] + DE
Ag	95	8	12	2	SE + DE, 2xSE[g] + DEmm[h]
Aa	311	3(1)[i]	6(2)[i]	1	SE + DE
Bm	20	3	4	1	SE + DE
Tc	4.1	1	1	1	SE + DE
Am	8.5	3	3	2	SE + DE
Nv	12	3	3	2 mm[h]	SE + DEmm[h]
Ap	31	1	1	1, 5 mm[h,j]	SE + DEmm[h,j]
Rp	17	1mm[h]	1mm[h]	1	SEmm[h] + DE

[a] Species symbols are as listed in the legend to Fig. 1.
[b] Indicates the size of the intergenic region upstream of *brk*, based on current genome annotations.
[c] Number of Dpp Silencer Elements (SE) upstream of *brk*, using the original (o) $GRCGNCN_5GTCTG$ motif definition (Pyrowolakis et al., 2004).
[d] Number of Dpp Silencer Elements (SE) upstream of *brk*, using the modified (m) $GNCKNCN_5GTCTG$ motif definition (Yao et al., 2008).
[e] Number of Downstream Elements (DE; $GCN_3GTTGCCRY$) upstream of *brk*.
[f] Nature of SE + DE motif combination(s) found.
[g] *Dm* and *Ag* both have two closely spaced SEs associated with a DE (*Dm*) or a DEmm (*Ag*); see Fig. 1A.
[h] One-base mismatch (mm) to the motif definition.
[i] Shown in parentheses is the number of SEs within the first 200 kb upstream of *brk*.
[j] *Ap* has one exact match to the DE motif definition, not associated with its single SE; the SE is associated with a single-base-mismatch DE motif (see Figs. 1A, C).

To investigate further the question of evolutionary conservation, we asked if it was possible to discern any exceptional similarity between *brk*-associated SE motifs in the different species that might be suggestive of site orthology. Indeed, using ClustalX for hierarchical motif clustering, we identified ten SE motifs from nine species that comprise a distinct subset of the total ensemble of occurrences (see Fig. S1). First, these ten sites define an unusually constrained version of the SE motif (Fig. 1B). Significant information content is evident at three positions flanking the motif (3, 20, and 21 in Fig. 1B), while two partially degenerate and several fully unconstrained positions within the motif show strong (8, 10, 12, 13) and even complete (5, 7, 14) sequence bias. By contrast, the remaining 22 (32 by the Yao et al. definition) SE motif occurrences upstream of *brk* in the nine species contain little more information than that embodied in the motif definitions (Fig. 1B) (Pyrowolakis et al., 2004; Yao et al., 2008).

A second and unique characteristic of this distinctive SE motif class is the presence, at a variable but typically quite short distance downstream from the SE, of a novel sequence we refer to as the Downstream Element (DE) (Figs. 1A, C, D). The DE is always found in the same orientation with respect to the SE, regardless of the SE's orientation with respect to the direction of *brk* transcription (see Fig. 1A). With the exception of *A. gambiae brk*, no *brk* gene includes more than one SE + DE combination, even when the more relaxed Yao et al. (2008) SE definition and single-base mismatches to the DE definition ($GCN_3GTTGCCRY$) are both permitted. Thus, in each species (again, *Anopheles* being the sole exception), the single sequence-constrained SE motif described above is paired uniquely with a DE.

The nine species we have considered in our analysis represent five insect orders (Diptera, Lepidoptera, Coleoptera, Hymenoptera, and Hemiptera). That the *brk* upstream region in each species includes a single SE (*Anopheles* has two) that is both a member of the sequence-constrained set and uniquely associated with a DE is likely to be highly significant both evolutionarily and functionally. Particularly noteworthy is the observation that the single SE found upstream of *brk* in three species (*Tribolium*, *Acyrthosiphon*, and *Rhodnius*) is in each case a member of this special set. We suggest that this phenomenon represents the long-term evolutionary conservation of an ancestral SE + DE unit that, by comparison to other SE motif occurrences, confers a unique regulatory functionality.

The brk Downstream Element (DE) functions in activation

The long-term conservation of the DE motif upstream of insect *brk* genes strongly implies its functionality, and we sought to test this expectation directly. Yao et al. (2008) showed previously that a 784-bp region upstream of *Drosophila brk* that includes three SEs drives a pattern of *lacZ* reporter expression in the wing imaginal disc similar to that of endogenous *brk* (Fig. 1A; Fig. 2). The DE of *Drosophila brk* lies adjacent to the middle SE in this fragment (see Figs. 1A, C; Fig. 2A). We compared the activities of reporter transgenes in which GFP expression is driven by either wild-type or DE-mutant versions of the fragment (Fig. 2). We find that, while their spatial patterns of GFP activity in imaginal discs appear very similar, the DE-mutant reporter (Figs. 2F–H) is expressed at a much lower level than the wild-type reporter (Figs. 2B–E). This result implies that the conserved DE does indeed have a functional role in the transcriptional activation of *brk* expression in this tissue.

Long-term evolutionary conservation of SPS-containing cis-regulatory architectures upstream of Notch pathway target genes

In previous reports, we have described the utilization of a "P + S" cis-regulatory code by genes that are activated via Notch signaling during lateral inhibition in proneural clusters (Bailey and Posakony, 1995; Castro et al., 2005; Nellesen et al., 1999; Singson et al., 1994). S sites mediate the activation and repression functions of the Notch-regulated transcription factor Su(H), while P sites mediate activation by proneural proteins. In *Drosophila*, Notch targets known to employ this code include bHLH repressor genes of the *Hes* class, as well as *Brd* family genes (Bailey and Posakony, 1995; Lai et al., 2000; Lecourtois and Schweisguth, 1995; Nellesen et al., 1999). The number and location of S and P motifs vary greatly from one target gene to another (Nellesen et al., 1999).

A subset of Notch target genes that use the "P + S" code is characterized by the presence of a special motif called the Su(H) Paired Site or SPS, which consists of two high-affinity Su(H) binding sites in opposite orientations, separated by 15–17 bp (Bailey and Posakony, 1995; Nellesen et al., 1999). This distinctive element is often accompanied by one or more single, or "lone", Su(H) sites.

Examination of the upstream regulatory regions of orthologous *Hes* bHLH repressor and *Brd* family genes in various arthropods reveals in each case the apparent long-term conservation of a particular P + SPS motif combination (Fig. 3; highlighted in red). In the case of the *Hes* genes, the SPS, which occurs at various locations with respect to the transcription start site, is closely flanked on the upstream side by an "upper strand" P site (Fig. 3A). *Brd* family genes, by contrast, are associated with an SPS (again at various distances upstream) accompanied by a "lower strand" P motif located closer to the transcription start site (Fig. 3B). *BFMs* are also characterized [except in *Drosophila* and other Brachyceran flies, such as the tsetse fly *Glossina morsitans* (not shown)] by the presence, upstream of the SPS, of an extended "E box" motif (RRCAGATGGY) that we have found by in vitro assays to be a variant proneural protein binding site (S.W. Miller, unpublished). Note that, in both *Hes* and *Brd* family genes, additional "lone" Su(H) sites and/or P sites may also be present, but these are not widely conserved, if at all. At a minimum, though, the distinctive P + SPS or E + SPS + P combination is present (e.g., *Ap bHLHR-1*, *Ag BFM*).

We suggest that these observations reflect the evolutionary conservation, over more than 400 My, of a specific P + SPS or E + SPS + P cis-regulatory architecture that was present in the common ancestors of these genes. This interpretation is strengthened by the fact that in most species we can clearly establish orthology between the genes we are comparing. Both the bHLH repressor and *Brd* family genes shown in Fig. 3 typically occupy the same positions in the respective *Enhancer of split* gene complexes [E(spl)-Cs] of these species; moreover, Bayesian phylogenetic analysis fully supports the orthology of the *Hes* genes

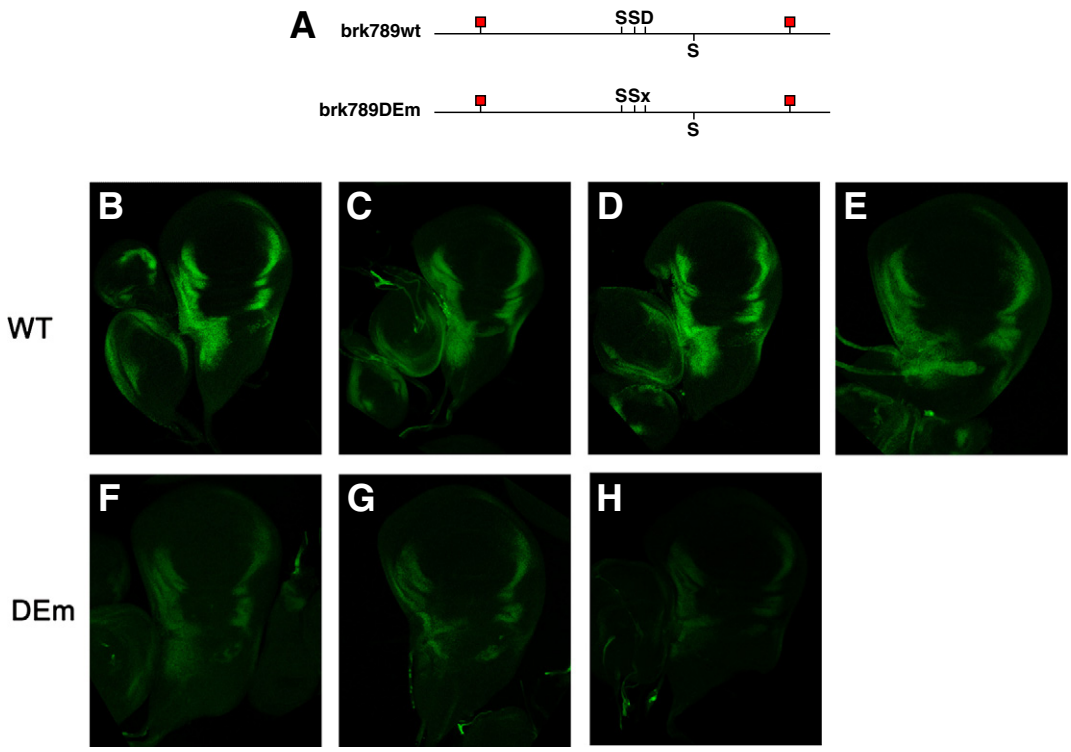


Fig. 2. The Downstream Element (DE) contributes to activation of a *brk* enhancer module in *Drosophila*. A: Diagram of the 0.8-kb "7/8/9" module from upstream of *Dm brk* (Yao et al., 2008). This fragment (bounded by red boxes) contains three SEs ("S"; the seventh, eighth, and ninth upstream of the transcription start site) plus the DE ("D") that accompanies the middle SE (#8; see Figs. 1A, C). Wild-type (*brk789wt*) and DE-mutant (*brk789DEm*) versions of the fragment are shown. B–H: Expression in late third-instar imaginal disc tissue of GFP reporters driven by either the wild-type (B–E; four independent transgene insertions) or the DE-mutant (F–H; three independent transgene insertions) version of the 7/8/9 module (see A). All transgenes are present in one copy, inserted into the attP2 docking site (see Materials and methods). Mutation of the DE results in severe reduction of the GFP reporter signal.

(Duncan and Dearden, 2010). We emphasize that, because of their low complexity, we cannot argue strongly for or against the conservation of the P and E sites that we have indicated as part of the shared architectures (see, however, Fig. S2), but we suggest that the highly constrained SPS is much more likely to have been conserved than repeatedly evolved anew.

Brd family genes have not been found in deuterostomes, but *Hes* repressor genes are widespread among metazoans, being present even in the placozoan *Trichoplax adherens* (see Discussion and Fig. 6). As in the arthropods, we find that an orthologous set of these Notch-regulated genes is associated in deuterostomes with the presence of three specific upstream cis-regulatory motifs (Fig. 4A; highlighted in red). First, an SPS element is found immediately upstream of the transcription start site. Next, at various distances upstream of the SPS, a predicted high-affinity binding site for bHLH repressors themselves (an "R" site) occurs. Finally, yet further upstream but often near the R site is a novel motif we refer to as the X element (XE) (Figs. 4B, C). As in the case of arthropod *Hes* repressor and *Brd* family genes, additional Su(H) lone sites may be present, but they do not show long-term conservation.

A number of lines of evidence support the interpretation that this shared cis-regulatory motif configuration upstream of deuterostome *Hes* repressor genes reflects long-term evolutionary conservation of an ancestral architecture. First, the genes themselves are generally unambiguous orthologs, so direct comparison of their putative regulatory motifs is valid and informative. Second, the nearly identical positioning of the SPS motif with respect to the transcription start sites of these genes is strongly suggestive of conservation. The frequent proximity of the X and R sites, as well as the common sequential order of the three motifs (X, R, SPS) is likewise consistent with this interpretation. Finally, strong conservation of the sequences of the various motifs, and even their flanking sequences, clearly suggests that the individual elements are orthologous. Thus, the various R sites denoted in Fig. 4A are identical in 10/10 positions (GGCACGTGCT), despite the fact that several variants of this motif are compatible with high-affinity binding by Hes repressor proteins (Jennings et al., 1999; Rebeiz et al., 2005; Van Doren et al., 1994). Even more strikingly, we observe strong sequence identity of the SPS elements and flanking sequences in various deuterostomes (Fig. 4D).

Conservation of "P+SPS" cis-regulatory function in Diptera

The foregoing analysis establishes the long-term evolutionary conservation of particular "P + SPS" motif combinations associated with Notch pathway target genes in arthropods. To investigate whether the functional properties of this cis-regulatory architecture are likewise conserved, we tested the behavior of a 1.0-kb non-coding DNA fragment from upstream of the *A. gambiae* (*Ag*) *BFM* gene (Schlatter and Maier, 2005) in a reporter assay in transgenic *Drosophila* (Fig. 5). As shown and described above, this region contains only a single P site and an SPS (Fig. 5A). When placed upstream of a minimal *Hsp70* promoter and an eGFP reporter gene in the pH-Stinger vector (Barolo et al., 2004), this fragment successfully recapitulates the specificity of previously studied Notch-regulated cis-regulatory modules from *Drosophila BFM* genes (Bailey and Posakony, 1995; Castro et al., 2005; Lai et al., 2000); i.e., it directs expression selectively in the non-SOP cells of proneural clusters (Figs. 5B–D). As in the case of the *Drosophila* enhancers, this expression is dependent on the integrity of the lone P site, since reporter gene expression is virtually abolished when this motif is mutated (Pm; Fig. 5E). Also mimicking the fly enhancers, two major effects are observed when the two Su(H) sites in the *Ag BFM* fragment's SPS are mutated: expression in non-SOP

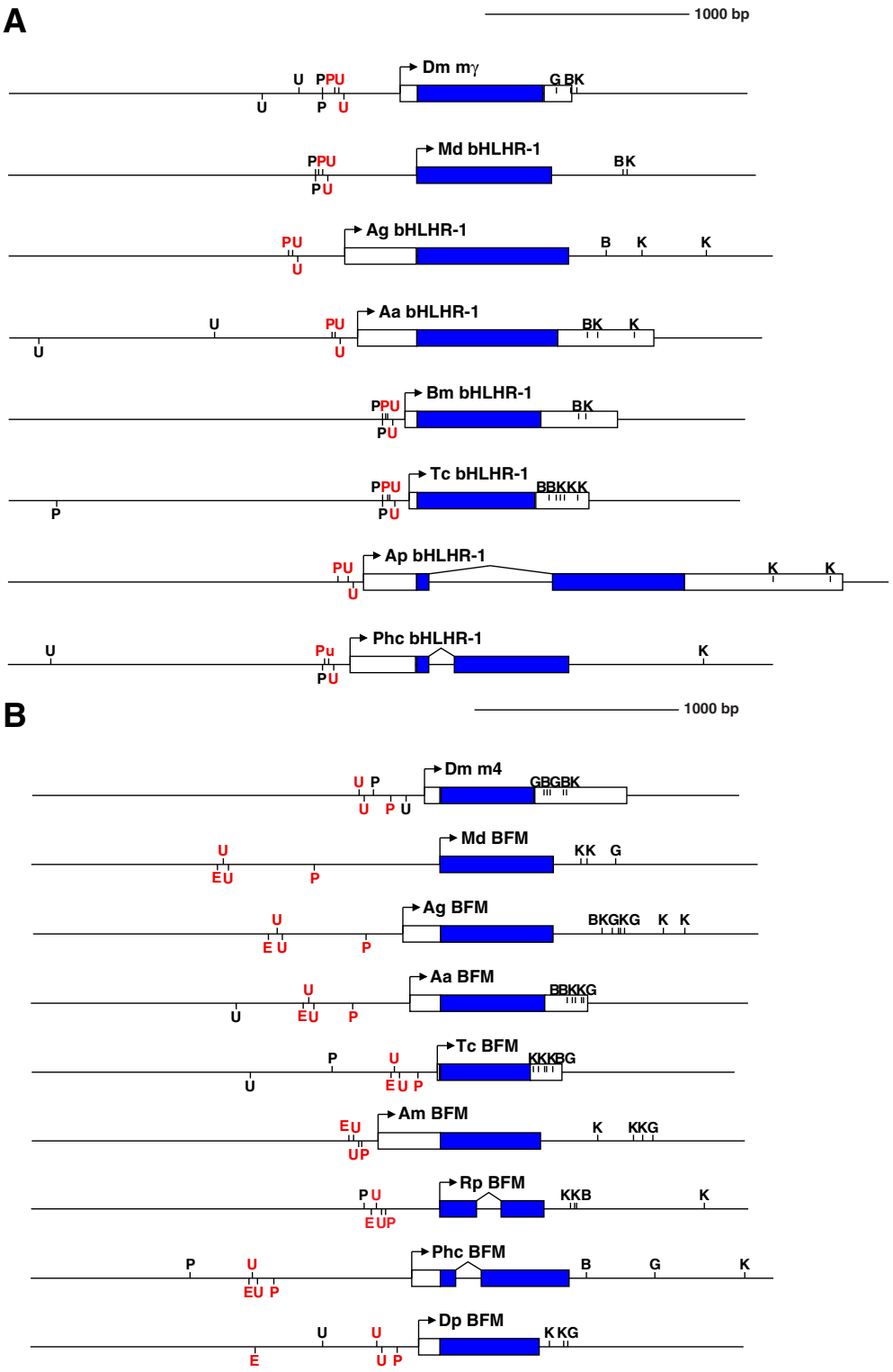


Fig. 3. Evolutionary conservation of "P + SPS" motif combinations upstream of orthologous arthropod *Hes* bHLH repressor and *Brd* family genes. A: Diagrams of orthologous *Hes*-class bHLH repressor genes located in the E(spl)-C of various insect species; these belong to the "E(spl)-C bHLH-1" clade described by Duncan and Dearden (2010), and are labeled here as "bHLHR-1" for clarity. B: Diagrams of orthologous *Brd* family genes located in the E(spl)-C of various arthropod species. Blue boxes represent protein coding sequences; white boxes represent untranslated regions (UTRs). Arrows denote direction of transcription and are positioned at the transcription start site of the gene (or at the start codon in the case of *Md* and *Rp* genes). Genes are aligned on the start codon. High-affinity Su(H) binding sites (YGTGDGAA) are indicated by "U"; Achaete/Scute-type proneural protein binding sites (RCAGSTG) are denoted by "P"; an extended E box motif identified by MEME (RRCAGATGGY) upstream of *Brd* family genes (see B) is represented by "E". Conserved P + SPS (A) and E + SPS + P (B) motif combinations are shown in red. In B, note the absence of the E motif in *Dm m4* (and in the ortholog in other Brachyceran flies), and its presence on the "upper" instead of the "lower" strand in *Am BFM* (and in the ortholog in other Hymenopterans). Lower-case "u" in the SPS upstream of *Phc bHLHR-1* (see A) denotes single-base mismatch to the U motif definition (CATGGGAA); Su(H) binds this site with somewhat reduced affinity (Nellesen et al., 1999). B, G, and K symbols in 3′ UTRs represent Brd box, GY box, and K box miRNA binding motifs (2005; Lai and Posakony, 1997, Lai et al., 1998). Species symbols are listed in the legend to Fig. 1A, except *Md*, *Mayetiola destructor* (Diptera); *Phc*, *Pediculus humanus corporis* (Phthiraptera); *Dp*, *Daphnia pulex* (Crustacea). Sequence scale is shown in upper left corner of each panel.

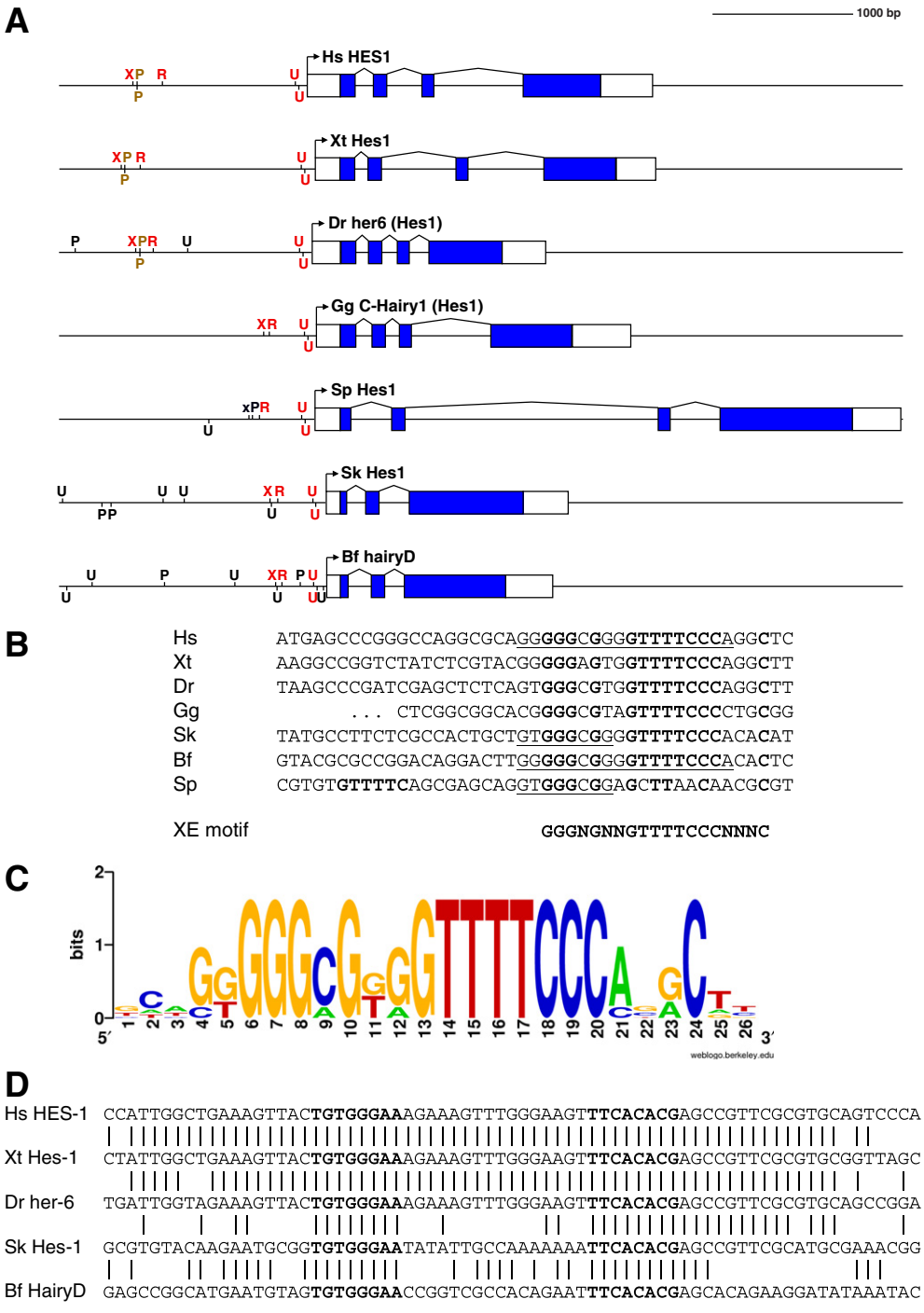


Fig. 4. Evolutionary conservation of an "X + R + SPS" motif combination upstream of *Hes1* bHLH repressor gene orthologs in various deuterostomes. A: Diagrams of *Hes1* orthologs in representative deuterostomes. Blue boxes represent protein coding sequences; white boxes represent untranslated regions. Arrows denote direction of transcription and are positioned at the transcription start site of the gene. Genes are aligned on the start codon. High-affinity Su(H)/CBF1 binding sites (YGTGDGAA) are indicated by "U"; Achaete/Scute-type proneural protein binding sites (RCAGSTG) are denoted by "P"; high-affinity bHLH repressor binding sites (GGCACGYGHY) are designated "R"; the novel X Element motif (see B, C) is also indicated (X). Only a truncated version of the X Element is present in *Sp*, indicated by a lowercase "x". Conserved X + R + SPS motif combinations are shown in red. Note also the conservation of a palindromic P site (brown) located between the X and R sites in mammals, amphibians, and fish. *Hs*, *Homo sapiens* (Chordata; Mammalia); *Xt*, *Xenopus tropicalis* (Chordata; Amphibia); *Dr*, *Danio rerio* (Chordata; Actinopteryggii); *Gg*, *Gallus gallus* (Chordata; Aves); *Sp*, *Strongylocentrotus purpuratus* (Echinodermata; Echinoidea); *Sk*, *Saccoglossus kowalevskii* (Hemichordata; Enteropneusta); *Bf*, *Branchiostoma floridae* (Cephalochordata). The *Bf* genome has undergone a large paralogous expansion of the *Hes* repressor gene family (Minguillon et al., 2003); the *hairyD* gene is shown here. Sequence scale is shown in upper left corner. B: Sequence alignment of the novel "X Element" motif upstream of deuterostome *Hes1* genes; strictly conserved nucleotides are shown in bold. Note the 18/18 sequence identity between *Hs* and *Bf* (underlined). As shown, *Sp* has a strong partial match to this motif in the corresponding location (see A); note that the missing component GTTTTC occurs a short distance upstream. Species symbols as in A. C: Logo plot corresponding to the XE motif alignment shown in B (*Sp* omitted). D: Sequence alignment of the *Hes1* SPS motif and flanking regions from *Hs*, *Xt*, *Dr*, *Sk*, and *Bf*, suggesting orthology between these elements in deuterostomes. Su(H)/CBF1 binding sites comprising the SPS are shown in bold; vertical bars indicate sequence identity. Note exceptionally strong sequence conservation between mammals, amphibians, and fish. Sequences of the two 8-bp Su(H)/CBF1 binding sites are strictly conserved in all seven species (not shown; see A). Species symbols as in A.

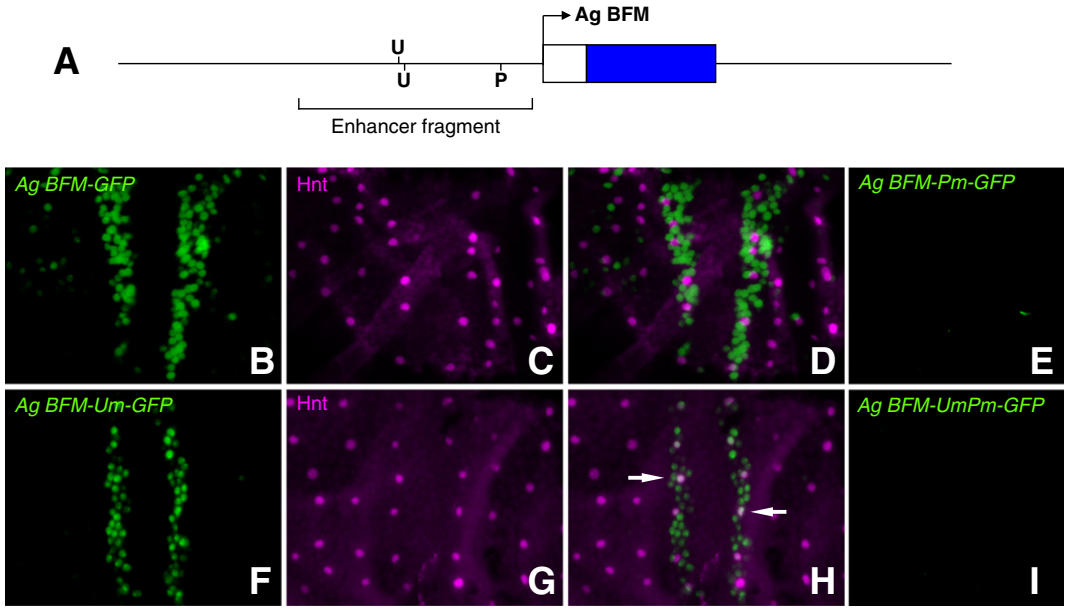


Fig. 5. Conservation of the regulatory activity of the SPS + P motif combination between *Anopheles* and *Drosophila*. A: Diagram of the lone member of the *Brd* gene family (*BFM*) in the mosquito *Anopheles gambiae* (*Ag*); shown also in Fig. 3B. The indicated upstream region (1.0 kb), which includes the conserved SPS + P motif combination, was tested for enhancer activity using a GFP reporter transgene in *Drosophila* (*Ag BFM-GFP*). The two Su(H) binding sites in the SPS are indicated by "U"; the lone proneural protein binding site is labeled "P". B–I: Images of pupal thoraces at 14 hours APF, centered on the midline; anterior is toward the top. GFP expressed by reporter transgenes is shown in green in B, D, E, F, H, and I. B: The wild-type *Ag BFM* upstream fragment drives reporter gene expression in the two microchaete "proneural rows" flanking the thoracic midline. Note "holes" in the GFP pattern. C, G: Microchaete SOPs are labeled with anti-Hindsight (Hnt) antibody (magenta). D: Merge of B and C, showing exclusion of reporter activity from SOPs. E: Mutation of the lone P site in the enhancer (*Ag BFM-Pm-GFP*) extinguishes expression. F: Mutation of the two Su(H) binding sites in the SPS (*Ag BFM-Um-GFP*) reduces expression in the non-SOPs of the proneural rows. H: Merge of F and G reveals ectopic expression of *Ag BFM-Um-GFP* in SOPs (white arrows). I: Both residual non-SOP and ectopic SOP expression displayed by *Ag BFM-Um-GFP* (see F, H) is dependent on the lone P site, as shown by the lack of activity of the triple-mutant construct *Ag BFM-UmPm-GFP*.

cells is greatly reduced, and ectopic expression in SOPs appears (Um; Figs. 5F–H). Finally, the ectopic SOP activity of the Um fragment is fully dependent on the P site, as this expression (along with the residual non-SOP expression) is lost in the UmPm triple site mutant (Fig. 5I).

We conclude that the *Ag BFM* non-coding DNA fragment, bearing a single P site and a single SPS, does indeed encompass an enhancer module that exhibits all of the key regulatory properties of similar modules from both *BFM* and *Hes* Notch pathway target genes in *Drosophila* (Castro et al., 2005): it directs expression specifically in non-SOP cells of proneural clusters; it requires activating inputs from both proneural proteins and Su(H); and it mediates "default repression" by Su(H) in SOP cells. Thus, not only has the core "P + SPS" architecture of the module been conserved for the ~235 My separating the fly and the mosquito, but so have its function and cis-regulatory logic.

Discussion

Deep origin and long-term evolutionary conservation of specific cis-regulatory motifs in developmental control genes

We have previously described the phylogenetically widespread occurrence of single, high-affinity bHLH repressor (R) binding sites upstream of bilaterian proneural genes (Rebeiz et al., 2005). We noted that we could not rule out the possibility that only the "linkage" (direct transcription factor-target gene relationship) has been maintained, and that the binding site itself has been replaced repeatedly in the course of animal evolution. However, we pointed to several lines of evidence suggesting that these R sites have been conserved from a deep common ancestor. These included the stability of the precise 10-bp sequence of the site over very long intervals, and the strong conservation of both the motif and flanking sequences in some instances, clearly suggesting that the sites are indeed orthologous.

The present report substantially expands the inventory of such apparently ancient and conserved cis-regulatory motifs in developmental control genes. We have described here five additional cases in which specific motif combinations have evidently been retained over hundreds of millions of years of evolution. With the exception of two novel elements (the insect *brk* DE and the deuterostome *Hes* XE), these motifs represent high-affinity binding sites for known transcription factors. The retention of these specific motif instances is especially striking when considered against the background of rapid appearance and disappearance of other binding sites for the same factors (Figs. 1A, 3A–B, 4A, S3B).

The conservation of the distinctive SE + DE motif combination upstream of insect *brk* genes extends over perhaps 270–300 My, reflecting the fact that the *brk* gene itself is found only in insects (Copley, 2008). A similar (minimum) age can be assigned to the P + SPS architecture found upstream of insect bHLH repressor genes, while the E + SPS + P combination associated with arthropod *BFM* genes is even older, in excess of 400 My, in view of its occurrence in the crustacean *D. pulex*. Finally, it is likely that the X + R + SPS ensemble upstream of deuterostome *Hes1* genes was present in the common ancestor, over 500 My ago. It is also possible that an SPS element was associated with an ancestral bilaterian *Hes* repressor gene, which would make this feature close to 600 My old.

Our analyses do not permit us to discern the population genetic/microevolutionary processes by which the distinctive cis-regulatory architectures we describe first arose and became fixed in an ancestral population (Lynch, 2007). However, we believe we can offer some useful insights into why these architectures have endured over such lengthy timescales.

Distinctive regulatory capabilities mediated by deeply conserved cis-regulatory motifs

What characteristics of ancient and conserved motifs drive their long-term preservation by selection, even as other binding sites for the same factors come and go rapidly in evolution? We first reiterate our earlier proposal that such deeply conserved motifs mediate

abstract or generic regulatory functions of fundamental utility to all or most members of an ancient clade (Rebeiz et al., 2005). It is certainly plausible that, once established, the capacity to repress *brk* transcription in response to a Dpp signal remained of great utility to all the descendants of the common insect ancestor, as diverse as they became. Similarly, the abstract ability to activate a *Hes* repressor gene via Notch signaling would remain of exceptional utility to descendants of a bilaterian (or earlier) ancestor that had evolved it. Finally, a generic capability for autorepression of a *Hes* bHLH repressor gene (Brend and Holley, 2009; Hirata et al., 2002; Lewis, 2003) might very well be retained by descendants of a deuterostome ancestor.

But it is certainly sensible to argue that, to retain such abstract and valuable regulatory capabilities, it would suffice to preserve only the linkage between the appropriate transcription factors and their targets. In this view, individual factor-binding motifs need not be retained; they would be free to turn over during evolution. However, the examples we have described here suggest a second important reason for the long-term evolutionary retention of particular motifs or motif combinations. We propose that these conserved sequence elements mediate a distinctive regulatory capability not conferred by other instances of the same motif or motifs. In the case of the SPS element, we can be quite confident that this perspective is correct. The SPS has been shown to mediate cooperative binding of two Su(H)/Mam/NICD trimers, thus conferring on the associated target gene unusually high sensitivity to Notch signaling (Arnett et al., 2010; Nam et al., 2007). While two "lone" Su(H) sites are indeed able to contribute to a target gene's response to activated Notch, they would not do so in a cooperative manner. In a similar vein, it seems plausible to suggest that while all SE motifs may be able to participate in signal-dependent repression of *brk*, the SE + DE combination offers a unique and valuable version of this capability (e.g., greater signal sensitivity), possibly conferring a fitness advantage. We hypothesize that in both cases, once the specialized motif architecture (SPS or SE + DE) had evolved to confer a distinctive capacity, it would be selectively retained in evolution. As we have seen, other instances of the SE or Su(H) binding motifs do arise and become fixed in individual clades, but these would not be expected to exhibit the same durability, since (according to the hypothesis) they confer no unique capability. The foregoing interpretation is particularly supported, we believe, by the frequent observation that if only one element mediating a particular response [either SE or Su(H) site] is present upstream of an orthologous gene in a given species, it is of the "special" type (SE + DE or SPS). Examples include the SE + DE combination in *T. castaneum brk* and the SPS motifs in the *A. gambiae bHLHR1* gene, the *A. mellifera BFM* gene, and *H. sapiens HES1*.

Another factor that may contribute to the long-term evolutionary conservation of the specialized motif architectures we have considered is their very complexity. Both the SE + DE unit and the SPS represent unusually extended and constrained motif combinations. While in principle this does not prevent them from turning over by duplication/degeneration, they are unlikely to evolve de novo.

Finally, we note an intriguing feature of the conserved motif architectures described here that involve the SPS: the apparently conserved order and even orientation of the individual sequence elements. The arthropod *BFM* genes are associated with a "lower-strand" E motif followed by an SPS followed by a "lower-strand" P site; insect *Hes* repressor genes bear an "upper-strand" P site followed by an SPS; and deuterostome *Hes1* genes have an "upper-strand" X site followed by an "upper-strand" R site followed by an SPS, which also has fixed orientation. Inter-site distances are often not conserved; consider the varying separation of the SPS and the P site in the *BFM* genes, or the different distances between the X + R combination and the SPS in the deuterostome *Hes1* genes. Evidently, the motif order and orientation of these architectures have functional significance, consistent with an "enhanceosome" model for the structure of these regions (Arnosti and Kulkarni, 2005). Alternatively, these features may suggest the existence of a "scanning" mechanism for optimal enhancer-promoter interaction. Such a property might be a particular characteristic of promoter-proximal cis-regulatory modules such as these, as contrasted with

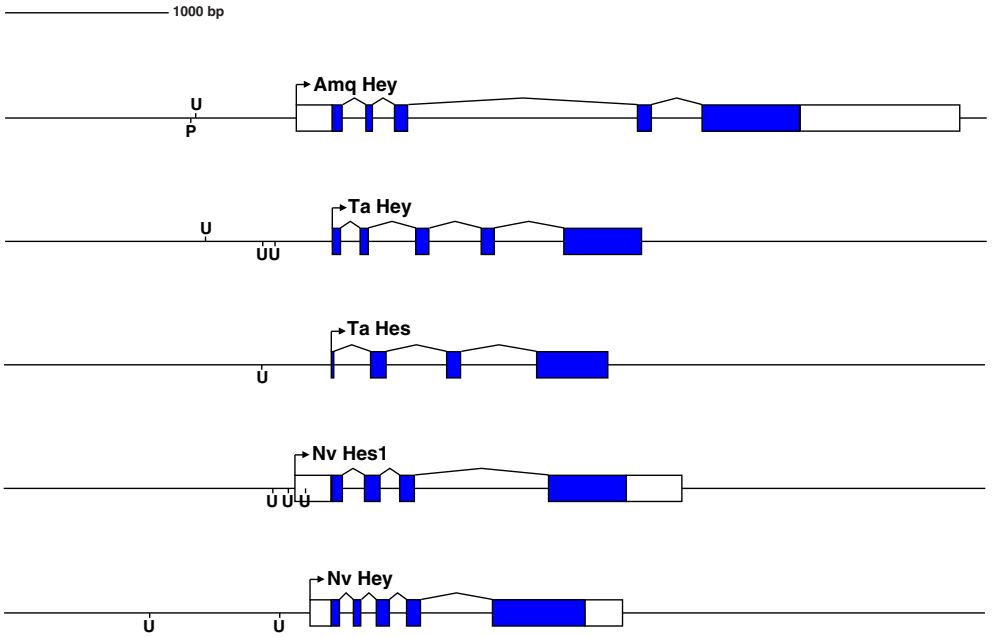


Fig. 6. Direct regulation of *Hey*- and *Hes*-class bHLH repressor genes by Su(H) is apparently ancient. Shown are diagrams of *Hey* and *Hes* genes in three non-bilaterians. Blue boxes represent protein coding sequences; white boxes represent untranslated regions. Arrows denote direction of transcription and are positioned at either the transcription start site or the start codon of the gene. Genes are aligned on the start codon. High-affinity Su(H) binding sites (YGTGDGAA) are indicated by "U"; Achaete/Scute-type proneural protein binding sites (RCAGSTG) are denoted by "P". Sequence scale is shown in upper left corner. The demosponge *Amphimedon queenslandica* (*Amq*) has a *Hey* gene, but no *Hes* genes (Simionato et al., 2007; Srivastava et al., 2010). The placozoan *Trichoplax adhaerens* (*Ta*) has one *Hey* ortholog, one *Hey*-related gene (not shown), and one *Hes* gene (Srivastava et al., 2008). The cnidarian *Nematostella vectensis* (*Nv*) has 11 *Hes* genes (Putnam et al., 2007); a representative example is shown [this corresponds to *Nem52* (Simionato et al., 2007)]. *Nv* also has one *Hey* ortholog (shown) and one *Hey*-related gene (not shown). Note conservation of exon-intron structure between *Hey* and *Hes* genes, respectively. High-affinity Su(H) sites occur in the proximal upstream regions of all of these genes, but no SPSs are found. Significantly, the three-Su(H)-site configuration shown for *Nv Hes1* is conserved in the orthologous gene of the distantly related anthozoan *Acropora digitifera* (stony coral) (see Fig. S3).

more distal enhancers. In the latter case, interaction with the promoter by "looping" may impose fewer architectural constraints.

Evolution and conservation of distinctive developmental regulatory capabilities

We have proposed here that the distinctive cis-regulatory architectures we describe are ancient ones that have been conserved from a deep ancestor. However, it also seems likely that, because of their very complexity, they may not represent the "original" version of their respective regulatory linkages. We believe that these two realizations can be reconciled via the following general evolutionary scenario.

The direct linkage of an ancestral *Hes* gene to Su(H) and the Notch pathway evidently originated in a deep metazoan ancestor, and was very likely mediated by a lone Su(H) binding site or sites. The genome of the demosponge *A. queenslandica* includes one member of the closely related *Hey* repressor family, but no *Hes* genes (Simionato et al., 2007; Srivastava et al., 2010); this *Amphimedon Hey* gene has one high-affinity Su(H) site 600 bp upstream of the transcription start site (Fig. 6). The placozoan *T. adhaerens* has one *Hey* ortholog, one *Hey*-related gene, and one *Hes* gene (Srivastava et al., 2008). The *Hey* ortholog has three high-affinity Su(H) sites in the first 800 bp upstream of the ATG start codon, while the *Hes* gene includes a single such site within 500 bp of its ATG (Fig. 6). The genome of the cnidarian *N. vectensis* (sea anemone) is endowed with a large paralogous family of 11 *Hes* genes (Putnam et al., 2007; Simionato et al., 2007), many of them with multiple lone Su(H) sites immediately upstream (Fig. 6; see also Fig. S3). Likewise, the *Nematostella Hey* ortholog has two upstream Su(H) sites. The SPS evidently did not appear upstream of a *Hey*/*Hes* gene until after the cnidarian–bilaterian divergence, but as we have seen, this association is now widespread among both protostomes and deuterostomes.

We suggest, then, that what appeared first was the simple capacity to regulate a *Hey*/*Hes* gene directly by Su(H) (presumably linked to the Notch pathway), via one or more lone Su(H) binding sites. Then, in a bilaterian ancestor, an SPS came into being upstream of an individual *Hes* gene, making possible a cooperative and thus highly sensitive response to Notch-activated Su(H). Once this novel regulatory capacity was established, it bestowed a sufficient selective advantage to ensure its subsequent retention in a wide variety of bilaterian taxa. Such a scenario can account for the phylogenetic distribution of the SPS-containing cis-regulatory architectures we have described. We cannot, however, rule out more complex histories, including the possibility that the SPS arose independently more than once in association with *Hes* genes.

Duplication–divergence of developmental control genes and their cis-regulatory architectures

It is important to note our finding that, in the case of target genes that are part of paralogous families (*Hes* repressor and *BFM*s), only one particular paralog in a given species is typically associated with the conserved motif architectures we have described. This is true even if other paralogs make use of the same overall cis-regulatory "code" (combination of transcription factor binding sites) to direct a similar expression specificity. For example, of the seven unambiguous *Hes* repressor paralogs in *H. sapiens* (Simionato et al., 2007), only *HES1* bears the X + R + SPS motif combination, though four others have upstream S sites and two of these also have upstream R sites. Likewise, the *D. melanogaster* genome includes nine *BFM* genes (Lai et al., 2000), most of which employ the S + P code, but only one, *E(spl)m4*, is associated with an SPS + P combination (Bailey and Posakony, 1995; Singson et al., 1994). It seems likely that, while the distinctive regulatory capability conferred by an ancient and conserved motif combination is of long-term selective value, it suffices for a single paralog in the genome to retain it.

This observation is consistent with a duplication–divergence model for the evolution of *Hes* and *BFM* paralogs. The special cis-regulatory architectures we have described, along with the associated protein coding sequences, comprise functional units that have been conserved from deep common ancestors because (we propose) of the unique regulatory capabilities they confer. Paralogous genes that arise by duplication within various taxa (this is a widespread phenomenon in the case of *Hes* genes) would not be subject to the same stringent constraints on their cis-regulatory architecture, since the ancestral gene would be present to provide the distinctive capabilities. The paralogs would thus be free to evolve their cis-regulatory motifs according to other selective pressures or genetic drift (Brown et al., 2007), yielding the many variations on a basic theme (e.g., S + P) that we observe within a single species today.

Acknowledgments

We are grateful to Steve Miller for providing his *attB-H-Stinger* GFP reporter vector, for permitting us to describe unpublished results, and for valuable comments on the manuscript. We thank Scott Rifkin for his very useful suggestions for improving the Discussion. The Developmental Studies Hybridoma Bank provided antibodies; the Bloomington Drosophila Stock Center provided fly stocks. This work was supported by NIH grant R01 GM046993 to J.W.P.

Appendix A. Supplementary data

Supplementary data to this article can be found online at doi:10.1016/j.ydbio.2011.12.011.

References

Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Scherer, S.E., Li, P.W., Hoskins, R.A., Galle, R.F., George, R.A., Lewis, S.E., Richards, S., Ashburner, M., Henderson, S.N., Sutton, G.G., Wortman, J.R., Yandell, M.D., Zhang, Q., Chen, L.X., Brandon, R.C., Rogers, Y.H., Blazej, R.G., Champe, M., Pfeiffer, B.D., Wan, K.H., Doyle, C., Baxter, E.G., Helt, G., Nelson, C.R., Gabor Miklos, G.L., Abril, J.F., Agbayani, A., An, H.J., Andrews-Pfannkoch, C., Baldwin, D., Ballew, R.M., Basu, A., Baxendale, J., Bayraktaroglu, L., Beasley, E.M., Beeson, K.Y., Benos, P.V., Berman, B.P., Bhandari, D., Bolshakov, S., Borkova, D., Botchan, M.R., Bouck, J., Brokstein, P., Brottier, P., Burtis, K.C., Busam, D.A., Butler, H., Cadieu, E., Center, A., Chandra, I., Cherry, J.M., Cawley, S., Dahlke, C., Davenport, L.B., Davies, P., de Pablos, B., Delcher, A., Deng, Z., Mays, A.D., Dew, I., Dietz, S.M., Dodson, K., Doup, L.E., Downes, M., Dugan-Rocha, S., Dunkov, B.C., Dunn, P., Durbin, K.J., Evangelista, C.C., Ferraz, C., Ferriera, S., Fleischmann, W., Fosler, C., Gabrielian, A.E., Garg, N.S., Gelbart, W.M., Glasser, K., Glodek, A., Gong, F., Gorrell, J.H., Gu, Z., Guan, P., Harris, M., Harris, N.L., Harvey, D., Heiman, T.J., Hernandez, J.R., Houck, J., Hostin, D., Houston, K.A., Howland, T.J., Wei, M.H., Ibegwam, C., Jalali, M., Kalush, F., Karpen, G.H., Ke, Z., Kennison, J.A., Ketchum, K.A., Kimmel, B.E., Kodira, C.D., Kraft, C., Kravitz, S., Kulp, D., Lai, Z., Lasko, P., Lei, Y., Levitsky, A.A., Li, J., Li, Z., Liang, Y., Lin, X., Liu, X., Mattei, B., McIntosh, T.C., McLeod, M.P., McPherson, D., Merkulov, G., Milshina, N.V., Mobarry, C., Morris, J., Moshrefi, A., Mount, S.M., Moy, M., Murphy, B., Murphy, L., Muzny, D.M., Nelson, D.L., Nelson, D.R., Nelson, K.A., Nixon, K., Nusskern, D.R., Pacleb, J.M., Palazzolo, M., Pittman, G.S., Pan, S., Pollard, J., Puri, V., Reese, M.G., Reinert, K., Remington, K., Saunders, R.D., Scheeler, F., Shen, H., Shue, B.C., Siden-Kiamos, I., Simpson, M., Skupski, M.P., Smith, T., Spier, E., Spradling, A.C., Stapleton, M., Strong, R., Sun, E., Svirskas, R., Tector, C., Turner, R., Venter, E., Wang, A.H., Wang, X., Wang, Z.Y., Wassarman, D.A., Weinstock, G.M., Weissenbach, J., Williams, S.M., Woodage, T., Worley, K.C., Wu, D., Yang, S., Yao, Q.A., Ye, J., Yeh, R.F., Zaveri, J.S., Zhan, M., Zhang, G., Zhao, Q., Zheng, L., Zheng, X.H., Zhong, F.N., Zhong, W., Zhou, X., Zhu, S., Zhu, X., Smith, H.O., Gibbs, R.A., Myers, E.W., Rubin, G.M., Venter, J.C., 2000. The genome sequence of *Drosophila melanogaster*. Science 287, 2185–2196.

Affolter, M., Basler, K., 2007. The Decapentaplegic morphogen gradient: from pattern formation to growth regulation. Nat. Rev. Genet. 8, 663–674.

Arnett, K.L., Hass, M., McArthur, D.G., Ilagan, M.X., Aster, J.C., Kopan, R., Blacklow, S.C., 2010. Structural and mechanistic insights into cooperative assembly of dimeric Notch transcription complexes. Nat. Struct. Mol. Biol. 17, 1312–1317.

Arnosti, D.N., Kulkarni, M.M., 2005. Transcriptional enhancers: intelligent enhanceosomes or flexible billboards? J. Cell. Biochem. 94, 890–898.

Bailey, T.L., Elkan, C., 1994. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Second International Conference on Intelligent Systems for Molecular Biology, pp. 28–36.

Bailey, A.M., Posakony, J.W., 1995. Suppressor of Hairless directly activates transcription of *Enhancer of split Complex* genes in response to Notch receptor activity. Genes Dev. 9, 2609–2622.

Bardin, A.J., Schweisguth, F., 2006. Bearded family members inhibit Neuralized-mediated endocytosis and signaling activity of Delta in *Drosophila*. Dev. Cell 10, 245–255.

Barolo, S., Castro, B., Posakony, J.W., 2004. New *Drosophila* transgenic reporters: insulated P-element vectors expressing fast-maturing RFP. Biotechniques 36, 436–442.

Bischof, J., Maeda, R.K., Hediger, M., Karch, F., Basler, K., 2007. An optimized transgenesis system for *Drosophila* using germ-line-specific phiC31 integrases. Proc. Natl. Acad. Sci. U. S. A. 104, 3312–3317.

Bray, S.J., 2006. Notch signalling: a simple pathway becomes complex. Nat. Rev. Mol. Cell Biol. 7, 678–689.

Brend, T., Holley, S.A., 2009. Expression of the oscillating gene *her1* is directly regulated by Hairy/Enhancer of Split, T-box, and Suppressor of Hairless proteins in the zebrafish segmentation clock. Dev. Dyn. 238, 2745–2759.

Brown, C.D., Johnson, D.S., Sidow, A., 2007. Functional architecture and evolution of transcriptional elements that drive gene coexpression. Science 317, 1557–1560.

Castro, B., Barolo, S., Bailey, A.M., Posakony, J.W., 2005. Lateral inhibition in proneural clusters: cis-regulatory logic and default repression by Suppressor of Hairless. Development 132, 3333–3344.

Chan, Y.F., Marks, M.E., Jones, F.C., Villarreal, G.J., Shapiro, M.D., Brady, S.D., Southwick, A.M., Absher, D.M., Grimwood, J., Schmutz, J., Myers, R.M., Petrov, D., Jonsson, B., Schluter, D., Bell, M.A., Kingsley, D.M., 2010. Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a *Pitx1* enhancer. Science 327, 302–305.

Colbourne, J.K., Pfrender, M.E., Gilbert, D., Thomas, W.K., Tucker, A., Oakley, T.H., Tokishita, S., Aerts, A., Arnold, G.J., Basu, M.K., Bauer, D.J., Caceres, C.E., Carmel, L., Casola, C., Choi, J.H., Detter, J.C., Dong, Q., Dusheyko, S., Eads, B.D., Frohlich, T., Geiler-Samerotte, K.A., Gerlach, D., Hatcher, P., Jogdeo, S., Krijgsveld, J., Kriventseva, E.V., Kultz, D., Laforsch, C., Lindquist, E., Lopez, J., Manak, J.R., Muller, J., Pangilinan, J., Patwardhan, R.P., Pitluck, S., Pritham, E.J., Rechtsteiner, A., Rho, M., Rogozin, I.B., Sakarya, O., Salamov, A., Schaack, S., Shapiro, H., Shiga, Y., Skalitzky, C., Smith, Z., Souvorov, A., Sung, W., Tang, Z., Tsuchiya, D., Tu, H., Vos, H., Wang, M., Wolf, Y.I., Yamagata, H., Yamada, T., Ye, Y., Shaw, J.R., Andrews, J., Crease, T.J., Tang, H., Lucas, S.M., Robertson, H.M., Bork, P., Koonin, E.V., Zdobnov, E.M., Grigoriev, I.V., Lynch, M., Boore, J.L., 2011. The ecoresponsive genome of *Daphnia pulex*. Science 331, 555–561.

International Aphid Genomics Consortium, 2010. Genome sequence of the pea aphid *Acyrthosiphon pisum*. PLoS Biol. 8, e1000313.

Copley, R.R., 2008. The animal in the genome: comparative genomics and evolution. Philos. Trans. R. Soc. Lond. B Biol. Sci. 363, 1453–1461.

Crooks, G.E., Hon, G., Chandonia, J.M., Brenner, S.E., 2004. WebLogo: a sequence logo generator. Genome Res. 14, 1188–1190.

Davidson, E.H., 2006. The regulatory genome. Gene Regulatory Networks in Development and Evolution. Academic Press, San Diego.

Duncan, E.J., Dearden, P.K., 2010. Evolution of a genomic regulatory domain: the role of gene co-option and gene duplication in the Enhancer of split complex. Genome Res. 20, 917–928.

Fiuza, U.M., Arias, A.M., 2007. Cell and molecular biology of Notch. J. Endocrinol. 194, 459–474.

Fontana, J.R., Posakony, J.W., 2009. Both inhibition and activation of Notch signaling rely on a conserved Neuralized-binding motif in Bearded proteins and the Notch ligand Delta. Dev. Biol. 333, 373–385.

Gao, S., Steffen, J., Laughon, A., 2005. Dpp-responsive silencers are bound by a trimeric Mad–Medea complex. J. Biol. Chem. 280, 36158–36164.

Gompel, N., Prud'homme, B., Wittkopp, P.J., Kassner, V.A., Carroll, S.B., 2005. Chance caught on the wing: *cis*-regulatory evolution and the origin of pigment patterns in *Drosophila*. Nature 433, 481–487.

Hare, E.E., Peterson, B.K., Iyer, V.N., Meier, R., Eisen, M.B., 2008. Sepsid *even-skipped* enhancers are functionally conserved in *Drosophila* despite lack of sequence conservation. PLoS Genet. 4, e1000106.

Hellsten, U., Harland, R.M., Gilchrist, M.J., Hendrix, D., Jurka, J., Kapitonov, V., Ovcharenko, I., Putnam, N.H., Shu, S., Taher, L., Blitz, I.L., Blumberg, B., Dichmann, D.S., Dubchak, I., Amaya, E., Detter, J.C., Fletcher, R., Gerhard, D.S., Goodstein, D., Graves, T., Grigoriev, I.V., Grimwood, J., Kawashima, T., Lindquist, E., Lucas, S.M., Mead, P.E., Mitros, T., Ogino, H., Ohta, Y., Poliakov, A.V., Pollet, N., Robert, J., Salamov, A., Sater, A.K., Schmutz, J., Terry, A., Vize, P.D., Warren, W.C., Wells, D., Wills, A., Wilson, R.K., Zimmerman, L.B., Zorn, A.M., Grainger, R., Grammer, T., Khokha, M.K., Richardson, P.M., Rokhsar, D.S., 2010. The genome of the Western clawed frog *Xenopus tropicalis*. Science 328, 633–636.

Hirata, H., Yoshiura, S., Ohtsuka, T., Bessho, Y., Harada, T., Yoshikawa, K., Kageyama, R., 2002. Oscillatory expression of the bHLH factor Hes1 regulated by a negative feedback loop. Science 298, 840–843.

Holt, R.A., Subramanian, G.M., Halpern, A., Sutton, G.G., Charlab, R., Nusskern, D.R., Wincker, P., Clark, A.G., Ribeiro, J.M., Wides, R., Salzberg, S.L., Loftus, B., Yandell, M., Majoros, W.H., Rusch, D.B., Lai, Z., Kraft, C.L., Abril, J.F., Anthouard, V., Arensburger, P., Atkinson, P.W., Baden, H., de Berardinis, V., Baldwin, D., Benes, V., Biedler, J., Blass, C., Bolanos, R., Boscus, D., Barnstead, M., Cai, S., Center, A., Chaturverdi, K., Christophides, G.K., Chrystal, M.A., Clamp, M., Cravchik, A., Curwen, V., Dana, A., Delcher, A., Dew, I., Evans, C.A., Flanigan, M., Grundschober-Freimoser, A., Friedli, L., Gu, Z., Guan, P., Guigo, R., Hillenmeyer, M.E., Hladun, S.L., Hogan, J.R., Hong, Y.S., Hoover, J., Jaillon, O., Ke, Z., Kodira, C., Kokoza, E., Koutsos, A., Letunic, I., Levitsky, A., Liang, Y., Lin, J.J., Lobo, N.F., Lopez, J.R., Malek, J.A., McIntosh, T.C., Meister, S., Miller, J., Mobarry, C., Mongin, E., Murphy, S.D., O'Brochta, D.A., Pfannkoch, C., Qi, R., Regier, M.A., Remington, K., Shao, H., Sharakhova, M.V., Sitter, C.D., Shetty, J., Smith, T.J., Strong, R., Sun, J., Thomasova, D., Ton, L.Q., Topalis, P., Tu, Z., Unger, M.F., Walenz, B., Wang, A., Wang, J., Wang, M., Wang, X., Woodford, K.J., Wortman, J.R., Wu, M., Yao, A., Zdobnov, E.M., Zhang, H., Zhao, Q., Zhao, S., Zhu, S.C., Zhimulev, I., Coluzzi, M., della Torre, A., Roth, C.W., Louis, C., Kalush, F., Mural, R.J., Myers, E.W., Adams, M.D., Smith, H.O., Broder, S., Gardner, M.J., Fraser, C.M., Birney, E., Bork, P., Brey, P.T., Venter, J.C., Weissenbach, J., Kafatos, F.C., Collins, F.H., Hoffman, S.L., 2002. The genome sequence of the malaria mosquito *Anopheles gambiae*. Science 298, 129–149.

Honeybee Genome Sequencing Consortium, 2006. Insights into social insects from the genome of the honeybee *Apis mellifera*. Nature 443, 931–949.

International Chicken Genome Sequencing Consortium, 2004. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. Nature 432, 695–716.

International Silkworm Genome Consortium, 2008. The genome of a lepidopteran model insect, the silkworm *Bombyx mori*. Insect Biochem. Mol. Biol. 38, 1036–1045.

Jarriault, S., Brou, C., Logeat, F., Schroeter, E.H., Kopan, R., Israel, A., 1995. Signalling downstream of activated mammalian Notch. Nature 377, 355–358.

Jennings, B.H., Tyler, D.M., Bray, S.J., 1999. Target specificities of Drosophila Enhancer of split basic helix-loop-helix proteins. Mol. Cell. Biol. 19, 4600–4610.

Jeong, S., Rebeiz, M., Andolfatto, P., Werner, T., True, J., Carroll, S.B., 2008. The evolution of gene regulation underlies a morphological difference between two *Drosophila* sister species. Cell 132, 783–793.

Werren, J.H., Richards, S., Desjardins, C.A., Niehuis, O., Gadau, J., Colbourne, J.K., Werren, J.H., Richards, S., Desjardins, C.A., Niehuis, O., Gadau, J., Colbourne, J.K., Beukeboom, L.W., Desplan, C., Elsik, C.G., Grimmelikhuijzen, C.J., Kitts, P., Lynch, J.A., Murphy, T., Oliveira, D.C., Smith, C.D., van de Zande, L., Worley, K.C., Zdobnov, E.M., Aerts, M., Albert, S., Anaya, V.H., Anzola, J.M., Barchuk, A.R., Behura, S.K., Bera, A.N., Berenbaum, M.R., Bertossa, R.C., Bitondi, M.M., Bordenstein, S.R., Bork, P., Bornberg-Bauer, E., Brunain, M., Cazzamali, G., Chaboub, L., Chacko, J., Chavez, D., Childers, C.P., Choi, J.H., Clark, M.E., Claudianos, C., Clinton, R.A., Cree, A.G., Cristino, A.S., Dang, P.M., Darby, A.C., de Graaf, D.C., Devreese, B., Dinh, H.H., Edwards, R., Elango, N., Elhaik, E., Ermolaeva, O., Evans, J.D., Foret, S., Fowler, G.R., Gerlach, D., Gibson, J.D., Gilbert, D.G., Graur, D., Grunder, S., Hagen, D.E., Han, Y., Hauser, F., Hultmark, D., Hunter, H.C.t., Hurst, G.D., Jhangian, S.N., Jiang, H., Johnson, R.M., Jones, A.K., Junier, T., Kadowaki, T., Kamping, A., Kapustin, Y., Kechavarzi, B., Kim, J., Kim, J., Kiryutin, B., Koevoets, T., Kovar, C.L., Kriventseva, E.V., Kucharski, R., Lee, H., Lee, S.L., Lees, K., Lewis, L.R., Loehlin, D.W., Logsdon, J.M.J., Lopez, J.A., Lozado, R.J., Maglott, D., Maleszka, R., Mayampurath, A., Mazur, D.J., McClure, M.A., Moore, A.D., Morgan, M.B., Muller, J., Munoz-Torres, M.C., Muzny, D.M., Nazareth, L.V., Neupert, S., Nguyen, N.B., Nunes, F.M., Oakeshott, J.G., Okwuonu, G.O., Pannebakker, B.A., Pejaver, V.R., Peng, Z., Pratt, S.C., Predel, R., Pu, L.L., Ranson, H., Raychoudhury, R., Rechtsteiner, A., Reese, J.T., Reid, J.G., Riddle, M., Robertson, H.M., Romero-Severson, J., Rosenberg, M., Sackton, T.B., Sattelle, D.B., Schluns, H., Schmitt, T., Schneider, M., Schuler, A., Schurko, A.M., Shuker, D.M., Simoes, Z.L., Sinha, S., Smith, Z., Solovyev, V., Souvorov, A., Springauf, A., Stafflinger, E., Stage, D.E., Stanke, M., Tanaka, Y., Telschow, A., Trent, C., Vattathil, S., Verhulst, E.C., Viljakainen, L., Wanner, K.W., Waterhouse, R.M., Whitfield, J.B., Wilkes, T.E., Williamson, M., Willis, J.H., Wolschin, F., Wyder, S., Yamada, T., Yi, S.V., Zecher, C.N., Zhang, L., Gibbs, R.A., 2010. Functional and evolutionary insights from the genomes of three parasitoid *Nasonia* species. Science 327, 343–348.

Lai, E.C., Posakony, J.W., 1997. The Bearded box, a novel 3′ UTR sequence motif, mediates negative post-transcriptional regulation of *Bearded* and *Enhancer of split* Complex gene expression. Development 124, 4847–4856.

Lai, E.C., Burks, C., Posakony, J.W., 1998. The K box, a conserved 3′ UTR sequence motif, negatively regulates accumulation of *Enhancer of split* Complex transcripts. Development 125, 4077–4088.

Lai, E.C., Bodner, R., Posakony, J.W., 2000. The *Enhancer of split* Complex of *Drosophila* includes four Notch-regulated members of the Bearded gene family. Development 127, 3441–3455.

Lai, E.C., Tam, B., Rubin, G.M., 2005. Pervasive regulation of Drosophila Notch target genes by GY-box-, Brd-box-, and K-box-class microRNAs. Genes Dev. 19, 1067–1080.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., Higgins, D.G., 2007. Clustal W and Clustal X version 2.0. Bioinformatics 23, 2947–2948.

Le Bras, S., Loyer, N., Le Borgne, R., 2011. The multiple facets of ubiquitination in the regulation of Notch signaling pathway. Traffic 12, 149–161.

Lecourtois, M., Schweisguth, F., 1995. The neurogenic Suppressor of Hairless DNA-binding protein mediates the transcriptional activation of the Enhancer of split Complex genes triggered by Notch signaling. Genes Dev. 9, 2598–2608.

Lewis, J., 2003. Autoinhibition with transcriptional delay: a simple mechanism for the zebrafish somitogenesis oscillator. Curr. Biol. 13, 1398–1408.

Ludwig, M.Z., Bergman, C., Patel, N.H., Kreitman, M., 2000. Evidence for stabilizing selection in a eukaryotic enhancer element. Nature 403, 564–567.

Lynch, M., 2007. The frailty of adaptive hypotheses for the origins of organismal complexity. Proc. Natl. Acad. Sci. U. S. A. 104 (Suppl. 1), 8597–8604.

Markstein, M., Zinzen, R., Markstein, P., Yee, K.P., Erives, A., Stathopoulos, A., Levine, M., 2004. A regulatory code for neurogenic gene expression in the Drosophila embryo. Development 131, 2387–2394.

Minguillon, C., Jimenez-Delgado, S., Panopoulou, G., Garcia-Fernandez, J., 2003. The amphioxus Hairy family: differential fate after duplication. Development 130, 5903–5914.

Mita, K., Kasahara, M., Sasaki, S., Nagayasu, Y., Yamada, T., Kanamori, H., Namiki, N., Kitagawa, M., Yamashita, H., Yasukochi, Y., Kadono-Okuda, K., Yamamoto, K., Ajimura, M., Ravikumar, G., Shimomura, M., Nagamura, Y., Shin-I, T., Abe, H., Shimada, T., Morishita, S., Sasaki, T., 2004. The genome sequence of silkworm, *Bombyx mori*. DNA Res. 11, 27–35.

Muller, B., Hartmann, B., Pyrowolakis, G., Affolter, M., Basler, K., 2003. Conversion of an extracellular Dpp/BMP morphogen gradient into an inverse transcriptional gradient. Cell 113, 221–233.

Nam, Y., Sliz, P., Pear, W.S., Aster, J.C., Blacklow, S.C., 2007. Cooperative assembly of higher-order Notch complexes functions as a switch to induce transcription. Proc. Natl. Acad. Sci. U. S. A. 104, 2103–2108.

Nellesen, D.T., Lai, E.C., Posakony, J.W., 1999. Discrete enhancer elements mediate selective responsiveness of *Enhancer of split* Complex genes to common transcriptional activators. Dev. Biol. 213, 33–53.

Nene, V., Wortman, J.R., Lawson, D., Haas, B., Kodira, C., Tu, Z.J., Loftus, B., Xi, Z., Megy, K., Grabherr, M., Ren, Q., Zdobnov, E.M., Lobo, N.F., Campbell, K.S., Brown, S.E., Bonaldo, M.F., Zhu, J., Sinkins, S.P., Hogenkamp, D.G., Amedeo, P., Arensburger, P., Atkinson, P.W., Bidwell, S., Biedler, J., Birney, E., Bruggner, R.V., Costas, J., Coy, M.R., Crabtree, J., Crawford, M., Debruyn, B., Decaprio, D., Eiglmeier, K., Eisenstadt, E., El-Dorry, H., Gelbart, W.M., Gomes, S.L., Hammond, M., Hannick, L.I., Hogan, J.R., Holmes, M.H., Jaffe, D., Johnston, J.S., Kennedy, R.C., Koo, H., Kravitz, S., Kriventseva, E.V., Kulp, D., Labutti, K., Lee, E., Li, S., Lovin, D.D., Mao, C., Mauceli, E., Menck, C.F., Miller, J.R., Montgomery, P., Mori, A., Nascimento, A.L., Naveira, H.F., Nusbaum, C., O'leary, S., Orvis, J., Pertea, M., Quesneville, H., Reidenbach, K.R., Rogers, Y.H., Roth, C.W., Schneider, J.R., Schatz, M., Shumway, M., Stanke, M., Stinson, E.O., Tubio, J.M., Vanzee, J.P., Verjovski-Almeida, S., Werner, D., White, O., Wyder, S., Zeng, Q., Zhao, Q., Zhao, Y., Hill, C.A., Raikhel, A.S., Soares, M.B., Knudson, D.L., Lee, N.H., Galagan, J., Salzberg, S.L., Paulsen, I.T., Dimopoulos, G., Collins, F.H., Birren, B., Fraser-Liggett, C.M., Severson, D.W., 2007. Genome sequence of *Aedes aegypti*, a major arbovirus vector. Science 316, 1718–1723.

Ong, C.T., Cheng, H.T., Chang, L.W., Ohtsuka, T., Kageyama, R., Stormo, G.D., Kopan, R., 2006. Target selectivity of vertebrate Notch proteins. Collaboration between discrete domains and CSL-binding site architecture determines activation probability. J. Biol. Chem. 281, 5106–5119.

Perriere, G., Gouy, M., 1996. WWW-query: an on-line retrieval system for biological sequence banks. Biochimie 78, 364–369.

Prud'homme, B., Gompel, N., Rokas, A., Kassner, V.A., Williams, T.M., Yeh, S.D., True, J.R., Carroll, S.B., 2006. Repeated morphological evolution through *cis*-regulatory changes in a pleiotropic gene. Nature 440, 1050–1053.

Putnam, N.H., Srivastava, M., Hellsten, U., Dirks, B., Chapman, J., Salamov, A., Terry, A., Shapiro, H., Lindquist, E., Kapitonov, V.V., Jurka, J., Genikhovich, G., Grigoriev, I.V., Lucas, S.M., Steele, R.E., Finnerty, J.R., Technau, U., Martindale, M.Q., Rokhsar, D.S., 2007. Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. Science 317, 86–94.

Putnam, N.H., Butts, T., Ferrier, D.E., Furlong, R.F., Hellsten, U., Kawashima, T., Robinson-Rechavi, M., Shoguchi, E., Terry, A., Yu, J.K., Benito-Gutierrez, E.L., Dubchak, I., Garcia-Fernandez, J., Gibson-Brown, J.J., Grigoriev, I.V., Horton, A.C., de Jong, P.J., Jurka, J., Kapitonov, V.V., Kohara, Y., Kuroki, Y., Lindquist, E., Lucas, S., Osoegawa, K., Pennacchio, L.A., Salamov, A.A., Satou, Y., Sauka-Spengler, T., Schmutz, J., Shin-I, T., Toyoda, A., Bronner-Fraser, M., Fujiyama, A., Holland, L.Z., Holland, P.W., Satoh, N., Rokhsar, D.S., 2008. The amphioxus genome and the evolution of the chordate karyotype. Nature 453, 1064–1071.

Pyrowolakis, G., Hartmann, B., Muller, B., Basler, K., Affolter, M., 2004. A simple molecular complex mediates widespread BMP-induced repression during *Drosophila* development. Dev. Cell 7, 229–240.

Rebeiz, M., Posakony, J.W., 2004. GenePalette: a universal software tool for genome sequence visualization and analysis. Dev. Biol. 271, 431–438.

Rebeiz, M., Stone, T., Posakony, J.W., 2005. An ancient transcriptional regulatory linkage. Dev. Biol. 281, 299–308.

Rebeiz, M., Jikomes, N., Kassner, V.A., Carroll, S.B., 2011. Evolutionary origin of a novel gene expression pattern through co-option of the latent activities of existing regulatory sequences. Proc. Natl. Acad. Sci. U. S. A. 108, 10036–10043.

Romano, L.A., Wray, G.A., 2003. Conservation of Endo16 expression in sea urchins despite evolutionary divergence in both cis and trans-acting components of transcriptional regulation. Development 130, 4187–4199.

Rubin, G.M., Spradling, A.C., 1982. Genetic transformation of *Drosophila* with transposable element vectors. Science 218, 348–353.

Schlatter, R., Maier, D., 2005. The *Enhancer of split* and *Achaete-Scute* complexes of Drosophilids derived from simple ur-complexes preserved in mosquito and honeybee. BMC Evol. Biol. 5, 67.

Sea Urchin Genome Sequencing Consortium, 2006. The genome of the sea urchin *Strongylocentrotus purpuratus*. Science 314, 941–952.

Shinzato, C., Shoguchi, E., Kawashima, T., Hamada, M., Hisata, K., Tanaka, M., Fujie, M., Fujiwara, M., Koyanagi, R., Ikuta, T., Fujiyama, A., Miller, D.J., Satoh, N., 2011. Using the *Acropora digitifera* genome to understand coral responses to environmental change. Nature 476, 320–323.

Simionato, E., Ledent, V., Richards, G., Thomas-Chollier, M., Kerner, P., Coornaert, D., Degnan, B.M., Vervoort, M., 2007. Origin and diversification of the basic helix-loop-helix gene family in metazoans: insights from comparative genomics. BMC Evol. Biol. 7, 33.

Singson, A., Leviten, M.W., Bang, A.G., Hua, X.H., Posakony, J.W., 1994. Direct downstream targets of proneural activators in the imaginal disc include genes involved in lateral inhibitory signaling. Genes Dev. 8, 2058–2071.

Srivastava, M., Begovic, E., Chapman, J., Putnam, N.H., Hellsten, U., Kawashima, T., Kuo, A., Mitros, T., Salamov, A., Carpenter, M.L., Signorovitch, A.Y., Moreno, M.A., Kamm, K., Grimwood, J., Schmutz, J., Shapiro, H., Grigoriev, I.V., Buss, L.W., Schierwater, B., Dellaporta, S.L., Rokhsar, D.S., 2008. The *Trichoplax* genome and the nature of placozoans. Nature 454, 955–960.

Srivastava, M., Simakov, O., Chapman, J., Fahey, B., Gauthier, M.E., Mitros, T., Richards, G.S., Conaco, C., Dacre, M., Hellsten, U., Larroux, C., Putnam, N.H., Stanke, M., Adamska, M., Darling, A., Degnan, S.M., Oakley, T.H., Plachetzki, D.C., Zhai, Y., Adamski, M., Calcino, A., Cummins, S.F., Goodstein, D.M., Harris, C., Jackson, D.J., Leys, S.P., Shu, S., Woodcroft, B.J., Vervoort, M., Kosik, K.S., Manning, G., Degnan, B.M., Rokhsar, D.S., 2010. The *Amphimedon queenslandica* genome and the evolution of animal complexity. Nature 466, 720–726.

Swanson, C.I., Schwimmer, D.B., Barolo, S., 2011. Rapid evolutionary rewiring of a structurally constrained eye enhancer. Curr. Biol. 21, 1186–1196.

Tribolium Genome Sequencing Consortium, 2008. The genome of the model beetle and pest *Tribolium castaneum*. Nature 452, 949–955.

Tun, T., Hamaguchi, Y., Matsunami, N., Furukawa, T., Honjo, T., Kawaichi, M., 1994. Recognition sequence of a highly conserved DNA binding protein RBP-Jk. Nucleic Acids Res. 22, 965–971.

Van Doren, M., Powell, P.A., Pasternak, D., Singson, A., Posakony, J.W., 1992. Spatial regulation of proneural gene activity: auto- and cross-activation of achaete is antagonized by extramacrochaetae. Genes Dev. 6, 2592–2605.

Van Doren, M., Bailey, A.M., Esnayra, J., Ede, K., Posakony, J.W., 1994. Negative regulation of proneural gene activity: hairy is a direct transcriptional repressor of *achaete*. Genes Dev. 8, 2729–2742.

Venken, K.J., He, Y., Hoskins, R.A., Bellen, H.J., 2006. P[acman]: a BAC transgenic platform for targeted insertion of large DNA fragments in *D. melanogaster*. Science 314, 1747–1751.

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., Gocayne, J.D., Amanatides, P., Ballew, R.M., Huson, D.H., Wortman, J.R., Zhang, Q., Kodira, C.D., Zheng, X.H., Chen, L., Skupski, M., Subramanian, G., Thomas, P.D., Zhang, J., Gabor Miklos, G.L., Nelson, C., Broder, S., Clark, A.G., Nadeau, J., McKusick, V.A., Zinder, N., Levine, A.J., Roberts, R.J., Simon, M., Slayman, C., Hunkapiller, M., Bolanos, R., Delcher, A., Dew, I., Fasulo, D., Flanigan, M., Florea, L., Halpern, A., Hannenhalli, S., Kravitz, S., Levy, S., Mobarry, C., Reinert, K., Remington, K., Abu-Threideh, J., Beasley, E., Biddick, K., Bonazzi, V., Brandon, R., Cargill, M., Chandramouliswaran, I., Charlab, R., Chaturvedi, K., Deng, Z., Di Francesco, V., Dunn, P., Eilbeck, K., Evangelista, C., Gabrielian, A.E., Gan, W., Ge, W., Gong, F., Gu, Z., Guan, P., Heiman, T.J., Higgins, M.E., Ji, R.R., Ke, Z., Ketchum, K.A., Lai, Z., Lei, Y., Li, Z., Li, J., Liang, Y., Lin, X., Lu, F., Merkulov, G.V., Milshina, N., Moore, H.M., Naik, A.K., Narayan, V.A., Neelam, B., Nusskern, D., Rusch, D.B., Salzberg, S., Shao, W., Shue, B., Sun, J., Wang, Z., Wang, A., Wang, X., Wang, J., Wei, M., Wides, R., Xiao, C., Yan, C., Yao, A., Ye, J., Zhan, M., Zhang, W., Zhang, H., Zhao, Q., Zheng, L., Zhong, F., Zhong, W., Zhu, S., Zhao, S., Gilbert, D., Baumhueter, S., Spier, G., Carter, C., Cravchik, A., Woodage, T., Ali, F., An, H., Awe, A., Baldwin, D., Baden, H., Barnstead, M., Barrow, I., Beeson, K., Busam, D., Carver, A., Center, A., Cheng, M.L., Curry, L., Danaher, S., Davenport, L., Desilets, R., Dietz, S., Dodson, K., Doup, L., Ferriera, S., Garg, N., Gluecksmann, A., Hart, B., Haynes, J., Haynes, C., Heiner, C., Hladun, S., Hostin, D., Houck, J., Howland, T., Ibegwam, C., Johnson, J., Kalush, F., Kline, L., Koduru, S., Love, A., Mann, F., May, D., McCawley, S., McIntosh, T., McMullen, I., Moy, M., Moy, L., Murphy, B., Nelson, K., Pfannkoch, C., Pratts, E., Puri, V., Qureshi, H., Reardon, M., Rodriguez, R., Rogers, Y.H., Romblad, D., Ruhfel, B., Scott, R., Sitter, C., Smallwood, M., Stewart, E., Strong, R., Suh, E., Thomas, R., Tint, N.N., Tse, S., Vech, C., Wang, G., Wetter, J., Williams, S., Williams, M., Windsor, S., Winn-Deen, E., Wolfe, K., Zaveri, J., Zaveri, K., Abril, J.F., Guigo, R., Campbell, M.J., Sjolander, K.V., Karlak, B., Kejariwal, A., Mi, H., Lazareva, B., Hatton, T., Narechania, A., Diemer, K., Muruganujan, A., Guo, N., Sato, S., Bafna, V., Istrail, S., Lippert, R., Schwartz, R., Walenz, B., Yooseph, S., Allen, D., Basu, A., Baxendale, J., Blick, L., Caminha, M., Carnes-Stine, J., Caulk, P., Chiang, Y.H., Coyne, M., Dahlke, C., Mays, A., Dombroski, M., Donnelly, M., Ely, D., Esparham, S., Fosler, C., Gire, H., Glanowski, S., Glasser, K., Glodek, A., Gorokhov, M., Graham, K., Gropman, B., Harris, M., Heil, J., Henderson, S., Hoover, J., Jennings, D., Jordan, C., Jordan, J., Kasha, J., Kagan, L., Kraft, C., Levitsky, A., Lewis, M., Liu, X., Lopez, J., Ma, D., Majoros, W., McDaniel, J., Murphy, S., Newman, M., Nguyen, T., Nguyen, N., Nodell, M., Pan, S., Peck, J., Peterson, M., Rowe, W., Sanders, R., Scott, J., Simpson, M., Smith, T., Sprague, A., Stockwell, T., Turner, R., Venter, E., Wang, M., Wen, M., Wu, D., Wu, M., Xia, A., Zandieh, A., Zhu, X., 2001. The sequence of the human genome. Science 291, 1304–1351.

Wray, G.A., 2007. The evolutionary significance of *cis*-regulatory mutations. Nat. Rev. Genet. 8, 206–216.

Yao, L.C., Phin, S., Cho, J., Rushlow, C., Arora, K., Warrior, R., 2008. Multiple modular promoter elements drive graded *brinker* expression in response to the Dpp morphogen gradient. Development 135, 2183–2192.