Evolution of the Bicoid homeodomain required amino acid substitutions in three subdomains and epistatic interactions between them

Pinar Onal[1]*, Himari Imaya Gunasinghe[1], Kristaley Yui Umezawa[1], Michael Zheng[1], Jia Ling[1], Stephen Small[1]

[1]New York University, Department of Biology

*Corresponding Author

pinar.onal@gmail.com

Summary

Gene duplications contribute to the evolution of biological diversity. The ancestor of the *Drosophila* Homeodomain (HD) protein Bicoid (AncBcd) emerged after a duplication of a Zerknullt (Zen)-like ancestral protein (AncZB) in a suborder of flies. AncBcd evolved and diverged from AncZB, gaining novel transcriptional and translational activities. Here, we focus on the evolution of the AncBcd HD, which binds DNA and RNA, and is comprised of four subdomains, an N-terminal arm (NT) and three helices: H1, H2, and Recognition Helix (RH). Using gene rescue assays in *Drosophila*, we show that robust patterning activity of the HD (high frequency rescue to adulthood) is achieved only when amino acid substitutions in three separate subdomains (NT, H1, and RH) are combined. Other combinations of subdomains also yield full rescue, but with lower and different penetrance. Our results suggest that there are multiple intermediate solutions for the evolution of the Bcd HD that constrain and enable further evolutionary changes. They also demonstrate how intramolecular epistasis and complementary but not completely redundant functions of subdomains contribute to the robust function of an evolving transcription factor.

Keywords: gene duplication, homeodomain, evolution, development, Bicoid, embryo, segmentation, anterior patterning

Introduction

How genetic changes lead to evolutionary divergence is one of the central questions of biology. Among many types of molecular events that alter genomes, gene duplication events are instrumental in creating genetic diversity by providing extra genetic material for the evolution of novel or modified functions. New functions can be gained (neofunctionalization) in either product of the duplication, through cis-regulatory changes that alter gene expression patterns, or coding sequence changes that alter protein activity [1], [2]. For example, multiple duplications in the Hox locus, followed by diversification of individual genes, were critical for establishing divergent body plans throughout the metazoa [3]–[7].

Here, we study the evolution of Bicoid (Bcd), a homeodomain (HD)-containing transcription factor that is critical for patterning anterior regions of the *Drosophila* embryo. The ancestor of *Drosophila* ***b****cd* gene (*ancbcd*) emerged 150 MYA in Cyclorrhaphan flies (a suborder in flies) after a duplication of an ancestral gene that it shares with *zerknullt* (*anc****zb***).[8]–[10]. In *Drosophila* and most other Cyclorrhaphan flies, *zen* maintained *anczb*'s role in extraembryonic patterning, while *ancbcd* evolved rapidly. In addition to evolution in regulatory sequences that led to maternal expression and anterior localization of *bcd* mRNA, coding sequence changes completely altered the DNA-binding activities of AncBcd [11] and allowed it to bind to RNA [12], [13]. In the early embryo, Bcd protein is distributed in an anterior to posterior (AP) gradient [14] and is essential for transcriptionally activating more than 50 genes in unique temporal and spatial patterns along the AP axis [15]–[17]. Most Bcd target genes encode transcription factors, which regulate each other in an integrated network that forms and positions seven head segments and three thoracic segments in the anterior half of the developing embryo [18]. Bcd also binds directly to the mRNA of the posterior determinant *caudal* (*cad*), and prevents its translation in anterior embryonic regions [19], [20]. Embryos lacking Bcd form no head or thoracic segments, and show variable defects in abdominal segmentation [21].

The Zen and Bcd proteins in *Drosophila* have completely different functions *in vivo*. Specifically, when expressed in a Bcd-like gradient in embryos lacking Bcd, Zen has no Bcd-like activity [22]. However, when the Bcd HD is swapped into the Zen protein, the chimeric ZenBcdHD partially rescues the morphological defects in Bcd-depleted embryos, and activates a

subset of Bcd target genes [22]. These results indicate that the unique patterning activities of Bcd are determined in large part by its DNA- and RNA-binding preferences. They suggest further that amino acid substitutions in the AncZB HD were critical for the evolution of Bcd's functions in anterior embryo patterning.

In a previous study, ancestral protein reconstruction (APR; [23]) was used to infer the amino acid sequences of the HDs that were present in AncZB and AncBcd. There were 31 high confidence differences between the two HDs, which were distributed among four HD subdomains: the N-terminal arm (NT) and three alpha helices [H1, H2, and the DNA recognition helix (RH)] (Figure 1A and B). When tested *in vivo*, a Bcd protein containing the AncZB HD failed to provide any Bcd-like activity [Figure 1D-G; [22]], while an identical construct carrying the AncBcd HD completely rescued embryos lacking endogenous Bcd to adulthood (Figure 1H-K). This study also tested the roles of two substitutions in the RH (`q50>K` and `m54>R`) because these had previously been shown to be critical for Bcd's DNA-binding specificity [24], [25] and RNA-binding activities [19]. Introducing the `q50>K` substitution alone into the AncZB HD (AncZB_**`K50`**) changed its *in vitro* DNA-binding specificity to one indistinguishable from that of AncBcd. However, when inserted into the Drosophila Bcd protein and tested *in vivo*, the AncZB_**`K50`** HD activated only a subset of Bcd target gene resposes, including *hb* and *eve stripe 2*, but failed to activate *gt, otd*, *btd*, and *eve* stripe 1. It also failed to repress Cad translation [22]. Adding the `m54>R` substitution (AncZB_**`K50R54`**) increased the number of target gene responses by additionally activating *gt* and *otd*, but still failed to activate *btd* and *eve* stripe 1, or repress Cad translation. Importantly, the double substitution failed to fully rescue the defects observed in embryos lacking Bcd [Figure 2A-D;[22]]. These results suggest that other substitutions in the RH or in other subdomains of the HD were required for the evolution of AncBcd HD's full transcriptional and post-transcriptional activities.

In this paper, we present experiments designed to identify these other substitutions (Table 1). We show that other forward substitutions in the RH collectively and synergistically contribute to HD function by increasing the number of target genes regulated by the AncZB HD. However, RH substitutions alone cannot fully rescue embryos lacking Bcd to adulthood. High frequency survival to adulthood is observed only if forward substitutions in the RH are combined with substitutions in two other subdomains (NT and H1). In contrast, combining substitutions in the

RH with those in helix H1 or NT alone generates suboptimal HDs that achieve full rescue to adulthood, but with much lower frequencies. Taken together, these results suggest a multi-step pathway to explain the evolutionary transition from a non-functional AncZB HD to an AncBcd HD with robust *in vivo* function.

Results:

Epistatic interactions within the RH and between subdomains of the AncBcd HD determine the functional outcome of evolutionary changes

Among the 31 amino acid differences between the AncZB and AncBcd HDs, 11 are highly conserved in all available Bcd HD sequences, and are not found in any available Zen HD sequences [[22]; Figure 1A]. Six of these "diagnostic" substitutions, including `q50>K` and `m54>R`, are present in the RH subdomain, which directly contacts DNA [26]. We hypothesized that one or more diagnostic RH substitutions besides `q50>K` and `m54>R` might augment the degree of rescue mediated by the AncZB_**K50R54** HD. To start, we added all four (residues 42, 43, 55, and 58) to the AncZB_**K50R54** HD to generate the AncZB_**RHdiag** HD (Table 1, Figure 2E-H). Surprisingly, when inserted into a *bcd* rescue transgene, the AncHox3_**RHdiag** HD showed a lower level of rescue than the AncZB_**K50R54** HD. For example, no head structures were observed in larvae carrying the AncZB_**RHdiag** HD, and only 40% formed two thoracic segments [compared to 80% for the AncZB_**K50R54** (Figure 2E, F, compared to Figure 2A, B)]. This lower level of rescue activity was also observed at the transcriptional level. Consistent with the missing head structures, expression of the head gap gene *otd* [easily detectable in embryos rescued by the AncZB_**K50R54** construct (Figure 2D)] was not detected in embryos rescued with the AncZB_**RHdiag** construct (Figure 2H). We also observed reductions in the expression patterns of *hb* and *gt*, and anterior shifts of these patterns compared to those activated by the AncZB_**K50R54** HD (Figure 2H, compared to 2D). For *hb*, we quantified this shift by measuring the posterior boundary position (pbp) as a percentage of embryo length (% EL, where 100% = the anterior tip) (see Methods). In embryos carrying the AncZB_**K50R54** construct, the average PBP was at 77% EL, while in AncZB_**RHdiag** embryos, the average position was at 82% EL

(Figure S1B compared to S1A). Finally, the AncZB_**RHdiag** construct did not detectably repress Cad translation (Figure 2G).

These experiments suggest that negative epistatic interactions exist among diagnostic residues in the AncZB_**RHdiag** HD, which reduce biological activity compared to the AncZB_**K50R54** double substitution. One possibility is that these negative interactions are mitigated by the other three non-diagnostic substitutions in the AncBcd RH (Figure 1C). To test this, we replaced the whole RH from AncZB HD with that of AncBcd (AncZB_**RH**; Table 1). The addition of three more substitutions in the RH substantially improved the *in vivo* activity compared to both the AncZB_**K50R54** and the AncZB_**RHdiag** constructs (Figure 2I-L). Around 95% of larvae containing AncZB_**RH** formed all three thoracic segments, and more than 80% formed cephalic structures [mouthhooks (MH) and lateralgraete (LG) only; Figure 2I, J]. However, no larvae carrying the AncZB_**RH** construct survived to adulthood. At the molecular level, early embryos activated transcription of the target genes *otd* and *btd* (Figure 2L), which were not activated by the AncZB_**RHdiag** (Figure 2H), but failed to activate *eve* stripe 1. Also, the expression patterns of *hb* and *gt* were more strongly activated in AncZB_**RH** embryos compared to AncZB_**RHdiag**. In particular, the average *hb* pbp in AncZB_**RH** embryos was at 73% EL (Figure S1D), which is more posteriorly localized compared to 82% EL in AncZB_**RHdiag** embryos (Figure S1C). We also could not detect any significant repression of Cad translation in AncZB_**RH** embryos (Figure 2K).

A second possibility is that the negative epistatic interactions observed in the AncZB_**RHdiag** experiment (Figure 2E-H) are mitigated by diagnostic substitutions in other subdomains. To test this, we replaced all 11 diagnostic residues in the nonfunctional AncZB HD with the corresponding residues in AncBcd HD (AncZB_**Alldiag**) (Table 1, Figure 2M-P). These substitutions in multiple subdomains caused an even stronger increase in patterning activity compared to all constructs containing substitutions in the RH alone (Figure 2). More than 90% of larvae containing the AncZB_**Alldiag** HD formed all three thoracic segments, and 80% formed detectable head structures such as mouthhooks (MH). 40% formed lateralgraete (LG) and around 10% of the larvae carrying AncZB_**Alldiag** formed dorsal and ventral arms (DA and VA; Figure 2M, N), which were never observed in embryos carrying AncZB_**RH**

(Figure 2I, J). Despite these improvements, none of the larvae appeared wild-type like (WTL), and no larvae containing the AncZB_**Alldiag** HD construct survived to adulthood. At the molecular level, the AncZB_**Alldiag** activated *otd* and *btd*, and weakly activated *eve* stripe 1 (Figure 2P), which was not detected in AncZB_**RH** embryos (Figure 2H). In addition, the *hb* and *gt* expression pattens were intensified (Figure 2P, compare with 2H), and the *hb* pbp was located at 70% EL (Figure S1D), slightly posterior compared to that of AncZB_**RH** (Figure S1C). Importantly, embryos carrying the AncZB_**Alldiag** construct suppressed translation of Cad in anterior regions (Figure 2O).

In summary, the preceding experiments indicate that epistatic interactions between amino acids in the RH can substantially modify the patterning activities of the AncBcd HD. They also show that an HD containing diagnostic substitutions in all four subdomains (AncZB_**Alldiag**) mediates increased patterning activity compared to an HD containing all nine forward substitutions (diagnostic and non-diagnostic) in the RH alone. From an evolutionary perspective, these experiments suggest that forward substitutions in multiple subdomains were critical for the acquisition of Bcd's novel patterning activities. However, none of these constructs could mediate full rescue of embryos lacking Bcd to adulthood.

Substitutions in three subdomains are required for the robust function of the AncBcd HD

To identify which subdomain substitutions are required and sufficient for the full function of the AncBcd HD, we generated a series of chimeric ancestral HDs containing all forward or reverse substitutions (diagnostic and non-diagnostic) in specific subdomains, and tested them using the *bcd* rescue strategy described above. For each construct, we first assayed cuticle patterns of first instar larvae. Because some experiments yielded highly variable phenotypes, we quantified the percentage of larvae in each of the following categories (see STAR Methods): 1. Larvae with easily detectable head structures (MH, LG, VA, DA, and DBr), three thoracic segments, and eight abdominal segments were classified as wild-type like (WTL). 2. Larvae missing any of those five head structures were classified as ΔHead. 3. Larvae with defects in abdominal segments were classified as ΔAbdomen. For all experiments that yielded WTL larvae, hatching tests were set up to monitor the frequency of survival past larval stages (See STAR

Methods). As positive controls, we assayed the rescue activities of constructs containing the AncBcd HD and the endogenous *Drosophila* Bcd HD, which showed 87% and 90% WTL larvae respectively (Figure 3A, B; Figure S2A and B). In both control experiments, the frequency of survival to adulthood was lower than the number of WTL larvae (~37% and 50%, respectively), and in general lower than the survival rate observed for wildtype embryos (>80%), perhaps due to the fact that these transgenes are inserted into an ectopic genomic position.

We first wanted to define subdomains required for full patterning function of the AncBcd HD. Because changing even a single amino acid (K50) completely abolishes AncBcd HD activity [22], we tested the other three subdomains individually by replacing the NT, H1, and H2 sequences of the fully functional AncBcd HD with those of the nonfunctional AncZB HD (Table 1). The co-crystal structure of Bcd bound to DNA shows that the NT arm contacts the minor groove [26], and this interaction, together with interactions between the RH and the major groove, contributes to the specificity and affinity of the HD for target DNA motifs. Consistent with this, reverse substitutions that change the AncBcd NT sequence to that seen in the AncZB HD [**CH(***nt***-H1-H2-RH**)] significantly reduced rescue activity compared to the AncBcd HD (Figure 3C; Figure S2C). Around 40% of larva containing this chimeric construct were classified as WTL, and 12% survived to adulthood (Figure 3C, H and Figure S2C). Reverse substitutions in the H1 helix [**CH(NT-***h1***-H2-RH**)] caused even stronger negative effects on the function of the AncBcd HD. While this construct partially rescued the morphological defects, no larvae were classified as WTL, and none survived to adulthood (Figure 3D, H; Figure S2D). 95% of the first instar larvae carrying **CH(NT-***h1***-H2-RH**) formed three thoracic and eight abdominal segments, but they consistently showed head defects (ΔHead, Figure 3D, Figure S2D). In contrast, reverse substitutions in H2, [**CH(NT-H1-***h2***-RH**)] had milder effects on the rescue activity of the AncBcd HD. More than 60% of the larvae carrying this construct were classified as WTL (Figure 3E; Figure S2E), and 43% survived to adulthood, Figure 3H). In summary, reverse substitutions in two different subdomains (NT and H1) strongly reduced the *in vivo* activity of the AncBcd HD, while reverse substitutions in H2 caused less severe effects.

Viewed another way, the strong patterning activity of the [**CH(NT-H1-***h2***-RH**)] construct shows that combining forward substitutions in NT, H1, and RH is sufficient to convert the inactive AncZB HD into a fully functional Bcd-like HD with control level survival rates. To

further investigate interactions between subdomains, we tested AncZB HDs containing two simpler combinations of forward substitutions: [**CH(NT-*h1*-*h2*-RH)**] and [CH(*nt*-**H1**-*h2*-**RH**)] (Table 1). When the [**CH(NT-*h1*-*h2*-RH)**] construct was introduced into embryos lacking Bcd, ~75% of first instar larvae were classified as WTL (Figure 3F; Figure S2F). In most of these larvae, however, specific head structures appeared shorter than normal, including the dorsal and ventral arms and the lateralgraete, possibly affecting the frequency of survival to adulthood, which was only ~3% (Figure 3H). When the [**CH(*nt*-H1-*h2*-RH)**] construct was tested in the *bcd* rescue assay, ~40% of the first instar larvae were classified as WTL (Figure 3G; Figure S2G). Curiously, another 40% formed all head segments, but showed defects in abdominal segments (ΔAbdomen), a phenotype seen only rarely with the other rescue constructs tested here (Figure 3; Figure S2). Abdominal defects ranged from a mild phenotype showing only four segments to a strong phenotype lacking all abdominal segments and poorly formed filzkörper, which are normally formed at the posterior end of the larva. Remarkably, despite these larval defects, ~7% of embryos rescued by the [CH(*nt*-**H1**-*h2*-**RH**)]construct survived to adulthood (Figure 3H).

Taken together, these experiments show that three different combinations of subdomains ([**CH(NT-*h1*-*h2*-RH)**], [**CH(*nt*-H1-*h2*-RH)**], and [**CH(*nt*-H1-H2-RH)**]) fully rescued embryos lacking Bcd with low penetrance (3-12%), while another combination of changes including three subdomains ([**CH(NT-H1-*h2*-RH)**]) fully rescued with higher penetrance (~43%).

The frequency of rescue to adulthood correlates with target gene position

We considered two hypotheses to explain the different levels of survival in embryos containing chimeric HDs with different combinations of AncZB and AncBcd subdomains. One hypothesis is that the frequency of full rescue is correlated with, and possibly determined by the percentage of embryos that activate all Bcd target genes or show suppression of Cad. In this scenario, lines that show low penetrance rescue would have fewer embryos that activate all Bcd target genes, and/or fewer embryos that show Cad suppression. To test this, we stained embryos containing each fully rescuing chimeric HD by *in situ* hybridization to detect expressions of *hb*, *otd*, *btd*, *gt* and *eve* (Figure S3) and suppression of the RNA target *cad.* All embryos in each low and high penetrance rescuing lines as well as the non-rescuing line [**CH(NT-*h1*-H2-RH)**]

expressed all Bcd target genes, and suppressed Cad translation (Figures S3 and S4), arguing strongly against the hypothesis that embryos surviving to adulthood are simply a small subset that have gained these molecular activities.

An alternative hypothesis is that the frequency of rescue correlates with the level and patterning of target genes along the AP axis. To test this, we measured the *hb* pbp and the position of *eve* stripe 1. In embryos containing the AncBcd rescue construct, which survive to adulthood at a frequency of 50%, the *hb* pbp was positioned at ~54% EL (Figure 4K). This boundary was positioned more anteriorly in all lines carrying chimeric HDs, but there were significant differences between individual constructs. At the extremes for this group of constructs, embryos carrying ([**CH(NT-*h1*-H2-RH**)], 0% rescue to adulthood) formed the *hb* pbp at 72% (Figure 4A); and embryos carrying the [**CH(NT-H1-*h2*-RH)**, 43% rescue to adulthood} formed this boundary at 63% EL (Figure. 4I). As expected for the boundary positioning hypothesis, embryos containing constructs that directed low frequency rescue to adulthood (3-12%) formed their *hb* pbps at positions between 72% and 63% EL. Similar results were observed for *eve* stripe 1, which is expressed in cells destined to become the cephalic furrow that separates the head from the thorax later in embryogenesis. Embryos carrying [**CH(NT-*h1*-H2-RH**)], consistently expressed *eve 1*, but expression was extremely weak and more anteriorly located (Figure 4A) compared to all constructs that directed rescue to adulthood (Figure 4D, F, H, J).

In summary, HD chimeras with more robust rescue activities extend the *hb* expression domain into more posterior regions of the embryo. Hb strongly represses posterior development; perhaps it must occupy a minimum amount of space in the anterior region, which is required to activate *eve* stripe 1 expression, and correctly position all the Bcd target genes involved in anterior development. Consistent with this, the expression patterns of *gt*, *otd*, and *btd* are more strongly expressed and positioned more posteriorly as the *hb* and *eve 1* patterns move posteriorly and in lines showing full rescue (Figure S3).

Discussion

Molecular requirements for the patterning activity of the AncBcd HD in Drosophila

In this paper, we used an *in vivo* rescue assay in *Drosophila* as a proxy to study the molecular evolution of the AncBcd HD. By making chimeric HDs between the AncZB (no function) and the AncBcd (full function) HDs (Table 1), we showed that the substitutions in at least three separate subdomains (NT, H1, and RH) must be combined for full patterning activity.

AncBcd evolved to suppress translation of Cad and activate transcription of a large number of target genes at different positions along the AP axis of the embryo. Our results shed light on the molecular requirements for both of these activities. The R54 residue in Bcd was previously shown to be required for Cad suppression [19], but our data suggest that it is not sufficient, even in combination with the other eight forward substitutions in the RH of AncBcd (AncZB_**RH**). However, by combining the RH substitutions with those in other subdomains, suppression was achieved. Several different combinations directed varying levels of suppression, which were in general lower than the control level suppression (Figure 2O, Figure S4). The impact of the level of suppression on the rescue potential and patterning is not known, and will be addressed by future experiments.

At the transcriptional level, it is possible to more definitively associate specific aspects of Bcd's gene activation program with the presence or absence of substitutions in different subdomains. It was previously shown that inserting K50 alone into the AncZB HD caused the activation of only three of eight tested target gene responses, while the double substitution (K50R54) increased that number to five [22]. In this paper, we show that substituting all nine amino acids from the AncBcd HD into AncZB resulted in the activation of all tested target gene responses except one (*eve stripe* 1). However, the expression patterns of the activated target genes in these experiments were confined to the region very near the anterior tip (Figure 2L). Combining substitutions in the RH with those in other subunits had two major effects, the activation of *eve* 1, and the extensions of target gene expression patterns into more posterior regions of the embryo, which correlates with the anterior development and rescue mediated by each combination.

This correlation is most easily observed for the target gene *hb*, which encodes a critical cofactor for activation of all Bcd-dependent target genes [27]–[30], and functions as an important repressor to prevent posterior gap gene expression in anterior regions of the embryo [31]–[34]. In embryos carrying constructs that fail to fully rescue to adulthood, *hb* pbps are located between 82 to 70 % EL (Figure S1A-D; Figure 4A). In contrast, embryos carrying constructs that fully rescue form *hb* pbps at more posterior positions (between 68 and 63% EL; Figure 4C, E, G, I). Interestingly, the **[CH(NT-H1-*h2*-RH)]** construct, which rescues to adulthood with a frequency similar to or greater than that observed for the AncBcd HD control, forms a *hb* pbp at 63%. While this position lies posterior to that observed in embryos that fail to survive or survive with low penetrance, it is significantly more anterior than the pbp seen in embryos carrying the control AncBcd construct (54% EL, Figure 4K) or wild type embryos (54%; [17]). However, the 63% EL position is very close to the *hb* pbp (~65% EL) in embryos laid by heterozygous *bcd* females, which survive with high penetrance [35]. We propose that ~63% EL represents a boundary position that establishes the minimum amount of embryonic space required for robust head formation and survival. Embryos with *hb* pbps posterior to 63% EL survive to adulthood; those with *hb* pbps anterior to 63% show head defects and die before the end of embryogenesis. How interactions between the RH and other HD subdomains cause posterior extensions of the zygotic *hb* domain is not clear; they could indirectly modify the DNA binding preferences of the HD or mediate interactions with maternal cofactors such as Hb or Zelda, both of which are critical for Bcd's *in vivo* functions in *Drosophila* [27], [28], [36]–[39].

Evolution of the AncBcd HD

An ancient duplication of AncZB led to the evolution of the K50 HD protein Bcd as a key regulator of anterior development in the Cyclorrhaphan suborder of the Diptera (two-winged insects) [10]. Other suborders of the Diptera and all other insects do not contain Bcd; In these insects, Bcd's roles in anterior patterning must be fulfilled by other gene(s). In the flour beetle *Tribolium* and the jewel wasp *Nasonia*, Bcd-like activity is partially provided by Orthodenticle (Otd), another K50 HD transcription factor that binds to DNA sequences similar to those bound by Bcd [40], [41]. Because many Bcd target genes are well conserved across all insects that have been studied so far, it has been proposed that Bcd evolved to take over regulation of an ancestral network of genes regulated by Otd [40]. In *Drosophila*, *otd* has evolved to become a Bcd target

gene [42], and its ancestral role in anterior patterning may have been diminished [43]. In addition, studies of other insects have identified unrelated proteins [a homolog of Odd-paired in the drain fly *Clogmia*, and a cysteine clamp protein in the midge *Chironomus*] that are required for patterning [44], [45]. These findings suggest that dramatic changes involving different molecular pathways have occurred in the evolution of the AP patterning network in different insects. The duplication of AncZB and neofunctionalization of AncBcd is one such change in the lineage that gave rise to the Cyclorrhaphous Diptera.

Our data show that robust patterning function of the AncBcd from AncZB is achieved by combining forward substitutions in three subdomains (RH, NT, and H1). It seems impossible critical amino acid substitutions in all three subdomains occurred simultaneously at some point in the evolution of the AncBcd HD. However, our data suggest that critical substitutions in each of the three might have occurred in a specific temporal order, and there were a series of intermediate HD configurations, each of which endowed the protein with a novel property that could be positively selected for in a subpopulation of evolving flies (Figure 5).

Assuming that the ancestral network was controlled by a K50 HD protein such as Otd [43], we propose that the first step involved multiple substitutions in the RH, including q50>K and m54>R. The codons for Q (Gln: CAA and CAG) and K (Lys: AAA and AAG) differ by only one base, so the q50>K transition involved only a single base pair substitution that would have dramatically changed the evolving protein's DNA-binding preference. Furthermore, reverse substituting or mutating K50 completely abolishes AncBcd HD function[22], which means that the effects of all other substitutions in the evolving HD were dependent on keeping the K50 residue intact. If this substitution occurred in an ancestral fly with an Otd-dependent anterior patterning network, the evolving protein would be immediately available to bind to many Otd-dependent target genes, which might have provided a selective advantage. The m54>R substitution, which also involves a single base change (AUG to AGG), might have refined DNA-binding specificity to increase the number of activated target genes, and set the stage for other substitutions that allowed the AncBcd HD to bind to RNA. K50 and R54 are present together only in Bcd HDs [25], consistent with the possibility that this combination might have been under positive selection. In addition to the q50>K and m54>R substitutions, there are seven other amino acid differences between the RH subdomains of AncZB and AncBcd. It is not clear which

of these are required for AncBcd HD function, or when they appeared historically. However, one combination of six substitutions tested here (AncZB_**RHdiag**) reduced HD activity compared to the K50R54 double substitution. This result suggests that interactions between amino acids constrained the historical order of substitutions in the RH subdomain.

While robust HD activity requires substitutions in three subdomains, forward substitutions in either NT or H1 substantially augment the rescue activity generated by changes in the RH alone. Specifically, AncZB HDs containing either combination (RH+NT or RH+H1) rescue a small percentage of embryos that survive to adulthood (Figure 3H). We propose that the addition of substitutions in either NT or H1 represent alternative second steps in the historical evolution of AncZB HD (Figure 5). Either combination (RH +NT or RH+H1) would have generated an intermediate HD configuration that could have been positively selected for and stabilized, perhaps by increasing fitness of a sub-population in specific physical/environmental conditions. Once stabilized, in a third step, substitutions in the other critical subdomain (H1 for the NT+RH intermediate, for example) would further increase HD activity and robustness of the evolving HD.

Limitations and challenges for the future.

Our results shed light on the mechanisms involved in the evolution of the AncBcd HD, but are limited by the fact that all chimeric HDs were inserted into the modern-day *Drosophila* Bcd protein. As such, these experiments do not take into account the evolution of other parts of the protein, which show even greater levels of amino acid sequence divergence. Further, all our experiments were performed in modern-day *Drosophila* embryos, and do not take into account changes in the cis-regulatory elements of target genes that co-evolved with the AncBcd protein. However, as the genome sequences of more insects become available, it should be possible to use reconstruction strategies to define the ancestral sequences of the complete AncBcd protein and the regulatory regions it interacts with. Furthermore, the ever-increasing use of CRISPR/Cas9 techniques for gene editing in non-model organisms should allow for testing ancestral protein and regulatory sequences in multiple insect species. While these methods cannot create the ancestral systems themselves, they should make it possible to discover general features that permit a transcription factor and its target regulatory sequences to co-evolve.

Acknowledgements

We thank Danyang Yu and Anecine Dalmeus for technical assistance, Rhea Datta and Patrick Lemaire for stimulating discussions throughout the course of this project, and Claude Desplan, Esteban Mazzoni, Erik Clark, and Matt Rockman for suggestions that greatly improved the manuscript. This project was supported by NIH Grant RO1 GM 51946 to SS.

Author contributions

Conceptualization, P.O. and S.S.; Methodology, P.O., M.Z., J.L., and S.S.; Investigation, P.O., H.I.G., K.Y.U., and M.Z.; Writing – Original Draft, P.O. and S.S.; Writing – Review & Editing, P.O. and S.S..; Funding Acquisition, S.S.; Resources, M.Z. and J.L..; Supervision, P.O. and S.S.

Figure Legends

Figure 1: 31 amino acids substitutions along the Homedomain (HD) led the divergence of the AncBcd HD from its ancestor AncZB HD

A. Amino acid sequences of the HDs of AncZB (blue) and AncBcd (orange), from Liu et al, 2018. The four subdomains of the HD are labeled above the corresponding residues. **B.** NMR structure of the *Drosophila* Bcd HD bound to DNA (adapted from [26]). **C.** Nomenclature used in this paper to denote the subdomain structures of the AncZB (top) and AncBcd HDs. AncZB sequences are shown as lower case and italicized, while AncBcd sequences are shown as regular and bold. Replacing AncZB amino acids with those in the AncBcd HD are referred to as forward substitutions; reciprocal replacements are referred to as reverse substitutions. **D-G.** Morphological and molecular activities provided by AncZB HD transgene. **D.** First instar larvae produced by *bcd* mutant females carrying two copies of the AncZB HD rescue transgene. For each construct, a DIC image of the cephalic region and a dark-field view of a whole larvae is shown. **E.** Percentages of first instar larvae that formed specific morphological structures in AncZB HD transgenic larvae. **F.** Cad immunostaining of the transgenic embryos at stage 5. **G.** RNA level detection of Bcd target genes in transgenic embryos with AncZB HD. **H-K.** Morphological and molecular activities provided by AncBcd HD transgene as described in D-G. Legend for structures in D-I and in plots I-M: Filzkoerper (Fk), dorsal arm (DA), ventral arm (VA), lateralgraete (LG), mouthhook (MH), thoracic segments (T1-T3), and abdominal segments (A1-A3), Dorsal Bridge (DBr). All larvae and embryos shown here and in all figures are oriented with anterior to the left and dorsal up. The genes tested were *hunchback* (*hb*), *giant* (*gt*), *orthodenticle* (*otd*), *buttonhead* (*btd*) and *even-skipped* (*eve*). The first three *eve* stripes are shown and Bcd-dependent stripes (#1 and #2) are labeled in red.

Figure 2: Morphological and molecular activities provided by diagnostic and nondiagnostic changes reveal synergistic and combinatorial interactions between them.

Morphological and molecular activities provided by AncZB_**K50R54** (A-D), AncZB_**RHdiag** (E-H), AncZB_**RH** (I-L) and AncZB_**Alldiag** (M-P). **A, E, I, M.** First instar larvae image produced by *bcd* mutant females carrying two copies of the rescue transgene. For each construct, a DIC image of the cephalic region and a dark-field view of a whole larvae is shown. **B, F, J, N.** Percentages of first instar larvae that formed specific morphological structures in the transgenic larvae. **C, G, K, O.** Cad immunostaining of the transgenic embryos at stage 5. **D, H, L, P.** RNA level detection of Bcd target genes in transgenic embryos with corresponding rescue construct. Each panel shows a representative stage 5 embryo stained by in situ hybridization to detect the expression pattern of a specific Bcd target gene. See also Figure S1.

Figure 3: Phenotypic variation in transgenic Drosophila larvae expressing ancestral HD proteins.

Morphologies of individual larvae were classified into three groups as follows: WTL: Wild-type Like, ΔHead: larvae missing one or more head segments with normal abdominal and thoracic segments: ΔAbdomen: larvae with variable defects in abdominal segments but normal head and thorax as described in Materials and Methods. **A-G.** Percentages of first instar larvae that show WTL, ΔHead and ΔAbdomen phenotypes within each transgenic line. **H.** Survival rates of embryos to adulthood in the corresponding transgenic lines. DmBcd: *D.melanogaster* Bcd. See also Figure S2.

Figure 4: hb and eve expression patterns in lines exhibiting different levels of phenotypic rescue.

A-L. HDs with significant partial to full rescue activities are tested for their *hb* and *eve* expression pattern differences. Tested HDs are labeled as in Figure 3. **A, C, E, G, I, K.** Representative stage 5 embryos (same embryos as in Figure S3) stained by *in situ* hybridization to detect *hb*. Panels show the analysis of multiple embryos to calculate the average posterior

boundary position (% EL; anterior tip = 100%) of *hb* at stage 5. Each horizontal line in each panel represents the anterior *hb* expression pattern in a single embryo, and the average pbp is denoted by a vertical red line. **B, D, F, H, J, L.** Representative stage 5 embryos stained for *eve* expression. *eve* stripe 1 positions were calculated from more 20 individual embryos and are denoted by vertical black lines. See also Figure S3.

Figure 5: A potential multi-step pathway for the evolution of the AncBcd HD.

Orange arrows represent amino acid substitutions in individual subdomains. In the first step, initial substitutions in the RH changed the DNA-binding preferences of the HD, and allowed it to bind to RNA. In step 2, these initial substitutions were followed by additional changes in either the NT or the H1 subdomain, each of which could have significantly augmented the *in vivo* activities of the evolving HD in a small percentage of embryos. In a third step, substitutions in the unchanged subdomain (H1 for RH+NT or NT for RH+H1) would further increase patterning activity and raise the survival rate to almost control levels.

Table 1: Constructs used in this paper are shown in the order of their appearance in the text.

HD constructs	**Description**
AncZB: *[nt-h1-h2-rh]*	Ancestor of Zen and Bcd [22]
AncZB_**K50R54**	AncZB with forward substitutions of q50>K and m54>R
AncZB_**RhDiag**	AncZB with diagnostic forward substitutions in RH
AncZB_**RH:** *[nt-h1-h2-***RH***]*	AncZB with all forward substitutions in RH
AncZB_**AllDiag**	AncZB with all diagnostic forward substitutions
[**CH(***nt***-H1-H2-RH)**]	AncBcd with reverse substitutions in N terminal arm (NT)
[**CH(NT-***h1***-H2-RH)**]	AncBcd with reverse substitutions in Helix 1
[**CH(NT-H1-***h2***-RH)**]	AncBcd with reverse substitutions in Helix 2
[**CH(NT-***h1***-***h2***-RH**)]	Forward substitutions in RH and NT combined
[**CH(***nt***-H1-***h2***-RH**)]	Forward substitutions in RH and H1 combined
AncBcd: [**NT-H1-H2-RH**]	ancestor of Bcd homologs[22]

STAR Methods:

Drosophila stocks, cloning, and transgenesis

We used the following stocks from our own lab for these experiments: yw (wild type), *Cyo* bcd^{+}/*Sco*; bcd^{E1}/bcd^{E1}, yw; TM3B, Sb, Ser/D and ΦC31 (y+); 38F1 (w+). We cloned an injection plasmid (piattB40-Bcd) containing two inverted ΦC31-specific recombination sequences, a Gmr-GFP reporter and a polylinker flanked by 1.9kb *bcd* promoter and 0.8kb 3'UTR. The *bcd* coding region was amplified by PCR from pBS-SK+ cDNA clones, digested with RsrII and AscI and ligated into piattB40-Bcd in between Bcd promoter and 3'UTR. This main plasmid was used to generate Dm Bcd protein with different ancestral HDs. We used standard cloning techniques to generate homeodomain swaps and residue changes. Gene Blocks coding for the ancestral and chimeric HD sequences together with the flanking Bcd coding sequence were obtained from Integrated DNA Technologies (IDT). They were digested with AscI and BspEI and ligated to the piattB40-Bcd vector digested with the same restriction enzymes. The cloned sequences were confirmed by sanger sequencing before and after transfection. All transgenic lines were generated using the ΦC31 integration system (Recombination mediated cassette exchange, RMCE), and constructs were integrated into the 38F1 landing site on the second chromosome [46]. Each transgene was crossed to *Cyo* bcd^{+}/*Sco*; bcd^{E1}/bcd^{E1} to generate *Cyo* bcd^{+}/*[transgene]*; bcd^{E1}/bcd^{E1} stocks. Embryos and larvae from homozygous transgenic females were assayed for gene expression and cuticle phenotype.

In situ hybridization, immunohistochemistry, and image processing

In situ hybridizations were performed as previously described [47]. Briefly, embryos 1-3 hours AEL (after egg laying) were dechorionated 2 minutes in 100% bleach, fixed and devitellinized in a biphasic fixation solution containing 3 ml 1X PBS, 1 ml 37% Formaldehyde and 4 ml Heptane for 25 minutes on a shaker at RT. Fixed and permeabilized embryos were incubated with DIG or Fluorescein-labeled RNA probes and the labeled probes were detected by Alkaline Phosphatase (AP)-conjugated primary antibodies (Roche Cat# 11093274910, RRID:AB_514497) and Roche Cat# 11426338910, RRID:AB_514504) by using NBT/BCIP solution (Roche Cat# 19315121). RNA expression was observed by Zeiss Axioskop microscopy.

Guinea pig anti-Cad [48] (1:400) and Alexa Fluor conjugated 647 donkey anti-guinea pig (1:500) (Molecular Probes Cat# A-21447, RRID:AB_141844) were used to examine Cad protein expression. All antibodies were diluted in PBT (1X PBS with 0.1% Tween). Data for immunostaining images were collected on a Leica TCS SP8 confocal microscope using the Leica confocal analysis software.

Larval Phenotype Analyses

Cuticle preparations were performed on embryos aged 24-30 hours at 25ºC as previously described [49]. Briefly, larvae were dechorionated for 2 minutes in 100% bleach, and a 1:1 mixture of methanol and heptane was used to remove the vitelline membrane and fix the larvae. Then these larvae were mounted in 1:1 mixture of Hoyer's medium [50] and lactic acid and incubated o/n at 65ºC to digest inner tissues.

Dark-field views of whole larvae were imaged at 200X magnification; DIC images of cephalic regions are imaged at 400X. Each image was sorted into one of three categories that encompassed the variation in phenotype of first instar larvae both within each transgenic fly line, and across all fly lines analyzed. The different phenotypic categories are; WTL: larvae containing all head segments [MH, LG, VA, DA, and DBr (Dorsal Bridge), three thoracic segments, and eight abdominal segments], ΔHead: larvae missing one or more head segments with normal abdominal and thoracic segments, ΔAbdomen: larvae with variable defects in abdominal segments but normal head and thorax. If an embryo showed both abdominal defects and head defects, the abdominal defect was given priority and that embryo was classified as ΔAbdomen. The number of WTL, ΔHead and ΔAbdomen larvae were counted for each transgenic line, tabulated and graphed as a percentage of the total number of embryos analyzed for that transgenic line.

Phylogenetics and Ancestral Sequence Construction

HD sequences were predicted as described in [22].

Measuring hb Posterior Boundary Positions (PBPs)

To measure PBPs of *hb* anterior expression, stained embryos of appropriate ages were imaged at 200× on a Zeiss Axioskop. Briefly, coordinates were established for each embryo so that the x- and y-axes were tangential to the ventral and anterior sides, respectively. A-P positions were displayed as percent of embryo length (EL%) with the anterior pole as 100%. PBPs were determined by visual estimation as the distance from the anterior tip to the most posterior position of *hb* anterior expression and these results were confirmed by ImageJ (ImageJ, RRID: SCR_003070) [51], [52] analyses.

Embryo images were loaded into ImageJ, and a Region of Interest (ROI) that was approximately 35% width of DV length from 95% to 60% (where 100% is most dorsal side) to analyze expression patterns was generated. The width of the ROI was kept constant when imaging all embryos, but the length was varied such that the length of the ROI spanned the length of the whole embryo. For each embryo, an intensity profile plot (intensity v. position along embryo length) was generated for the ROI. The midpoint of the curve that represents the edge of the boundary of expression of target gene was selected as the position of the boundary of gene expression. This numerical position was divided by the total embryo length to normalize the PBPs by % EL, and PBPs from individual embryos were averaged. 100% EL denotes the anterior tip, and 0% represents the posterior tip of the embryo.

Measuring eve patterns

To measure *eve* stripe patterns, stained embryos at stage 5 were positioned as described above and imaged at 200X. For each embryo, an intensity profile plot (intensity vs position along embryo length, where 100% denotes the anterior tip) was generated using ImageJ (ImageJ, RRID: SCR_003070) [51], [52], and analyzed using our Embryo Analyzer tool, which serves to take ImageJ plots of fly embryos, and convert them to produce a single file of normalized intensities along the AP axis of n number embryos.

Resource Availability

Lead contact: Pinar Onal

Materials availability: All unique materials generated in this study are available upon request.

Data and code availability: The code for hb and eve stripe pattern measurement supporting the current study have not been deposited in a public repository because of the specificity of the measurement to our case but are available from the corresponding author on request.

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Dr. Pinar Onal (pinar.onal@gmail.com)

References

[1] S. Ohno, *Evolution by Gene Duplication*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1970.

[2] F. A. Kondrashov, I. B. Rogozin, Y. I. Wolf, and E. V Koonin, "Selection in the evolution of gene duplications," *Genome Biol.*, vol. 3, no. 2, p. research0008.1, Jan. 2002, doi: 10.1186/gb-2002-3-2-research0008.

[3] F. R. Schubert, K. Nieselt-Struwe, and P. Gruss, "The Antennapedia-type homeobox genes have evolved from three precursors separated early in metazoan evolution," *Proc. Natl. Acad. Sci.*, vol. 90, no. 1, pp. 143–147, Jan. 1993, doi: 10.1073/PNAS.90.1.143.

[4] J. Zhang and M. Nei, "Evolution of Antennapedia-Class Homeobox Genes," *Genetics*, vol. 142, no. 1, p. 295, 1996, Accessed: May 21, 2020. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1206958/.

[5] C. Kappen, K. Schughart, and F. H. Ruddle, "Two steps in the evolution of Antennapedia-class vertebrate homeobox genes," *Proc. Natl. Acad. Sci.*, vol. 86, no. 14, pp. 5459–5463, Jul. 1989, doi: 10.1073/PNAS.86.14.5459.

[6] D. Duboule and P. Dollé, "The structural and functional organization of the murine HOX gene family resembles that of Drosophila homeotic genes.," *EMBO J.*, vol. 8, no. 5, pp. 1497–1505, May 1989, doi: 10.1002/j.1460-2075.1989.tb03534.x.

[7] J. M. Greer, J. Puetz, K. R. Thomas, and M. R. Capecchi, "Maintenance of functional equivalence during paralogous Hox gene evolution," *Nature*, vol. 403, no. 6770, pp. 661–665, Feb. 2000, doi: 10.1038/35001077.

[8] F. Falciani *et al.*, "Class 3 Hox genes in insects and the origin of zen.," *Proc. Natl. Acad. Sci.*, vol. 93, no. 16, pp. 8479–8484, Aug. 1996, doi: 10.1073/pnas.93.16.8479.

[9] U. Schmidt-ott, A. M. Rafiqi, and S. Lemke, "Hox3/zen and the Evolution of Extraembryonic Epithelia in Insects."

[10] M. Stauber *et al.*, "The anterior determinant bicoid of Drosophila is a derived Hox class 3 gene.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 96, no. 7, pp. 3786–9, Mar. 1999, doi: 10.1073/pnas.96.7.3786.

[11] M. Stauber, A. Prell, and U. Schmidt-Ott, "A single Hox3 gene with composite bicoid and zerknullt expression characteristics in non-Cyclorrhaphan flies.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 99, no. 1, pp. 274–9, Jan. 2002, doi: 10.1073/pnas.012292899.

[12] S.-K. Chan and G. Struhl, "Sequence-specific RNA binding by Bicoid," *Nature*, vol. 388, no. 6643, pp. 634–634, Aug. 1997, doi: 10.1038/41692.

[13] R. Rivera-Pomar, D. Niessing, U. Schmidt-Ott, W. J. Gehring, and H. Jacklë, "RNA binding and translational suppression by bicoid," *Nature*, vol. 379, no. 6567, pp. 746–749, Feb. 1996, doi: 10.1038/379746a0.

[14] W. Driever and C. Nüsslein-Volhard, "A gradient of bicoid protein in Drosophila embryos.," *Cell*, vol. 54, no. 1, pp. 83–93, Jul. 1988, [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/3383244.

[15] W. Driever, G. Thoma, and C. Nüsslein-Volhard, "Determination of spatial domains of

zygotic gene expression in the Drosophila embryo by the affinity of binding sites for the bicoid morphogen," *Nature*, vol. 340, no. 6232, pp. 363–367, Aug. 1989, doi: 10.1038/340363a0.

[16] G. Struhl, K. Struhl, and P. M. Macdonald, "The gradient morphogen bicoid is a concentration-dependent transcriptional activator," *Cell*, vol. 57, no. 7, pp. 1259–1273, Jun. 1989, doi: 10.1016/0092-8674(89)90062-7.

[17] H. Chen, Z. Xu, C. Mei, D. Yu, and S. Small, "A system of repressor gradients spatially organizes the boundaries of Bicoid-dependent target genes.," *Cell*, vol. 149, no. 3, pp. 618–29, Apr. 2012, doi: 10.1016/j.cell.2012.03.018.

[18] A. Nasiadka, B. H. Dietrich, and H. M. Krause, "Anterior – posterior patterning in the Drosophila embryo," vol. 12, 2002.

[19] D. Niessing, W. Driever, F. Sprenger, H. Taubert, H. Jäckle, and R. Rivera-Pomar, "Homeodomain position 54 specifies transcriptional versus translational control by Bicoid.," *Mol. Cell*, vol. 5, no. 2, pp. 395–401, Feb. 2000, [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/10882080.

[20] D. Niessing, S. Blanke, and H. Ja, "Bicoid associates with the 5 ˜-cap-bound complex of caudal mRNA and represses translation," pp. 2576–2582, 2002, doi: 10.1101/gad.240002.2576.

[21] H. G. Frohnhöfer and C. Nüsslein-Volhard, "Organization of anterior pattern in the Drosophila embryo by the maternal gene bicoid," *Nature*, vol. 324, no. 6093, pp. 120–125, Nov. 1986, doi: 10.1038/324120a0.

[22] Q. Liu *et al.*, "Ancient mechanisms for the evolution of the bicoid homeodomain's function in fly development.," *Elife*, vol. 7, Oct. 2018, doi: 10.7554/eLife.34594.

[23] M. J. Harms and J. W. Thornton, "Analyzing protein structure and function using ancestral gene reconstruction," *Current Opinion in Structural Biology*, vol. 20. pp. 360–366, 2010, doi: 10.1016/j.sbi.2010.03.005.

[24] J. Treisman, P. Gönczy, M. Vashishtha, E. Harris, and C. Desplan, "A single amino acid can determine the DNA binding specificity of homeodomain proteins.," *Cell*, vol. 59, no. 3, pp. 553–62, Nov. 1989, doi: 10.1016/0092-8674(89)90038-X.

[25] M. B. Noyes, R. G. Christensen, A. Wakabayashi, G. D. Stormo, M. H. Brodsky, and S. A. Wolfe, "Analysis of Homeodomain Specificities Allows the Family-wide Prediction of Preferred Recognition Sites," *Cell*, 2008, doi: 10.1016/j.cell.2008.05.023.

[26] J. M. Baird-Titus, K. Clark-Baldwin, V. Dave, C. A. Caperelli, J. Ma, and M. Rance, "The solution structure of the native K50 Bicoid homeodomain bound to the consensus TAATCC DNA-binding site.," *J. Mol. Biol.*, vol. 356, no. 5, pp. 1137–51, Mar. 2006, doi: 10.1016/j.jmb.2005.12.007.

[27] M. Simpson-Brose, J. Treisman, and C. Desplan, "Synergy between the hunchback and bicoid morphogens is required for anterior patterning in Drosophila.," *Cell*, vol. 78, no. 5, pp. 855–65, Sep. 1994, [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/8087852.

[28] A. Porcher and N. Dostatni, "The Bicoid Morphogen System," *Curr. Biol.*, vol. 20, no. 5, pp. 249–254, 2010, doi: 10.1016/j.cub.2010.01.026.

[29] A. Ochoa-Espinosa *et al.*, "The role of binding site cluster strength in Bicoid-dependent patterning in Drosophila.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 102, no. 14, pp. 4960–5, Apr. 2005, doi: 10.1073/pnas.0500373102.

[30] M. D. Schroeder, C. Greer, and U. Gaul, "How to make stripes: deciphering the transition from non-periodic to periodic patterns in Drosophila segmentation.," *Development*, vol. 138, no. 14, pp. 3067–78, Jul. 2011, doi: 10.1242/dev.062141.

[31] G. Struhl, P. Johnston, and P. A. Lawrence, "Control of Drosophila body pattern by the hunchback morphogen gradient," *Cell*, vol. 69, no. 2, pp. 237–249, Apr. 1992, doi: 10.1016/0092-8674(92)90405-2.

[32] X. Wu, V. Vasisht, D. Kosman, J. Reinitz, and S. Small, "Thoracic Patterning by the Drosophila Gap Gene hunchback," *Dev. Biol.*, vol. 237, no. 1, pp. 79–92, Sep. 2001, doi: 10.1006/DBIO.2001.0355.

[33] M. Hülskamp, C. Pfeifle, and D. Tautz, "A morphogenetic gradient of hunchback protein organizes the expression of the gap genes Krüppel and knirps in the early Drosophila embryo," *Nature*, vol. 346, no. 6284, pp. 577–580, Aug. 1990, doi: 10.1038/346577a0.

[34] D. Yu and S. Small, "Precise registration of gene expression boundaries by a repressive morphogen in Drosophila.," *Curr. Biol.*, vol. 18, no. 12, pp. 868–76, Jun. 2008, doi: 10.1016/j.cub.2008.05.050.

[35] F. Liu, A. H. Morrison, and T. Gregor, "Dynamic interpretation of maternal inputs by the Drosophila segmentation gene network," *Proc. Natl. Acad. Sci.*, vol. 110, no. 17, pp. 6724–6729, Apr. 2013, doi: 10.1073/pnas.1220912110.

[36] Z. Xu, H. Chen, J. Ling, D. Yu, P. Struffi, and S. Small, "Impacts of the ubiquitous factor Zelda on Bicoid-dependent DNA binding and transcription in Drosophila.," *Genes Dev.*, vol. 28, no. 6, pp. 608–21, Mar. 2014, doi: 10.1101/gad.234534.113.

[37] M. Mir *et al.*, "Dense Bicoid hubs accentuate binding along the morphogen gradient.," *Genes Dev.*, vol. 31, no. 17, pp. 1784–1794, Sep. 2017, doi: 10.1101/gad.305078.117.

[38] R. R. Datta *et al.*, "A feed-forward relay integrates the regulatory activities of Bicoid and Orthodenticle via sequential binding to suboptimal sites.," *Genes Dev.*, vol. 32, no. 9–10, pp. 723–736, May 2018, doi: 10.1101/gad.311985.118.

[39] C. E. Hannon, S. A. Blythe, and E. F. Wieschaus, "Concentration dependent chromatin states induced by the bicoid morphogen gradient," *Elife*, vol. 6, p. e28275, Sep. 2017, doi: 10.7554/eLife.28275.

[40] R. Schröder, "The genes orthodenticle and hunchback substitute for bicoid in the beetle Tribolium," *Nature*, vol. 422, no. 6932, pp. 621–625, Apr. 2003, doi: 10.1038/nature01536.

[41] J. A. Lynch, A. E. Brent, D. S. Leaf, M. Anne Pultz, and C. Desplan, "Localized maternal orthodenticle patterns anterior and posterior in the long germ wasp Nasonia," *Nature*, vol. 439, no. 7077, pp. 728–732, Feb. 2006, doi: 10.1038/nature04445.

[42] Q. Gao and R. Finkelstein, "Targeting gene expression to the head: the Drosophila orthodenticle gene is a direct target of the Bicoid morphogen," *Development*, vol. 125, no. 21, pp. 4185–4193, Nov. 1998, Accessed: Jan. 12, 2016. [Online]. Available: http://dev.biologists.org/content/125/21/4185.abstract?ijkey=93cb26aef4cb3eef47761ecc2

ee591bbc2345234&keytype2=tf_ipsecsha.

[43] J. Lynch and C. Desplan, "Evolution of development: beyond bicoid.," *Curr. Biol.*, vol. 13, no. 14, pp. R557-9, Jul. 2003, doi: 10.1016/s0960-9822(03)00472-x.

[44] Y. Yoon *et al.*, "Embryo polarity in moth flies and mosquitoes relies on distinct old genes with localized transcript isoforms," *Elife*, vol. 8, Oct. 2019, doi: 10.7554/eLife.46711.

[45] J. Klomp *et al.*, "A cysteine-clamp gene drives embryo polarity in the midge Chironomus," *Science (80-.).*, vol. 348, no. 6238, pp. 1040–1042, May 2015, doi: 10.1126/science.aaa7105.

[46] J. R. Bateman, A. M. Lee, and C. Wu, "Site-specific transformation of Drosophila via phiC31 integrase-mediated cassette exchange.," *Genetics*, vol. 173, no. 2, pp. 769–77, Jun. 2006, doi: 10.1534/genetics.106.056945.

[47] S. Small, "In vivo analysis of lacZ fusion genes in transgenic Drosophila melanogaster," *Methods Enzymol.*, 2000, doi: 10.1016/s0076-6879(00)26052-7.

[48] D. Kosman, S. Small, and J. Reinitz, "Rapid preparation of a panel of polyclonal antibodies to Drosophila segmentation proteins," *Dev. Genes Evol.*, vol. 208, no. 5, pp. 290–294, 1998, doi: 10.1007/s004270050184.

[49] X. Wu, R. Vakani, and S. Small, "Two distinct mechanisms for differential positioning of gene expression borders involving the Drosophila gap protein giant," *Development*, vol. 125, no. 19, 1998.

[50] L. E. Anderson, "Hoyer's Solution as a Rapid Permanent Mounting Medium for Bryophytes," *Bryologist*, vol. 57, no. 3, p. 242, Sep. 1954, doi: 10.2307/3240091.

[51] C. T. Rueden *et al.*, "ImageJ2: ImageJ for the next generation of scientific image data," *BMC Bioinformatics*, vol. 18, no. 1, p. 529, Dec. 2017, doi: 10.1186/s12859-017-1934-z.

[52] C. A. Schneider, W. S. Rasband, and K. W. Eliceiri, "NIH Image to ImageJ: 25 years of image analysis," *Nat. Methods*, vol. 9, no. 7, pp. 671–675, Jul. 2012, doi: 10.1038/nmeth.2089.

Figure 1

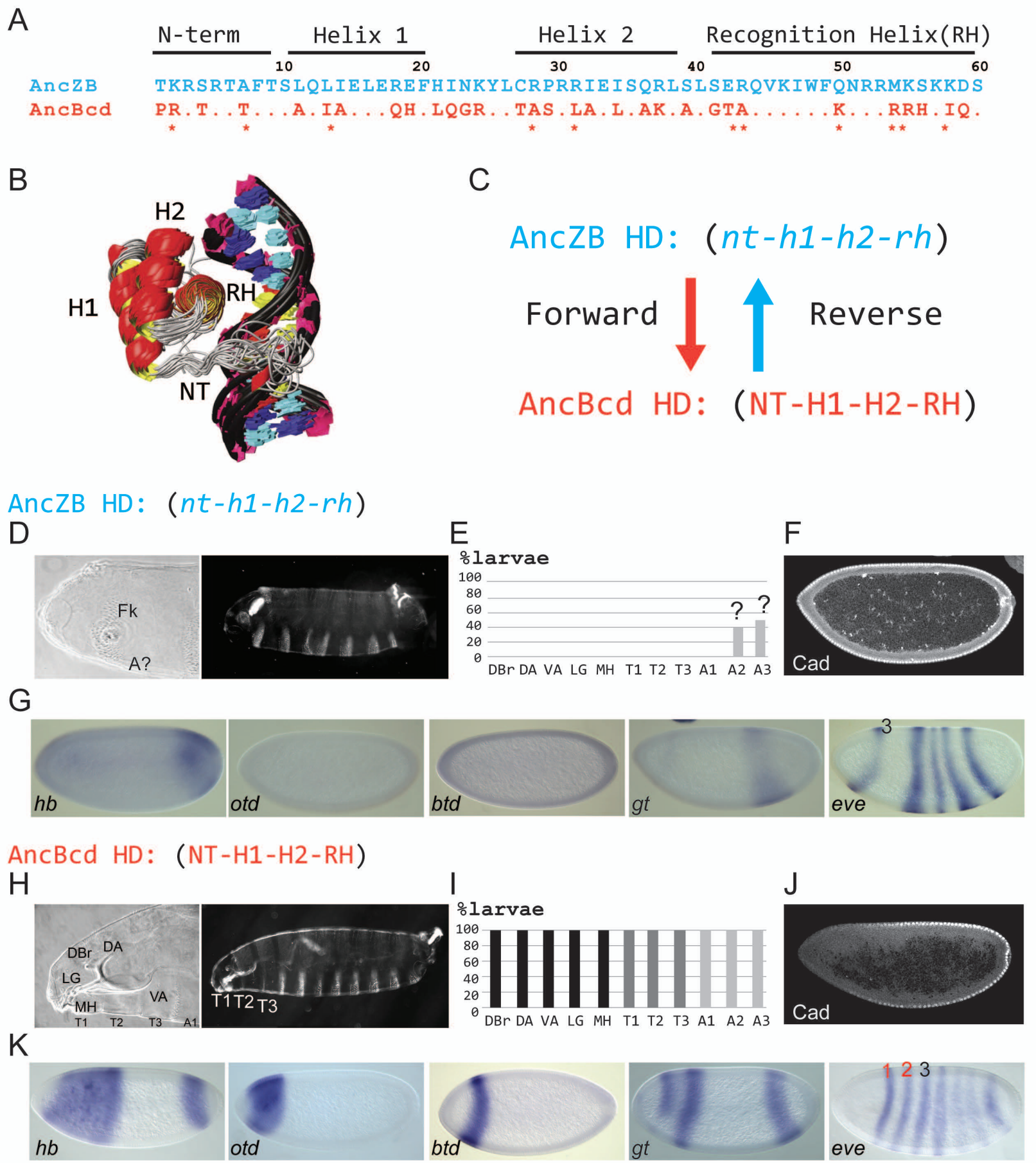
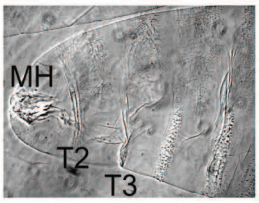
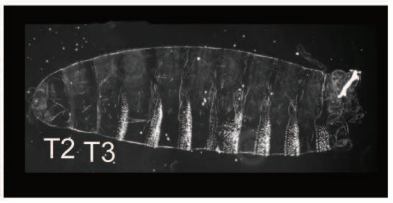

A
N-term
Helix 1
Helix 2
Recognition Helix(RH)
10
20
30
40
50
60
AncZB TKRSRTAFTSLQLIELEREFHINKYLCRPRRIEISQRLSLSERQVKIWFQNRRMKSKKDS
AncBcd PR.T..T...A.IA...QH.LQGR..TAS.LA.L.AK.A.GTA.....K...RRH.IQ.
B
H2
H1
RH
NT
C
AncZB HD: (nt-h1-h2-rh)
Forward
Reverse
AncBcd HD: (NT-H1-H2-RH)
AncZB HD: (nt-h1-h2-rh)
D
Fk
A?
E
%larvae
100
80
60
40
20
0
? ?
DBr DA VA LG MH T1 T2 T3 A1 A2 A3
F
Cad
G
hb
otd
btd
gt
3
eve
AncBcd HD: (NT-H1-H2-RH)
H
DBr
DA
LG
MH
VA
T1 T2 T3 A1
T1 T2 T3
I
%larvae
100
80
60
40
20
0
DBr DA VA LG MH T1 T2 T3 A1 A2 A3
J
Cad
K
hb
otd
btd
gt
1 2 3
eve

Figure 2

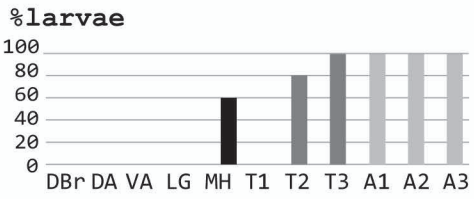
AncZB_K50R54

A

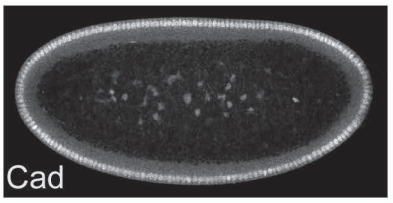




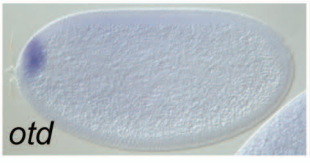
B

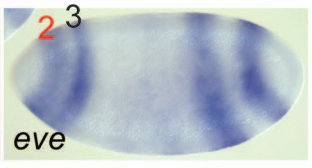


C



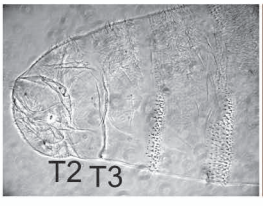
D

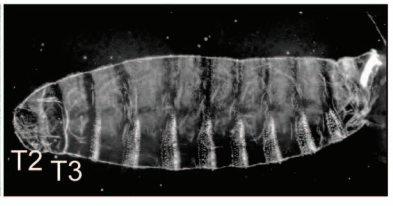




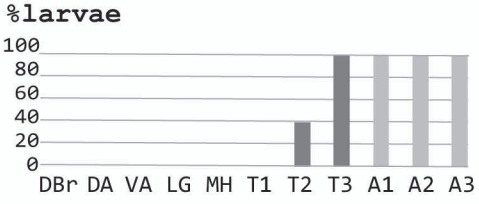
AncZB_RHdiag

E

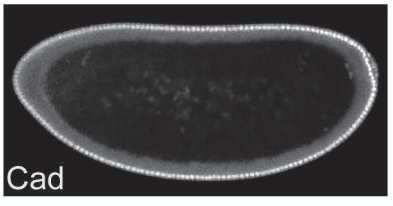




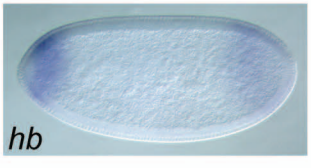
F

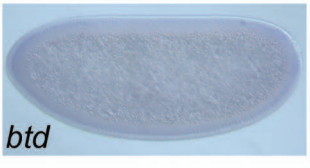


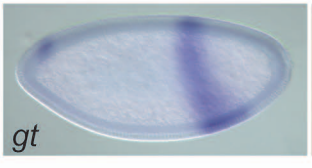
G

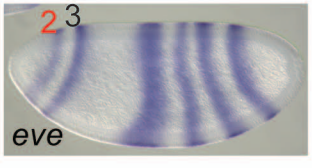


H



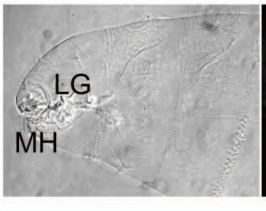




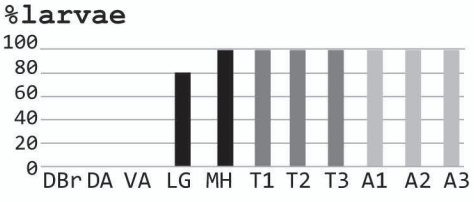


AncZB_RH

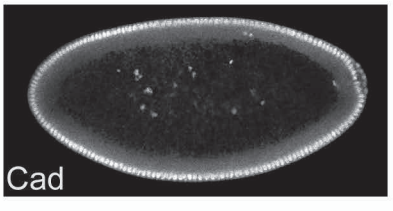
I



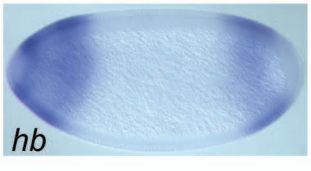
J

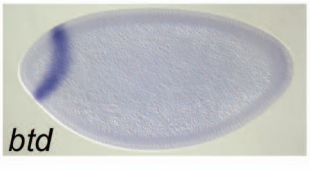


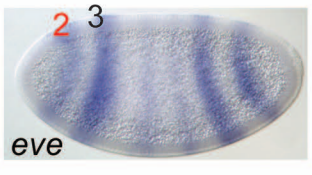
K



L

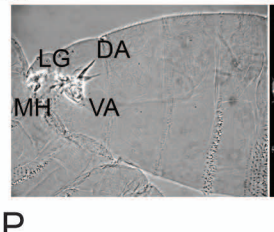


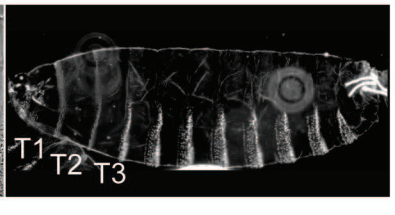




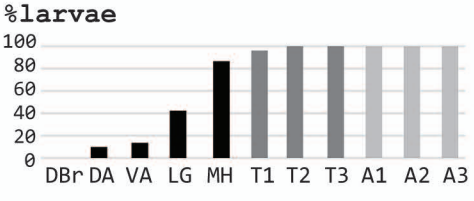
AncZB_Alldiag

M

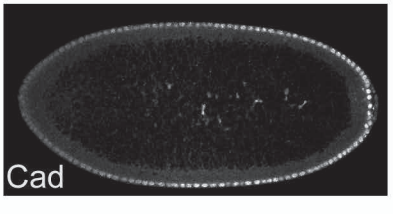




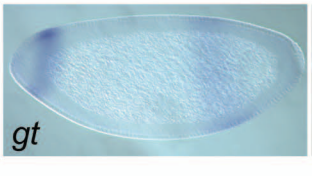
N



O



P



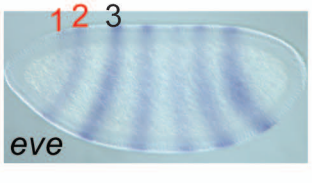


Figure 3

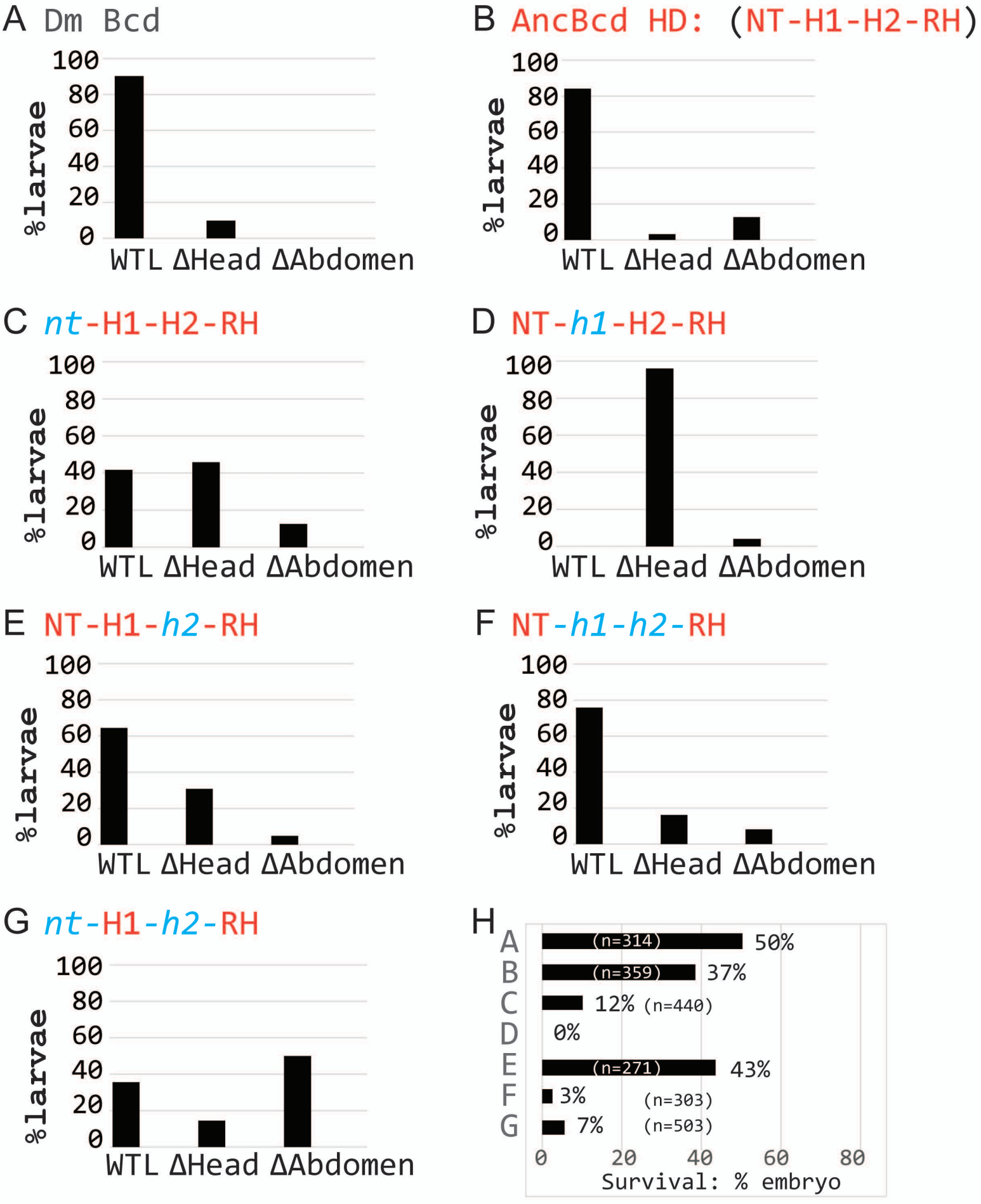
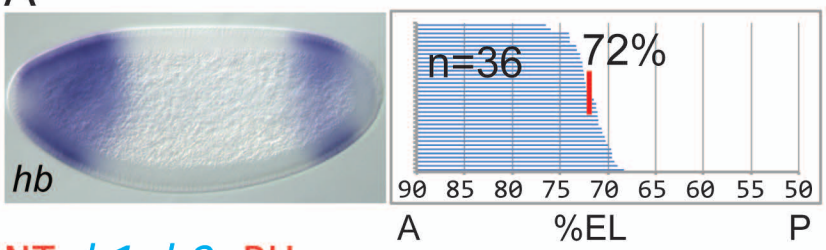

A Dm Bcd
B AncBcd HD: (NT-H1-H2-RH)
C nt-H1-H2-RH
D NT-h1-H2-RH
E NT-H1-h2-RH
F NT-h1-h2-RH
G nt-H1-h2-RH
%larvae
100
80
60
40
20
0
WTL ΔHead ΔAbdomen
H
A (n=314) 50%
B (n=359) 37%
C 12% (n=440)
D 0%
E (n=271) 43%
F 3% (n=303)
G 7% (n=503)
0 20 40 60 80
Survival: % embryo

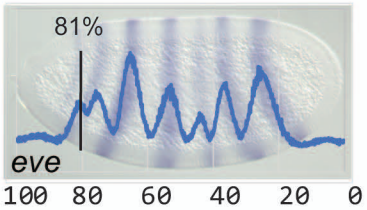
Figure 4

NT-*h1*-H2-RH

A

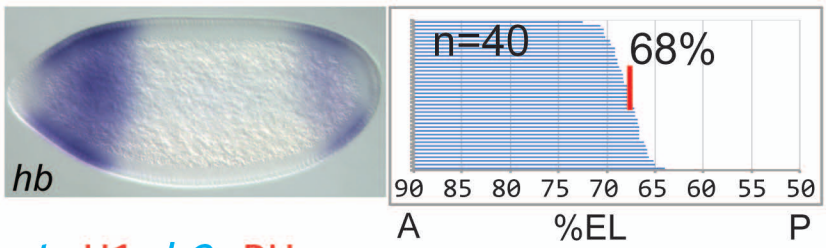


B

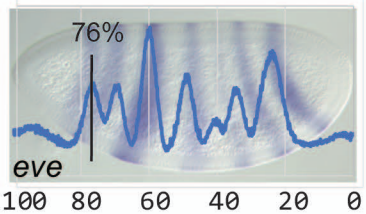


NT-*h1*-*h2*-RH

C

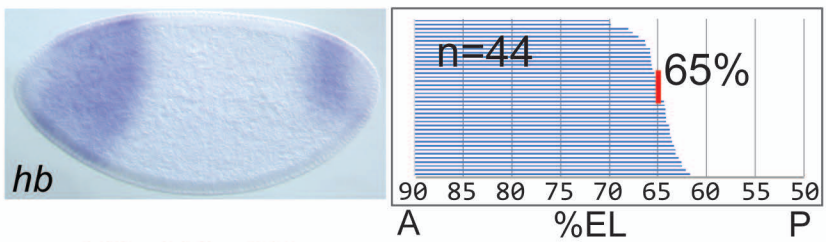


D

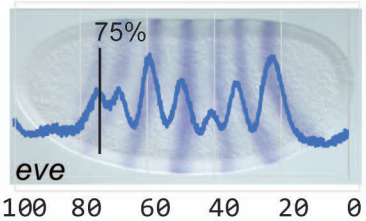


nt-H1-*h2*-RH

E

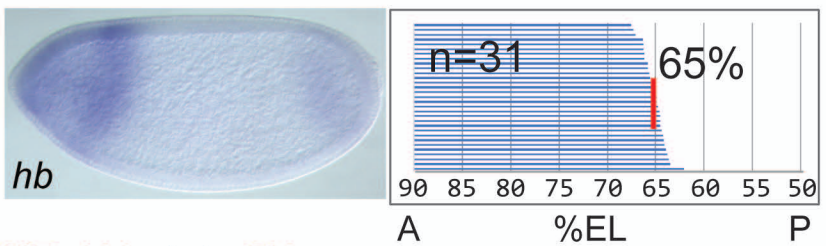


F

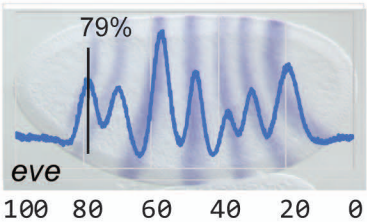


nt-H1-H2-RH

G

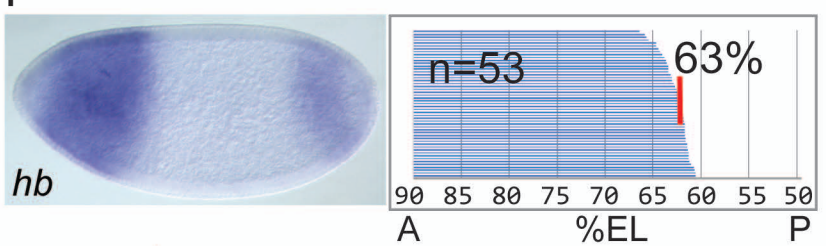


H

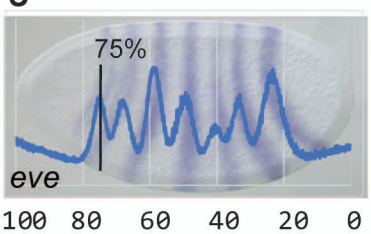


NT-H1-*h2*-RH

I

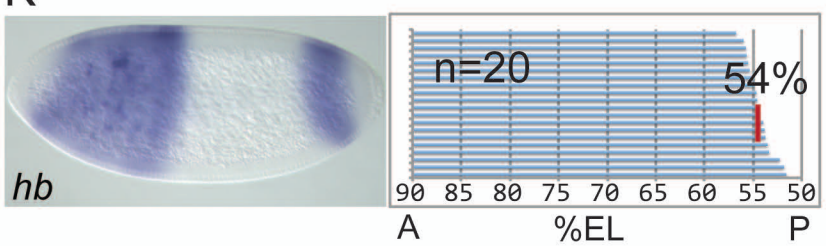


J



AncBcd HD: (NT-H1-H2-RH)

K



L

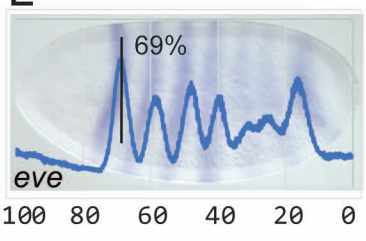
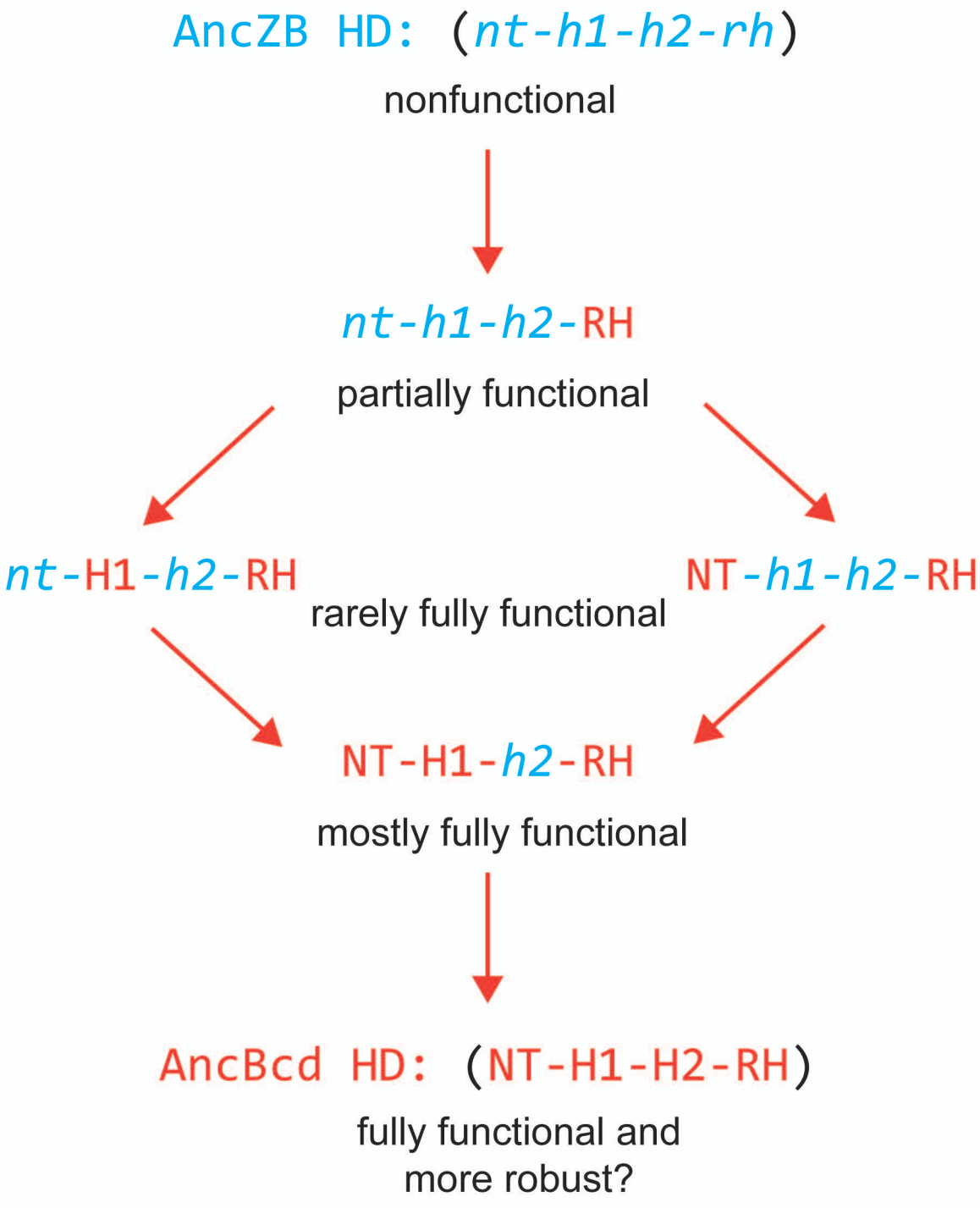


Figure 5


AncZB HD: (nt-h1-h2-rh)
nonfunctional
nt-h1-h2-RH
partially functional
nt-H1-h2-RH
NT-h1-h2-RH
rarely fully functional
NT-H1-h2-RH
mostly fully functional
AncBcd HD: (NT-H1-H2-RH)
fully functional and
more robust?