Supplementary material

Supplemental Figure 1 Phylogram of E(spl)-C bHLH protein from Drosophilid genomes.

Unrooted Bayseain phylogram of E(spl)-C and related bHLH proteins from *Drosophila melanogaster* (labelled with gene abbreviations), *Drosophila virilis* (labelled with Flybase identifiers prefixed with 'vir'), *Drosophila erecta* (ere), *Drosophila pseudoobscura* (pse), *Drosophila grimshawi* (gri), *Drosophila mojavensis* (moj) and *Drosophila anassanae* (ana). Clades are labelled with *melanogaster* gene names. E(spl)-C bHLH proteins are shown in shades of green.

Supplemental Figure 2 – Phylogeny of Kazal class protease inhibitors.

An unrooted Bayesian phylogeny of Kazal protease inhibitor domains. Kazal protease inhibitor domains were extracted from annotated protein sequences of *D. melanogaster, D. virilis, D. persmilis, A. aegypti, A. gambiae, C. pipens, B. mori, P. humanus, T. caseneum, A. mellifera, A. pisum, and D. pulex* using the hmmer suite of programs (Eddy 1998), against the Pfam Kazal 2 hmm model (PF07648) (Finn et al. 2008). Phylogeny was constructed using MrBayes (Ronquist and Huelsenbeck 2003) under the Jones amino acid substitution model. Posterior probabilities are shown on internal branches. The Jones model was chosen as the most appropriate model of amino acid substitution after preliminary analyses using MrBayes with mixed models. The Monte Carlo Markov Chain search was run with four chains over 1000000 generations with trees sampled every 1000 generations. The first 250000 trees were discarded as 'burn-in'. In general the tree is well resolved and orthologues of m1 are only detected in the other Drosophilid species sampled; *D. persimilis* and *D. virilis.* m1 is part of a *Drosophila* specific clade of Kazal protease inhibitors that include 5 additional *Drosophila* proteins, consistent with expansion of this family in *Drosophila.* The clade containing m1 has been extracted and enlarged for clarity and is shown on the right hand side of the full tree.

Supplemental Figure 3 - Bearded protein sequence features.

The genomes of the crustacean *D. pulex* and the hemimetabolous insect *A. pisum* encode a bearded family protein that is linked to bHLH transcription factors of the E(spl)-C family. A) putative amino acid sequence of the *Daphnia* and pea aphid bearded proteins. The N-terminal amphipathic ⟨-helix (B domain) is shown in blue and the NXANE(K/R)L motif (N motif) in red. The *Daphnia* sequence also has a partial G motif (GTFFWT) in green. No sequence corresponding to this motif is seen in the *Acyrthosiphon* ortholog. B) Helical wheel diagrams of the most likely amphipathic ⟨-helices for the bearded family proteins from *A. pisum* and *D. pulex*. Both helices have a strong hydrophobic face and no interspersed polar and non-polar residues. Lysine and arginine residues (blue) are located predominantly on one side of the putative helix in this protein. The opposite face is enriched in hydrophobic residues (yellow), the uncharged polar residues are shown in pink and mauve. Both the *Daphnia* and *Acyrthosiphon* helices have a proline residue within this region, making the helix most similar to that seen in the *Drosophila* Tom protein (Lai et al., 2000).

Supplemental Figure 4 - Putative cis-regulatory motifs around arthropod E(spl) complexes

Gene models were defined by comparing the predicted transcript with the appropriate genomic sequence using SPIDEY (Wheelan et al. 2001) and transcriptional orientation is indicated by the arrow in the diagram. Conserved binding sites were identified for Su(H) high affinity (TGTGRGAA), Su(H) paired (YGTGRGAAMN{10,60}KTTCYCACR (Bailey and Posakony 1995), Bearded box (AGCTTTA), GY box (GTCTTC) (Lai and Posakony 1997; Leviten et al. 1997) K box (TGTGAT) as well as A box proneural motif (GCAGSTG) (Singson et al. 1994). Models were visualised using Gene Palette (v 1.38)(Rebeiz and Posakony 2004). Note that for clarity some of the intervening genomic regions have been omitted from the diagram, this is indicated with a scale break. In *Drosophila* many Notch responsive genes, including many of the genes of the E(spl)-C are known to be regulated by miRNA binding sites in the 3' UTR. miRNAs are proposed to bind to the GY, Brd and K-box motifs (Lai et al. 2005). Identical motifs were identified in the genomic sequence

immediately downstream of the coding regions of a number of these genes in both honeybee and *Daphnia*, indicating that regulation of these transcripts by miRNA may be an evolutionary conserved feature. Although no Brd box sequences were found associated with *Daphnia* genes, indicating that regulation by this miRNA may have evolved in the lineage leading to insects or alternatively that both the seed sequence and miRNA have diverged such that it can not be detected. The genes encoded by the aphid E(spl)-C also have motifs corresponding to the Brd-box, GY box and K box. But these motifs in general lie much further away from the putative stop codon than they do in other species examined (between 380 bp and 1685 bp downstream). This may indicate that regulation of these genes in the pea aphid is not dependent on miRNA or alternatively that this species has unusually long 3' UTRs.

In both aphid and honeybee the orthologue of *Her* is regulated by both a paired Su(H) site and a A-box proneural site at the 5' region of the gene, and a GY box and K box at the 3' of the gene. However, in the lineage leading to *Drosophila* these regulatory sites appear to have been lost with no paired Su(H) site, and a GY box being located 833 nt downstream of the putative stop site. This is consistent with the data presented here that Her is not Notch responsive in *Drosophila* and implies a loss of functional constraint around this genomic region following the divergence of Diptera from Hymenoptera 300 mya (Gaunt and Miles 2002).

Supplemental Table 1 –Names of proteins and species used in this study.

Supplemental Table 2 -Sequences of oligonucleotide primers.