Figures and figure supplements

Ancient mechanisms for the evolution of the bicoid homeodomain’s function in fly development

Qinwen Liu et al
Figure 1. Role of the homeodomain in functional divergence of *D. melanogaster* Zen and Bcd. Offspring of transgenic *bcd*-deficient *D. melanogaster* containing two copies of various replacement transgenes are shown in cuticle preparations of larval head (A, K, U) and whole body (B, L, V); in situ
hybridization experiments and antibody stains in transgenic embryos show expression of Bcd’s translational (C, M, W) and transcriptional target genes (D–J, N–T, X–DD). Replacement transgenes were Bcd from D. melanogaster (DmBcd, A–J), Zen from D. melanogaster (DmZen, K–T), and a chimera of DmZen in which the homeodomain (HD) has been replaced by the HD of DmBcd (DmZen-BcdHD, U–DD). All three transgenes contain the D. melanogaster bcd promoter (BP), which directs maternal expression, and the bcd 3’UTR sequence required for anterior localization of the transgenic RNA. In the icons for each construct, protein-coding sequences from DmBcd are light blue and DmZen are red. Triangles show AttB sites used for recombination-mediated cassette exchange. For cuticle preps, mouthhooks (MH), dorsal arm (DA), ventral arm (VA), cephalic structures (Ceph), thoracic segments (T1–T3), abdominal segments (A1–A8), and Filzkörper (Fk) are indicated when present. An anti-Cad antibody was used to detect Cad protein. Antisense probes used in in situ hybridization experiments are indicated (hb: hunchback, kni: knirps, gt: giant, otd: orthodenticle, ems: empty-spiracle, btd: buttonhead, eve: even-skipped mRNAs). For comparison to wild-type, bcd-deficient, and a reciprocal construct of Bcd containing the ZenHD, see Figure 1—figure supplement 1.

DOI: https://doi.org/10.7554/eLife.34594.002
Figure 1—figure supplement 1. Phenotypes of wildtype, bcd-null, and transgenic D. melanogaster. (A–J) Flies of genotype yw (wild-type) are not modified at the bcd locus. Shown are cuticle preparations of larval head (A) and whole body (B); embryonic expression patterns are shown for Bcd’s.
translational regulatory target Caudal (Cad, C) and transcriptional targets (D–J). (K–T) Flies of genotype bcd\(^{E1}\) carry null alleles at the bcd locus. (U–DD) Flies of genotype DmBcdZenHD are bcd\(^{E1}\) null but carry transgenic alleles consisting of the D. melanogaster bcd locus, including the coding region (blue) but with the HD replaced by that of the D. melanogaster zen gene (red). (EE–GG) Embryonic expression of bcd and zen mRNAs in wild-type and transgenic embryos: (EE) bcd mRNA in yw flies, (FF) transgenic bcd mRNA in bcd\(^{E1}\) flies transformed with D. melanogaster bcd, and (GG) zen mRNA in bcd\(^{E1}\) flies transformed with the D. melanogaster zen.
Figure 2. Reconstruction of ancestral HD sequences. (A) Phylogeny of Bcd, Zen, and Hox3 HD amino acid sequences from insects and other arthropods. One representative sequence is shown among sequence entries with no sequence diversity. Reconstructed ancestral nodes are labeled; the mean posterior probability across sites for each maximum likelihood ancestral sequence is shown. Branch support labels show the approximate likelihood ratio statistic. Scale bar, 0.7 amino acid changes per site. The tree was constrained to reproduce well-corroborated species relationships; for the maximum likelihood phylogeny, see Figure 2—figure supplement 1. For species names and sequences, see Supplementary file 1. (B) Amino acid sequences of AncZB, AncBcd and Bcd from D. melanogaster (Dm) HDs. For AncBcd and DmBcd, only residues that differ from the AncZB HD are shown. Of these, 11 substitutions (shown in blue) are diagnostic differences conserved among all sequences in the Bcd vs. Zen clades. Differences between AncBcd and DmBcd are labeled red. For site-specific posterior probabilities of ancestral states, see Supplementary file 2. DOI: https://doi.org/10.7554/eLife.34594.004
Figure 2—figure supplement 1. Unconstrained phylogeny. The maximum likelihood phylogeny of the Hox3/Bcd/Zen homeodomain protein family obtained with no topological constraint. Scale bar, expected amino acid substitutions per site. Branch labels, statistical clade support shown as the approximate likelihood ratio statistic, defined as two times the difference in log-likelihoods of the ML tree and the next-best tree containing a topological arrangement around that node.

DOI: https://doi.org/10.7554/eLife.34594.005
Figure 3. In vitro and in vivo functions of AncBcd and AncZB HDs. (A, B) Protein binding microarray (PBM) analysis of the DNA specificity of AncBcd, DmBcd, and AncZB HDs. Each dot represents one DNA 6-mer, plotted according to its binding enrichment (E-scores) in PBMs performed with two
different HDs; Pearson’s correlation coefficient for each comparison is shown. (C–E) Energy logos showing the site-specific sequence determinants of DNA binding by AncZB, AncBcd, and DmBcd HDs as determined from PBM data. The size of each letter is proportional to the absolute value of that base’s main effect on the energy of binding relative to the mean of all bases at that position. The y-axis is unitless after division by \(RT\). (F) Affinity of ancestral HDs for consensus Zen and Bcd DNA motifs (ZM and BM, sequences shown) in a quantitative fluorescence electrophoretic mobility shift assay. Several noncanonical Bcd targets previously shown to bind \(Drosophila\) Bcd were also tested. Column heights and error bars show the mean and SEM for three replicates. (G–Z) Cuticle phenotypes and gene expression patterns for offspring produced by \(bcd\)\(^{E1}\) females carrying two copies of a rescue construct expressing \(Dm\) Bcd protein (light blue) in which the HD has been replaced by the reconstructed AncBcd HD (dark blue, panels G–P) or the reconstructed AncZB HD (purple, panels Q–Z). Individual panels and constructs are labeled as in Figure 1. For replicates and alternative ancestral reconstructions, see Figure 3—figure supplement 1, Figure 4—figure supplement 1 and Supplementary files 3 and 4. For SPR analysis, see Figure 3—figure supplement 2 and Supplementary file 4.
Figure 3—figure supplement 1. Binding site preference is robust to uncertainty about the ancestral sequence reconstruction. Quantitative fluorescent EMSAs were performed using alt-all versions of AncZB and AncBcd HDs, which contain the state with the second highest posterior probability at all ambiguously reconstructed sites, defined as all those sites at which more than one state has posterior probability >0.20. Plots show the fraction of ZM and BM bound as the concentrations of AncZB HD-AltAll (A) or AncBcd HD-AltAll (B) were increased.

DOI: https://doi.org/10.7554/eLife.34594.007
Figure 3—figure supplement 2. Bcd evolved novel binding specificity by changes in sequence-specific dissociation rates. (A) Association and dissociation rates of reconstructed ancestral HDs to canonical Zen and Bcd motifs (ZM, BM) were measured using a surface plasmon resonance (SPR) assay. Raw binding traces (color in black) show SPR response when different concentrations of HD were applied to sensors with immobilized BM or ZM DNA elements. Binding modes with parameters fit to the experimental data are in red. (B) Isoaffinity plots show the dissociation ($k_{\text{off}}$) and association ($k_{\text{on}}$) rates of AncBcd (square) and AncZB (triangle) HDs on DNA elements ZM (cyan), and to BM (red-orange) obtained from SPR. Arrows indicate the evolutionary change from AncZB to AncBcd HD. Diagonal dotted lines mark isoaffinity lines of association constants ($K_a$, blue labels) that differ by factors of ten. (C) Effect of historical substitutions on off- and on-rates of binding to BM. Solid arrows show the effect of q50K; dotted arrows show the effect of all 30 other substitutions from the same historical interval.

DOI: https://doi.org/10.7554/eLife.34594.008
Figure 4. A single amino acid substitution changes AncZB’s in vitro binding preference. (A) Effect of reversing diagnostic historical substitutions on AncBcd-HD’s affinity for consensus Bcd and Zen motifs (BM and ZM), in a fluorescent EMSA assay. Column height and error bars show mean and SEM of three replicates. Upper and lower cases denote derived and ancestral states, respectively. (B) Effect of introducing historical substitutions to the derived state at site 50 and 54 into AncZB HD. (C) Binding of DNA motifs by AncZBHD-q50K is tightly correlated with binding by AncBcdHD. Each point is a DNA 6-mer plotted by the average of E-scores for 8-mers containing that 6-mer. Pearson correlation coefficient is shown. (D) Binding of DNA motifs by AncBcdHD-K50q is tightly correlated with binding by AncZB-HD. (E, F) The PBM energy logos of AncZBHD-q50K and AncBcdHD-K50q (compare to Figure 3C). (G) Effect of historical substitutions on global binding specificities in PBMs. Each column represents an 8-mer, colored by its PBM E-score, according to the color scale shown. Each row shows PBM of a reconstructed HD protein with or without residue 50 swapped between ancestral and derived states. 8-mers were clustered using the Manhattan distance metric. Only 8-mers with E-score ≥ 0.45 for at least one HD variant are shown. Logos at the bottom of the heatmap were created from the 8-mers in the two large clusters.

DOI: https://doi.org/10.7554/eLife.34594.009
**Figure 4—figure supplement 1.** 8mer binding profiles for replicate HD constructs. Each row represents one 8mer sequence, and each column represents one HD protein variant, colored by E-score using the heatmap scale shown at left. Rows and columns were clustered by similarity using the Manhattan distance metric. Each protein was tested in two replicate experiments, indicated by the suffix _1 or _2. Ancestral reconstructions represent the maximum likelihood ancestral sequence (unlabeled) or the ‘alt-all’ reconstruction, which contains the state with the second-highest posterior probability at all ambiguously reconstructed sites, defined as those in which more than one site has posterior probability >0.20. All HDs were tested with 15 additional flanking residues from the DmBcd protein (labeled B) or the DmZen protein (unlabeled). Only 8-mers with E-score ≥0.45 for at least one protein variant are shown.

DOI: https://doi.org/10.7554/eLife.34594.010
Figure 5. Effect of historical substitutions on Bcd functions in vivo. (A–J) Testing the necessity of q50K for AncBcd function. Cuticle phenotypes and Bcd-target gene expression patterns of embryos from bcd<sup>E1</sup> females carrying two copies of a Bcd rescue construct containing AncBcd-HD (dark blue)
with an amino acid reversion to the AncZB ancestral state at residue 50 (purple arrow). (K–NN) Testing the sufficiency of historical substitutions of q50K and m54R. Embryos expressing Bcd rescue constructs containing AncZB-HD (purple) with historical substitutions to the derived AncBcd state (blue arrows) at site 50 (K–T), 54 (U–DD), or both (EE–NN). Individual panels and constructs are labeled as in Figure 1. Stars indicate Bcd-dependent target gene activation induced by the indicated constructs.

DOI: https://doi.org/10.7554/eLife.34594.011