Hairless as a novel component of the Notch signaling pathway

Steven W Miller[1], Artem Movsesyan[1], Sui Zhang[1], Rosa Fernández[2,3], James W Posakony[1]

[1]Division of Biological Sciences, Section of Cell and Developmental Biology, University of California San Diego, La Jolla, United States

[2]Bioinformatics and Genomics Unit, Center for Genomic Regulation, Barcelona, Spain

[3]Current address: Department of Life Sciences, Barcelona Supercomputing Center, Barcelona, Spain

For correspondence: jposakony@ucsd.edu

Abstract

Suppressor of Hairless [Su(H)], the transcription factor at the end of the Notch pathway in *Drosophila*, utilizes the Hairless protein to recruit two co-repressors, Groucho (Gro) and C-terminal Binding Protein (CtBP), indirectly. Hairless is present only in the Pancrustacea, raising the question of how Su(H) in other protostomes gains repressive function. We show that Su(H) from a wide array of arthropods, molluscs, and annelids includes motifs that directly bind Gro and CtBP; thus, direct co-repressor recruitment is ancestral in the protostomes. How did Hairless come to replace this ancestral paradigm? Our discovery of a protein (S-CAP) in Myriapods and Chelicerates that contains a motif similar to the Su(H)-binding domain in Hairless has revealed a likely evolutionary connection between Hairless and Metastasis-associated (MTA) protein, a component of the NuRD complex. Sequence comparison and widely conserved microsynteny suggest that S-CAP and Hairless arose from a tandem duplication of an ancestral MTA gene.

Introduction

A very common paradigm in the regulation of animal development is that DNA-binding transcriptional repressors bear defined amino acid sequence motifs that permit them to recruit, by direct interaction, one or more common co-repressor proteins that are responsible for conferring repressive activity. Two such universal co-repressors are Groucho (Gro) and C-terminal Binding Protein (CtBP).

The ancient and highly conserved transcription factor Suppressor of Hairless [Su(H)] functions at the terminus of the widely utilized Notch cell-cell signaling pathway. Su(H) is converted into an activator by signaling through the Notch receptor, but in the absence of signaling it functions as a repressor. Earlier studies have revealed that in many settings in *Drosophila*, Su(H)'s repressive activity depends on binding to the Hairless protein (Figure 1). Hairless includes separate Gro- and CtBP-binding motifs, which permit it to function as an adaptor to bring these two corepressors to Su(H) (Figure 1B) (Barolo et al. 2002). Thus, the Su(H)/H partnership in the fly represents a notable exception to the rule of direct co-repressor recruitment.

As genome and transcriptome sequences have become available for more and more insects and other arthropods, we have searched for possible Hairless orthologs in a wide variety of species, in an attempt to determine the protein's phylogenetic distribution. We have found that Hairless is confined to the Pancrustacea (or Tetraconata), a clade of arthropods that includes the Crustacea and Hexapoda (Misof et al. 2014; Kjer et al. 2016). While this indicates that Hairless arose over 500 Mya, it also raises the question of how Su(H) in other protostomes gains repressive activity.

Here we present evidence that direct co-repressor recruitment by Su(H) is likely to be ancestral in the protostomes. We show that Su(H) in a broad range of protostomes,

including arthropods, molluscs, and annelids, bears both a short linear motif that mediates binding of CtBP and a novel motif for direct recruitment of Gro. Thus, the evolutionary appearance of Hairless has permitted the replacement of an ancient and predominant regulatory mechanism (direct co-repressor recruitment) with a novel one (indirect recruitment).

What can we learn about the evolutionary history of Hairless? While Hairless itself is found only in the Pancrustacea, we show that the genomes of Myriapods and Chelicerates encode a protein with clear sequence and functional similarities to Hairless. These proteins include a motif that strongly resembles the Su(H)-binding domain of Hairless, and we demonstrate that this motif from the house spider *Parasteatoda tepidariorum* does indeed bind Su(H). In addition, these Myriapod and Chelicerate proteins also include one or more canonical motifs for recruitment of CtBP. Accordingly, we designate these factors as "Su(H)-Co-repressor Adaptor Proteins" (S-CAPs).

Finally, further sequence analyses, along with the discovery of conserved microsynteny, have provided substantial evidence that Hairless and the S-CAPs are likely to be orthologous and that they arose from a duplication of the gene encoding Metastasis-associated (MTA) protein, a component of the nucleosome remodeling and deacetylase (NuRD) complex.

An intriguing question in evolutionary biology concerns the path by which a particular clade has escaped a strongly selected character that has been conserved for hundreds of millions of years. We believe that our study has yielded valuable insight into both the emergence of an evolutionary novelty and its replacement of an ancestral paradigm.

Results

Hairless is present only in the Pancrustacea

We have conducted extensive BLAST searches of genome and transcriptome sequence data for a wide variety of metazoa in an attempt to define the phylogenetic distribution of Hairless. We find that Hairless as originally described (Bang and Posakony 1992; Maier et al. 1992; Maier et al. 2008) is confined to the Pancrustacea (or Tetraconata), and occurs widely within this clade, including the Hexapoda, Vericrustacea, and Oligostraca (Figure 2A). By contrast, no evidence for a true *Hairless* gene has been detected in either Myriapods or Chelicerates, even in cases where substantially complete genome sequence assemblies are available.

The enormous variation in the size of the Hairless protein in various Pancrustacean clades is worthy of note (Figure 1A). The known extremes are represented by the Diplostracan (shrimp) *Eulimnadia texana* (343 aa) (Baldwin-Brown et al. 2018) and the Dipteran (fly) *Protophormia terraenovae* (1614 aa) (Hase et al. 2017), a 4.7-fold difference. There is a broad tendency for the size of the protein to be relatively stable within an order (Supplementary file 1). Thus, as noted previously (Maier et al. 2008), the Hymenoptera generally have a small Hairless (of the order of 400 aa; see Figure 1A), while the Diptera typically have a much larger version (of the order of 1000 aa or more). Notable exceptions to this pattern of uniformity are aphids, where Hairless is typically ~900 aa compared to ~400 aa in other Hemiptera, and chalcid wasps, where the protein is over 500 aa instead of the Hymenoptera-typical ~400 aa noted above (Supplementary file 1). Smaller Hairless proteins retain all five conserved motifs/domains characteristic of this factor (Maier et al. 2008), while the regions that flank and lie between these sequences are reduced in size (Figure 1A; Supplementary file 2).

A known CtBP-binding motif is present in the non-conserved N-terminal region of Su(H) in a wide variety of protostomes

The apparent confinement of the Hairless co-repressor adaptor protein to the Pancrustacea raises the question of the mechanism(s) by which Su(H) in other protostomes might recruit co-repressor proteins to mediate its repressor function. Of course, other protostomes need not utilize the Gro and CtBP co-repressors for this purpose; different co-repressors might substitute. Nevertheless, we first sought to identify known binding motifs for Gro and CtBP in Su(H) from arthropods lacking Hairless. As shown in Table 1, we found a canonical CtBP recruitment motif of the form PϕDϕS (where ϕ = I, L, M, or V) in predicted Su(H) proteins from a variety of Myriapods and Chelicerates, including the centipede *Strigamia maritima*, the tick *Ixodes scapularis*, the spider *Parasteatoda tepidariorum*, the horseshoe crab *Limulus polyphemus*, and the scorpion *Centruroides sculpturatus*. These motifs are all located in the non-conserved N-terminal region of Su(H) (Supplementary file 3).

Extending this sequence analysis to other protostome phyla led to the finding that a similar PϕDϕS motif occurs in the N-terminal region of Su(H) from a large number of molluscs and annelids, as well as from multiple Nemertea, Brachiopoda, Phoronida, and monogonont rotifers, and also from some flatworms (Table 1). It is notable, by contrast, that we do not find CtBP-binding motifs present in Su(H) from nematodes. Nevertheless, given the broad phylogenetic distribution of the PϕDϕS motif in Su(H) from both Ecdysozoa and Lophotrochozoa, our observations strongly suggest that direct recruitment of CtBP by Su(H) is ancestral in the protostomes.

To verify that the shared PϕDϕS motif in protostome Su(H) proteins can indeed mediate direct recruitment of CtBP, we carried out an *in vitro* pulldown assay using

GST-tagged *Drosophila* CtBP (bound to Glutathione Sepharose beads) and a His-tagged fragment of *Strigamia maritima* Su(H) (Figure 3A). We found that the two proteins do interact directly and robustly, in a manner that is dependent on the integrity of the PVDLS motif in *Strigamia* Su(H).

A novel conserved motif in protostome Su(H) binds the Gro co-repressor

In addition to a PϕDϕS CtBP-binding motif, we have found that Su(H) from a wide variety of protostomes includes a novel motif similar to GSLTPPDKV (Table 1). Where present, this sequence typically lies a short (but variable) distance C-terminal to the PϕDϕS motif, also within the non-conserved N-terminal region of the protein (Supplementary file 3). The GSLTPPDKV motif is particularly prevalent in Su(H) from the Trochozoa, which includes annelids, sipunculans, molluscs, nemerteans, brachiopods, and phoronids (Kocot et al. 2017). Among the Ecdysozoa, it appears consistently in Su(H) from Crustacea and Myriapoda, and in small subsets of both Hexapoda (Ephemeroptera, Odonata, Zygentoma, Archaeognatha, Diplura, and Collembola) and Chelicerata [harvestmen (Opiliones) and Scorpiones]. The motif is absent from Su(H) in all other insect orders, and we have not found it so far in Su(H) from nematodes, flatworms, rotifers, or tardigrades; it is, however, found in the Onychophoran *Euperipatoides kanangrensis* (Table 1). Perhaps surprisingly, the motif is present in Su(H) from the acorn worms *Saccoglossus kowalevskii* and *Ptychodera flava* (Simakov et al. 2015), which are hemichordates (deuterostomes).

Using an *in vitro* pulldown assay, we tested the possibility that the GSLTPPDKV motif mediates binding of the Gro co-repressor (Figure 3B). Indeed, we find that GST-tagged Gro protein interacts strongly with a His-tagged protein bearing this motif at its C-terminus, and that this binding is abolished when the motif is replaced by alanine

residues. We conclude that Su(H) from a broad range of protostomes is capable of directly recruiting both CtBP and Gro (Table 1), and that this capacity is hence very likely to be ancestral in this clade.

Retention of the hybrid state: Species that have both Hairless and the co-repressor-binding motifs in Su(H)

The evolutionary emergence of Hairless as an adaptor protein capable of mediating the indirect recruitment of both Gro and CtBP to Su(H) might be expected to relieve a selective pressure to retain the ancestral Gro- and CtBP-binding motifs in Su(H) itself. And indeed, we find that Su(H) from multiple insect orders comprising the Neoptera lacks both of these sequences (Figure 2B). Strikingly, however, we have observed that Crustacea and a small group of Hexapoda retain both traits (Figure 2B). Thus, multiple representatives of the Branchiopoda, Malacostraca, and Copepoda, along with Ephemeroptera, Odonata, Zygentoma, Archaeognatha, Diplura, and Collembola, have both a canonical Hairless protein (including its Gro- and CtBP-binding motifs) and Gro- and CtBP-binding motifs within Su(H). These clades, then, appear to have retained a "hybrid intermediate" state (Baker et al. 2012) characterized by the presence of both co-repressor recruitment mechanisms.

Myriapods and Chelicerates encode a protein with similarity to Hairless

While canonical Hairless proteins are confined to the Pancrustacea, we have discovered that the genomes of Myriapods and Chelicerates nevertheless encode a protein with intriguing similarities to Hairless. Most notable is the presence of a motif that strongly resembles the "Su(H) binding domain" (SBD) of Hairless, which mediates its high-affinity direct interaction with Su(H) (Figure 1; Figure 4A). We will refer to these proteins as “S-CAPs”; the basis for this designation will be made clear in forthcoming

figures. We note that the occurrence of this protein in the centipede *Strigamia maritima* has also recently been reported by Maier (Maier 2019). In the Pancrustacea, the N-terminal and C-terminal halves of the Hairless SBD are encoded by separate exons (Figure 4B). Strikingly, the related motif in Myriapod and Chelicerate S-CAPs is likewise encoded by separate exons, with exactly the same splice junction as in Hairless (Figure 4B). We believe that this is highly unlikely to be coincidental, and is instead strongly suggestive of an evolutionary relationship between Hairless and S-CAPs.

A recent structural analysis of the Su(H)-Hairless protein complex identified several residues in the Hairless SBD that are involved in binding to the C-terminal domain (CTD) of Su(H) (Yuan et al. 2016) (Figure 4A). These include four hydrophobic amino acids in the main body of the SBD (L235, F237, L245, and L247; these are highlighted in red in Figure 4A). Note that the Myriapod and Chelicerate S-CAP motifs share these same residues. In addition, a tryptophan (W258) C-terminal to the main body of the Hairless SBD also participates in binding to Su(H) (Figure 4A). Myriapod and Chelicerate S-CAPs all include a tryptophan residue at a similar position C-terminal to the main SBD-like domain (Figure 4A). Moreover, this particular W residue in both Hairless and the S-CAPs is followed by a hydrophobic residue, typically V or I. These sequence features, we suggest, is further strong evidence of a common ancestry for the respective segments of Hairless and S-CAPs.

A third structural similarity between Hairless and S-CAPs is the presence in the latter of one or more short linear motifs capable of binding the CtBP co-repressor (Figure 5A). These motifs typically reside in the C-terminal half of the S-CAPs, superficially resembling the C-terminal location of Hairless's CtBP recruitment motif.

A table listing representative examples of Myriapod and Chelicerate S-CAPs is provided as Supplementary file 4, and an annotated FASTA file of their amino acid sequences is included as Supplementary file 5.

Spider S-CAP binds to *Drosophila* Su(H)

Given the clear sequence similarity between the Hairless SBD and the SBD-like motif in Myriapod and Chelicerate S-CAPs, we investigated whether the latter motif is likewise capable of mediating direct binding to Su(H). As noted above, the Hairless SBD interacts specifically with the CTD of Su(H). Since this domain in Su(H) is very highly conserved throughout the Bilateria and Cnidaria, we thought it reasonable to utilize *Drosophila* Su(H) for this binding assay. As shown in Figure 4C, we find that a 200-amino-acid segment of S-CAP from the spider *Parasteatoda tepidariorum* binds directly to *Drosophila* Su(H) *in vitro*. This interaction depends strictly on the integrity of the five residues that in Hairless have been shown to contact the Su(H) CTD (highlighted in red in Figure 4A).

Given the presence of one or more CtBP recruitment motifs in the Myriapod and Chelicerate S-CAP proteins (Figure 5A), along with the ability of their SBD-like domains to bind Su(H) (Figure 4C), we have designated these as "Su(H)-Co-repressor Adaptor Proteins" (S-CAPs).

Chelicerate S-CAP proteins are related to Metastasis-Associated (MTA) proteins

In addition to their similarities to Hairless, the S-CAP proteins of Chelicerates include two regions with strong sequence homology to the Metastasis-associated (MTA) protein family, which is highly conserved among Metazoa. The MTA proteins play an important role in transcriptional regulation via their function as core components of the nucleosome remodeling and deacetylase (NuRD) complex (Allen et al. 2013). The N-

terminal half of MTAs includes four well-defined functional domains: BAH (Bromo-Adjacent Homology), ELM2 (Egl-27 and MTA1 homology), SANT (Swi3, Ada2, N-CoR, and TFIIIB), and GATA-like zinc finger (Millard et al. 2014) (Figure 5B). Of these, the ELM2 and SANT domains are retained at the N-terminal end of Chelicerate S-CAPs (Figure 5B; Figure S1A). This is highly likely to have functional significance, as the ELM2 and SANT domains of MTA proteins work together to recruit and activate the histone deacetylases HDAC1 and HDAC2 (Millard et al. 2013). Further suggesting homology between Chelicerate S-CAPs and MTAs is the observation that their shared ELM2 and SANT domains are each encoded by two exons with exactly the same splice junction (Figure 5C).

It is noteworthy that, despite sharing the SBD-like and CtBP recruitment motifs of Chelicerate S-CAPs, the available Myriapod S-CAP protein sequences lack the N-terminal ELM2 and SANT homologies with MTA proteins (Figure 5B). Consistent with this, the SBD motif in Myriapod S-CAPs lies much closer to the protein's N terminus than the SBD motif in Chelicerate S-CAPs, suggesting that simple loss of the ELM2/SANT-encoding exons might underlie this difference between the two S-CAP clades. Likewise, Hairless proteins are devoid of clear similarities to MTAs.

In addition to their SBD and ELM2/SANT domains, Chelicerate S-CAPs share a third region of homology that lies between the ELM2 and SANT sequences (Figure 5-figure supplement 1A). This region is absent from both Hairless and the Myriapod S-CAPs. Conversely, Myriapod S-CAPs include three segments of sequence similarity in their C-terminal regions that are not found in either Hairless or Chelicerate S-CAPs (Figure 5-figure supplement 1B).

Conserved microsynteny between *MTA* and *S-CAP*/*Hairless* genes

Our analysis of the genomic locations of genes encoding MTA proteins in Arthropoda, Hairless in Pancrustacea, and S-CAPs in Myriapods and Chelicerates has yielded the surprising finding that proximate or near-proximate linkage between *MTA* and *Hairless* genes or between *MTA* and *S-CAP* genes is broadly conserved among arthropods (Figure 6; Supplementary file 1; Supplementary file 4). Thus, in the centipede *Strigamia maritima*, the gene encoding S-CAP lies immediately upstream of that encoding MTA, in the same orientation (Figure 6; Supplementary file 4). A similar linkage relationship between *S-CAP* and *MTA* genes is seen in many arachnids, including the spiders *Nephila clavipes* (Supplementary file 4) and *Parasteatoda tepidariorum* (Figure 6; Supplementary file 4) and the mites *Achipteria coleoptrata* and *Sarcoptes scabiei* (Supplementary file 4). Likely due at least in part to its history of whole-genome duplication (Nossa et al. 2014; Kenny et al. 2016), the horseshoe crab *Limulus polyphemus* (representing the Merostomata/Xiphosura) has three paralogous copies of this same *S-CAP-MTA* linkage pairing (Supplementary file 4). Some exceptions to this pattern do exist. In the genomes of the mites *Metaseiulus occidentalis* (Supplementary file 4) and *Varroa destructor* (Techer et al. 2019), for example, the genes encoding S-CAP and MTA are far separated from each other.

Close, typically adjacent, linkage between *Hairless* and *MTA* genes is likewise widely observed in the genomes of Pancrustacea. Among the Hexapoda, this pattern can be found in many different orders (Supplementary file 1), including Diptera, Lepidoptera, Coleoptera (Figure 6), Hymenoptera (Figure 6), Psocodea, Hemiptera (Figure 6), Thysanoptera, Blattodea, Orthoptera, Odonata, and Collembola. Among the Vericrustacea, adjacent linkage of *Hairless* and *MTA* is seen in the shrimp *Triops cancriformis* (Notostraca) (Supplementary file 1). Nevertheless, exceptions are readily found, even within the same orders as above (Supplementary file 1). Examples include

Drosophila melanogaster, *Ceratitis capitata*, and *Lucilia cuprina* (Diptera; Supplementary file 1), *Bicyclus anynana* (Lepidoptera), *Anoplophora glabripennis*, *Dendroctonus ponderosae*, and *Nicrophorus vespilloides* (Coleoptera), and *Cimex lectularius* (Hemiptera; Supplementary file 1).

Interestingly, in some instances *Hairless*/*MTA* microsynteny is preserved, but the genes' relative orientation is different (Figure 6; Supplementary file 1). Thus, in the aphids — in contrast to other Hemiptera — *MTA* lies downstream of *Hairless*, but in the opposite orientation (Figure 6). In the beetle *Harmonia axyridis* (Coleoptera), *MTA* lies upstream of *Hairless* (Figure 6).

Despite the multiple instances in which it has been lost, we believe that the most parsimonious interpretation of our analysis is that close linkage between *MTA* and *S-CAP*/*Hairless* genes is ancestral in the respective taxa (Myriapods/Chelicerates and Pancrustacea). We leave for the Discussion our proposed interpretation of the evolutionary significance of this adjacency.

Discussion

The evolution of Hairless represents a shift from the ancestral and dominant paradigm of direct co-repressor recruitment by Su(H)

Our analysis of sequences from a broad range of protostomes strongly suggests that direct recruitment of the CtBP and Gro co-repressors by Su(H) is ancestral in this clade. This is consonant with the fact that direct co-repressor recruitment by DNA-binding repressor proteins in general is a dominant paradigm among Metazoa. This evokes the intriguing question of what might have led to the loss of direct recruitment by Su(H) in the Neoptera (see Figure 1B) and its replacement by Hairless-mediated indirect recruitment? Does Hairless provide some advantageous functional capacity? Note that this is not intended to suggest that Hairless must be an evolutionary adaptation *per se* (Lynch 2007); rather, we are asking: What capability might it have conferred that would lead to its retention and the subsequent loss of the recruitment motifs in Su(H)?

One appealing possibility is that Hairless may have permitted Su(H) for the first time to recruit both CtBP and Gro simultaneously to the same target genes. As we have noted, the apparently ancestral PϕDϕS and GSLTPPDKV motifs in protostome Su(H) typically lie quite close to each other in the protein's linear sequence (Supplementary file 3). CtBP (~400 aa) and Gro (~700 aa) are both large proteins that engage in oligomerization as part of their functional mechanism (Song et al. 2004; Bhambhani et al. 2011). It is very unlikely that both could bind stably to DNA-bound Su(H) at the same time. In contrast, the Gro and CtBP recruitment motifs in Hairless are far apart in the linear sequence (Figure 1A) and are separated by a region predicted to be largely disordered (Figure 1-figure supplement 1). We suggest that this might be compatible with simultaneous recruitment of the two co-repressors.

Whatever may have been the selective forces that led to the loss of direct co-repressor recruitment by Su(H) in the Neoptera and its replacement by Hairless-mediated indirect recruitment, Hairless is a notable evolutionary novelty for having permitted the unusual abandonment of an ancestral and highly conserved paradigm. We suggest that this represents a striking example of "developmental system drift" (True and Haag 2001), in which a common output (widespread "default repression" of Notch pathway target genes) is achieved via distinct molecular mechanisms in different species.

A possible evolutionary pathway for the appearance of Hairless

We have described here several findings that we believe have important implications for an attempt to reconstruct the history of Hairless as an evolutionary novelty. First, we observe that Hairless is apparently confined to the Pancrustacea, wherein it is widely distributed among diverse taxa (Figure 2A; Supplementary file 1). Second, we have discovered in the sister groups Myriapoda and Chelicerata a protein (S-CAP) with clear sequence homology to the Su(H) binding domain (SBD) of Hairless (Figure 4A). Significantly, in both Hairless and the S-CAPs these motifs are encoded by contributions from two exons, with the associated splice junction in precisely the same location (Figure 4B; Supplementary file 4). Third, we find that S-CAPs in the Chelicerata include in their N-terminal region strong homology to the ELM2 and SANT domains of MTAs, which themselves are highly conserved among Metazoa, and therefore would have been present in the arthropod common ancestor (Figure 5B,C). Finally, our analysis indicates that close, usually adjacent, linkage of *Hairless* and *MTA* genes (in the Pancrustacea) and between *S-CAP* and *MTA* genes (in the Myriapoda and

Chelicerata) is widespread (Figure 6; Supplementary file 1; Supplementary file 4), and hence very likely to be ancestral, in these taxa.

While any attempt to infer the sequence of evolutionary events that led to the appearance of Hairless is necessarily speculative, we believe that the above findings offer substantial support for the following hypothetical pathway. We propose that in a deep arthropod ancestor a tandem duplication of the *MTA* gene occurred. One copy retained the strong sequence conservation (and presumably ancestral function) of metazoan *MTA* genes, while the second copy diverged very substantially, eventually encoding a protein that had lost all but the ELM2 and SANT domains of the MTA ancestor. The extensive reconfiguration of this paralog also included the eventual acquisition of the SBD motif and the addition of one or more CtBP recruitment motifs (see Figure 7 for some possible sources of these components). In the Myriapod lineage, even the ELM2 and SANT domains were eventually lost. In the Pancrustacea, we suggest that this same divergent *MTA* paralog evolved to become *Hairless*. Beyond the alterations described for the Myriapoda, this would have involved the acquisition of sequences encoding additional now-conserved domains and motifs, including the Gro recruitment motif (Supplementary file 2). This radical evolutionary transformation resulted in a protein with little or no remaining homology to its MTA ancestor, and with an entirely novel regulatory function (Holland et al. 2017).

In this context, it is of interest that the *Drosophila* Mi-2/Nurd complex — which includes the MTA protein — has recently been shown to engage in direct repression of multiple Notch pathway target genes, independent of both Su(H) and Hairless (Zacharioudaki et al. 2019). Whether this activity preceded the emergence of Hairless is unknown, but the possibility that it is in some way connected to Hairless's evolutionary history is indeed intriguing.

Materials and methods

Sequence searches, analysis, and annotation

Genome and transcriptome sequences encoding Hairless, Suppressor of Hairless, S-CAP, and MTA proteins from a wide variety of species were recovered via BLAST searches, using either the online version at the NCBI website (Boratyn et al. 2013) or the version implemented by the BlastStation-Local64 desktop application (TM Software, Inc.). Sequences were analyzed and annotated using the GenePalette (Rebeiz and Posakony 2004; Smith et al. 2017) and DNA Strider (Marck 1988; Douglas 1995) desktop software tools. Analysis of predicted disordered regions in Hairless was conducted using DISOPRED3 on the PSIPRED server (Buchan et al. 2013; Jones and Cozzetto 2015).

Generation of constructs for GST pulldown experiments
Strigamia maritima Su(H) protein constructs to test CtBP binding

A codon-optimized fragment corresponding to exons 2 and 3 from *S. maritima* Su(H) mRNA was synthesized by Genewiz, Inc., and cloned into pRSET-C using Acc65I and BamHI restriction sites. The CtBP-motif mutant was subsequently generated by overlap extension PCR (Ho) using the primers HISsmarSUH-f (CGCTGGATCCGCGGCCAGTATGAC), HISsmarSUH-r (CCATGGTACCAGTTATGCGTGGTG), HISsmarSUHctbpm-f (AACCACgCCGcaGcTGcGgCTAACAGCCATCGCGGTGAAGGCGGCCAC), HISsmarSUHctbpm-r (GCTGTTAGcCgCAgCtgCGGcGTGGTTGTCGGCGAAGTGAGGGGTCAG). After sequence confirmation, this fragment was also cloned into pRSET-C using the same enzymes. Binding of these constructs to *Drosophila melanogaster* CtBP was assayed using GST alone and a GST-CtBP fusion protein (Nibu et al. 1998).

Constructs to test potential Gro-binding motif in *Strigamia maritima* Su(H)

A truncated version of HLHmβ (HLHmβ-WRPWtrunc) was amplified from a pRSET-HLHmβ-WT construct using the primers HISmbeta-f (cgatggatccgaATGGTTCTGGAAATGGAGATGTCCAAG) and HISmbetatrunc-r (ccatggtaccagTCACATGGGGCCagaggtggagctggcctcgctgggcgc); a version of HLHmβ with the WRPW motif replaced with the amino acids GSLTPPDKV (HLHmβ+Smar-motifWT) was amplified from the WT construct with HISmbeta-f and mbetaSmarSuH-r (ccatggtaccagTCACACTTTATCAGGTGGAGTGAGAGAACCCATGGGGCCagaggtggagctggcc); and a version of HLHmβ with the WRPW motif replace with a stretch of 9 alanine residues (HLHmβ+Smar-motifMUT) was amplified using HISmbeta-f and mbetaSmarSuHmut-r (ccatggtaccagTCAggctgccgctgcggctgccgctgctgcCATGGGGCCagaggtggagctggcc). Each construct was then subsequently cloned into pRSET-C using the restriction enzymes BamHI and Acc65I and sequence verified. Binding of these constructs to *Drosophila melanogaster* Gro was assayed using GST alone and a GST-Gro fusion protein. The latter construct was made by cloning full-length Gro coding sequence into the pGEX expression vector.

S-CAP/Hairless constructs for Su(H) interaction analysis

Codon-optimized fragments from *Drosophila melanogaster* Hairless (residues 192-389), and *Parasteatoda tepidariorum* cS-CAP (residues 233-432) as well as 5-alanine mutant substitutions (Dmel: G**GRLQFFKDGKFILEL**ARSKDGDKSG**W** -> G**GR<u>A</u>Q<u>A</u>FKDGKFI<u>A</u>E<u>A</u>**ARSKDGDKSG**<u>A</u>**; Ptep: V**GSLKFFLGGRLVLKL**NAQQDGGSGNKCQ**W** ->

V**GS<u>A</u>K<u>A</u>FLGGRLV<u>A</u>K<u>A</u>**NAQQDGGSGNKCQ**<u>A</u>**) were synthesized by Genewiz, Inc. Inserts were subsequently cloned into pRSET-C using the restriction enzymes BamHI and Acc65I. Binding of these constructs to *Drosophila melanogaster* Su(H) was assayed using GST alone and a GST-Su(H) fusion protein (Bailey and Posakony 1995).

GST pulldowns using each of these constructs were performed as previously described (Fontana and Posakony 2009).

Synthesized, codon-optimized sequences

>Smar Su(H)ex2-3 WT (116 aa)
CGCTGGATCCGCGGCCAGTATGACTACCCGCCGCCGTTAGCCAGCACATACAGCCGCGAGGCCG
ACCTGTGGAACGTGAACCTGGCCACCTACAGCAGCGCACCGACCACATGCACCGGTGCAACCCC
GGCACCTAGCGTTACCGGTTTCTACGCCCAGGCCACCGGCAGCAACAGCGTTAGCCCGAGTAGC
GTGAGCCTGACCACCCTGACCCCTCACTTCGCCGACAACCACCCGGTGGACCTGAGCAACAGCC
ATCGCGGTGAAGGCGGCCACCTGGATCTGGTGCGCTTCCAGAGCGACCGCGTGGATGCCTACAA
GCACGCCAACGGCCTGAGCGTGCATATCCCGGACCACCACGCATAACTGGTACCATGG

>Smar Su(H)ex2-3 mut
CGCTGGATCCGCGGCCAGTATGACTACCCGCCGCCGTTAGCCAGCACATACAGCCGCGAGGCCG
ACCTGTGGAACGTGAACCTGGCCACCTACAGCAGCGCACCGACCACATGCACCGGTGCAACCCC
GGCACCTAGCGTTACCGGTTTCTACGCCCAGGCCACCGGCAGCAACAGCGTTAGCCCGAGTAGC
GTGAGCCTGACCACCCTGACCCCTCACTTCGCCGACAACCACgCCGcaGcTGcGgCTAACAGCC
ATCGCGGTGAAGGCGGCCACCTGGATCTGGTGCGCTTCCAGAGCGACCGCGTGGATGCCTACAA
GCACGCCAACGGCCTGAGCGTGCATATCCCGGACCACCACGCATAACTGGTACCATGG

>Dmel Hairless192-389 WT
CGATGGATCCGAGCAGTGGTTGCAGCAGCAGCTGGCACTGCCAAAATTGGTAAAGGCAGCAACA
GCGGTGGCAGTTTTGATATGGGCCGCACACCGATCAGCACCCACGGCAACAATAGTTGGGGTGG
CTATGGCGGCCGTTTACAGTTCTTTAAAGATGGCAAGTTTATTTTAGAACTGGCCCGCAGCAAA
GATGGCGATAAAAGCGGCTGGGTGAGTGTGACCCGCAAAACCTTTCGCCCGCCGAGTGCAGCAA
CCAGCGCAACCGTGACCCCTACCAGTGCCGTGACCACCGCCTACCCGAAGAATGAAAACAGCAC
CTCTTTAAGCTTCAGCGACGACAATAGCAGCATTCAGAGCAGCCCGTGGCAGCGTGATCAGCCG
TGGAAACAGAGTCGTCCGCGCCGTGGCATCAGCAAAGAACTGTCTTTATTTTTCCACCGCCCGC
GCAATAGTACACTGGGTCGTGCAGCCTTACGTACCGCAGCCCGCAAACGTCGTCGTCCGCATGA
ACCGCTGACCACCAGCGAAGATCAGCAGCCGATCTTTGCCACCGCAATCAAAGCCGAGAACGGT
GATGATACTTTAAAAGCCGAAGCAGCCGAATAACTGGTACCATGG

>Dmel Hairless192-389 5Amut
CGATGGATCCGAGCCGTTGTGGCAGCAGCAGCTGGCACTGCCAAAATCGGCAAAGGCAGCAATA
GCGGTGGTAGCTTTGACATGGGCCGCACCCCGATTAGCACCCATGGCAACAACAGCTGGGGTGG

TTATGGTGGTCGTGCCCAAGCTTTTAAAGACGGCAAGTTCATCGCCGAAGCCGCACGCAGCAAA
GATGGCGACAAAAGCGGTGCCGTGAGCGTGACCCGCAAAACCTTTCGTCCGCCGAGTGCAGCAA
CCAGCGCAACCGTTACCCCGACCAGCGCAGTTACCACCGCCTACCCGAAAAACGAAAACAGCAC
CTCTTTAAGCTTTAGCGACGACAACAGCAGCATTCAGAGCAGCCCGTGGCAGCGCGATCAGCCG
TGGAAACAGAGCCGTCCTCGTCGCGGCATCAGCAAAGAGCTGTCTTTATTCTTTCATCGCCCGC
GCAATAGCACTTTAGGTCGTGCAGCACTGCGCACAGCAGCACGTAAACGTCGTCGCCCGCATGA
ACCGCTGACCACCAGCGAAGACCAGCAGCCGATTTTTGCCACCGCAATCAAAGCCGAGAACGGC
GATGATACTTTAAAAGCAGAAGCAGCCGAATAACTGGTACCATGG

>Ptep S-CAP233-432 WT
CGATGGATCCGAACCGTGAATACCGAAGATCCGCCGAAGGATAGCATCAACTTTCTGGACCACA
GCCGCGTGACCGATCCGTGTAGTGCCGCAAGCGAAACCAGCCTGCCGCAGGATGTGCCGGCAAC
AAGCACCGTGGGCAGCCTGAAATTTTTTCTGGGCGGTCGCCTGGTGCTGAAATTAAACGCCCAG
CAGGATGGCGGCAGCGGCAATAAATGCCAGTGGGTGCAGAGCAACGATCTGCCGAAACATAGCA
ACCATAACAAAAAAGATAAACATAAGAAAAAATTTGCACCGTATAGCTATAGCAGCAGCGGCAC
TCAGAAACCGCTGAAGAAAGGCGACGATACCAGTGCCGTGCCGGACTGTGATCCGAGCGGCATC
AAAAAGCCGCGCCTGAAAGAGTACGAGACCAGCGAGAATAGCGCCCTGGGTCTGCTGCTGTGCA
GCAGCAGTTGGACCCCGCCGGTTGCAGATGGTCAGGAGAGCATTGACGTGGACGATACCAGCAG
CAAAACCAGCGAGGGCTATATTAGCCCGATCCTGAGCAACAATAGCCGCACCAGCAAAATCGAC
ACCATCAAGCACGATTTTGCCAGCAACCCGAACACCTAACTGGTACCATGG

>Ptep S-CAP233-432 5Amut
CGATGGATCCGAACCGTGAACACCGAAGACCCGCCGAAAGATAGCATCAACTTTTTAGACCATA
GCCGCGTGACAGACCCGTGCAGTGCCGCAAGTGAAACCTCTTTACCGCAAGATGTGCCGGCAAC
CAGCACCGTGGGTAGCGCCAAAGCCTTTCTGGGCGGTCGTCTGGTGGCCAAAGCCAATGCCCAG
CAAGATGGTGGTAGTGGTAACAAATGCCAAGCTGTGCAGAGCAACGATCTGCCGAAACACAGCA
ATCACAATAAGAAAGACAAACACAAGAAAAAATTTGCCCCGTATAGCTATAGCAGCAGCGGCAC
CCAGAAACCGCTGAAAAAAGGCGATGACACCAGCGCAGTGCCGGATTGCGATCCGAGCGGCATT
AAGAAACCGCGTTTAAAGGAGTACGAGACCAGCGAAAACAGTGCTTTAGGTTTACTGCTGTGCA
GCAGCAGTTGGACACCGCCGGTGGCCGATGGTCAAGAAAGTATCGATGTGGACGACACCAGCAG
CAAAACCAGCGAAGGCTACATCAGCCCGATTCTGAGCAACAATAGCCGCACCAGCAAAATTGAT
ACCATTAAACATGATTTTGCAAGCAATCCGAATACCTAACTGGTACCATGG

Acknowledgements

We thank Scott Rifkin for very useful discussion and input during the preparation of the manuscript. We are grateful to the following artists for making available the illustrations shown in Figure 2A: crab, by Firkin (https://openclipart.org/detail/270221/crab-silhouette); monarch butterfly, by carolemagnet (https://openclipart.org/detail/263384/monarch-butterfly); centipede, by Firkin (https://openclipart.org/detail/261126/centipede-3); spider, by liftarn (https://openclipart.org/detail/179190/spider); horseshoe crab, by Gosc (https://openclipart.org/detail/174556/horseshoe-crab); tick, by Juhele (https://openclipart.org/detail/279073/simple-tick-ixodes-ricinus-silhouette). This work was supported by NIH Grants R01GM046993 and R01GM120377 (to JWP).

References

Albertin CB, Simakov O, Mitros T, Wang ZY, Pungor JR, Edsinger-Gonzales E, Brenner S, Ragsdale CW, Rokhsar DS. 2015. The octopus genome and the evolution of cephalopod neural and morphological novelties. *Nature* **524**: 220-224.

Allen HF, Wade PA, Kutateladze TG. 2013. The NuRD architecture. *Cell Mol Life Sci* **70**: 3513-3524.

Almada AA, Tarrant AM. 2016. Vibrio elicits targeted transcriptional responses from copepod hosts. *FEMS Microbiol Ecol* **92**: fiw072.

Ando T, Matsuda T, Goto K, Hara K, Ito A, Hirata J, Yatomi J, Kajitani R, Okuno M, Yamaguchi K, Kobayashi M, Takano T et al. 2018. Repeated inversions within a pannier intron drive diversification of intraspecific colour patterns of ladybird beetles. *Nat Commun* **9**: 3843.

Andrade SC, Novo M, Kawauchi GY, Worsaae K, Pleijel F, Giribet G, Rouse GW. 2015. Articulating "Archiannelids": Phylogenomics and Annelid Relationships, with Emphasis on Meiofaunal Taxa. *Mol Biol Evol* **32**: 2860-2875.

Anstead CA, Korhonen PK, Young ND, Hall RS, Jex AR, Murali SC, Hughes DS, Lee SF, Perry T, Stroehlein AJ, Ansell BR, Breugelmans B et al. 2015. Lucilia cuprina genome unlocks parasitic fly biology to underpin future interventions. *Nat Commun* **6**: 7344.

Babb PL, Lahens NF, Correa-Garhwal SM, Nicholson DN, Kim EJ, Hogenesch JB, Kuntner M, Higgins L, Hayashi CY, Agnarsson I, Voight BF. 2017. The Nephila clavipes genome highlights the diversity of spider silk genes and their complex expression. *Nat Genet* **49**: 895-903.

Bailey AM, Posakony JW. 1995. Suppressor of Hairless directly activates transcription of *Enhancer of split Complex* genes in response to Notch receptor activity. *Genes Dev* **9**: 2609-2622.

Baker CR, Booth LN, Sorrells TR, Johnson AD. 2012. Protein modularity, cooperative binding, and hybrid regulatory states underlie transcriptional network diversification. *Cell* **151**: 80-95.

Baldwin-Brown JG, Weeks SC, Long AD. 2018. A New Standard for Crustacean Genomes: The Highly Contiguous, Annotated Genome Assembly of the Clam Shrimp Eulimnadia texana Reveals HOX Gene Order and Identifies the Sex Chromosome. *Genome Biol Evol* **10**: 143-156.

Bang AG, Posakony JW. 1992. The Drosophila gene Hairless encodes a novel basic protein that controls alternative cell fates in adult sensory organ development. *Genes Dev* **6**: 1752-1769.

Barolo S, Posakony JW. 2002. Three habits of highly effective signaling pathways: principles of transcriptional control by developmental cell signaling. *Genes Dev* **16**: 1167-1181.

Barolo S, Stone T, Bang AG, Posakony JW. 2002. Default repression and Notch signaling: Hairless acts as an adaptor to recruit the corepressors Groucho and dCtBP to Suppressor of Hairless. *Genes Dev* **16**: 1964-1976.

Barreto FS, Pereira RJ, Burton RS. 2015. Hybrid dysfunction and physiological compensation in gene expression. *Mol Biol Evol* **32**: 613-622.

Barreto FS, Watson ET, Lima TG, Willett CS, Edmands S, Li W, Burton RS. 2018. Genomic signatures of mitonuclear coevolution across populations of Tigriopus californicus. *Nat Ecol Evol* **2**: 1250-1257.

Bast J, Schaefer I, Schwander T, Maraun M, Scheu S, Kraaijeveld K. 2016. No Accumulation of Transposable Elements in Asexual Arthropods. *Mol Biol Evol* **33**: 697-706.

Battelle BA, Ryan JF, Kempler KE, Saraf SR, Marten CE, Warren WC, Minx PJ, Montague MJ, Green PJ, Schmidt SA, Fulton L, Patel NH et al. 2016. Opsin Repertoire and Expression Patterns in Horseshoe Crabs: Evidence from the Genome of Limulus polyphemus (Arthropoda: Chelicerata). *Genome Biol Evol* **8**: 1571-1589.

Bhambhani C, Chang JL, Akey DL, Cadigan KM. 2011. The oligomeric state of CtBP determines its role as a transcriptional co-activator and co-repressor of Wingless targets. *EMBO J* **30**: 2031-2043.

Blankers T, Oh KP, Bombarely A, Shaw KL. 2018. The Genomic Architecture of a Rapid Island Radiation: Recombination Rate Variation, Chromosome Structure, and Genome Assembly of the Hawaiian Cricket Laupala. *Genetics* **209**: 1329-1344.

Boratyn GM, Camacho C, Cooper PS, Coulouris G, Fong A, Ma N, Madden TL, Matten WT, McGinnis SD, Merezhuk Y, Raytselis Y, Sayers EW et al. 2013. BLAST: a more efficient report with usability improvements. *Nucleic Acids Res* **41**: W29-33.

Brandt A, Schaefer I, Glanz J, Schwander T, Maraun M, Scheu S, Bast J. 2017. Effective purifying selection in ancient asexual oribatid mites. *Nat Commun* **8**: 873.

Brauchle M, Bilican A, Eyer C, Bailly X, Martínez P, Ladurner P, Bruggmann R, Sprecher SG. 2018. Xenacoelomorpha Survey Reveals That All 11 Animal Homeobox Gene Classes Were Present in the First Bilaterians. *Genome Biol Evol* **10**: 2205-2217.

Buchan DW, Minneci F, Nugent TC, Bryson K, Jones DT. 2013. Scalable web services for the PSIPRED Protein Analysis Workbench. *Nucleic Acids Res* **41**: W349-57.

Bunnefeld L, Hearn J, Stone GN, Lohse K. 2018. Whole-genome data reveal the complex history of a diverse ecological community. *Proc Natl Acad Sci U S A* **115**: E6507-E6515.

Cao Z, Yu Y, Wu Y, Hao P, Di Z, He Y, Chen Z, Yang W, Shen Z, He X, Sheng J, Xu X et al. 2013. The genome of Mesobuthus martensii reveals a unique adaptation model of arthropods. *Nat Commun* **4**: 2602.

Carmona-Antoñanzas G, Carmichael SN, Heumann J, Taggart JB, Gharbi K, Bron JE, Bekaert M, Sturm A. 2015. A Survey of the ATP-Binding Cassette (ABC) Gene Superfamily in the Salmon Louse (Lepeophtheirus salmonis). *PLoS One* **10**: e0137394.

Chen W, Hasegawa DK, Kaur N, Kliot A, Pinheiro PV, Luan J, Stensmyr MC, Zheng Y, Liu W, Sun H, Xu Y, Luo Y et al. 2016. The draft genome of whitefly Bemisia tabaci MEAM1, a global crop pest, provides novel insights into virus transmission, host adaptation, and insecticide resistance. *BMC Biol* **14**: 110.

Cheng T, Wu J, Wu Y, Chilukuri RV, Huang L, Yamamoto K, Feng L, Li W, Chen Z, Guo H, Liu J, Li S et al. 2017. Genomic adaptation to polyphagy and insecticides in a major East Asian noctuid pest. *Nat Ecol Evol* **1**: 1747-1756.

Chipman AD, Ferrier DE, Brena C, Qu J, Hughes DS, Schröder R, Torres-Oliva M, Znassi N, Jiang H, Almeida FC, Alonso CR, Apostolou Z et al. 2014. The first myriapod genome sequence reveals conservative arthropod gene content and genome organisation in the centipede Strigamia maritima. *PLoS Biol* **12**: e1002005.

Christie AE, Chi M, Lameyer TJ, Pascual MG, Shea DN, Stanhope ME, Schulz DJ, Dickinson PS. 2015. Neuropeptidergic Signaling in the American Lobster Homarus americanus: New Insights from High-Throughput Nucleotide Sequencing. *PLoS One* **10**: e0145964.
Clarke TH, Garb JE, Hayashi CY, Arensburger P, Ayoub NA. 2015. Spider Transcriptomes Identify Ancient Large-Scale Gene Duplication Event Potentially Important in Silk Gland Evolution. *Genome Biol Evol* **7**: 1856-1870.

Colbourne JK, Pfrender ME, Gilbert D, Thomas WK, Tucker A, Oakley TH, Tokishita S, Aerts A, Arnold GJ, Basu MK, Bauer DJ, Caceres CE et al. 2011. The ecoresponsive genome of *Daphnia pulex*. *Science* **331**: 555-561.

Cong Q, Shen J, Li W, Borek D, Otwinowski Z, Grishin NV. 2017. The first complete genomes of Metalmarks and the classification of butterfly families. *Genomics* **109**: 485-493.

Dong X, Chaisiri K, Xia D, Armstrong SD, Fang Y, Donnelly MJ, Kadowaki T, McGarry JW, Darby AC, Makepeace BL. 2018. Genomes of trombidid mites reveal novel predicted allergens and laterally transferred genes associated with secondary metabolism. *Gigascience* **7**: 1-33.

Douglas SE. 1995. DNA Strider. An inexpensive sequence analysis package for the Macintosh. *Mol Biotechnol* **3**: 37-45.

Egekwu N, Sonenshine DE, Garman H, Barshis DJ, Cox N, Bissinger BW, Zhu J, M Roe R. 2016. Comparison of synganglion neuropeptides, neuropeptide receptors and

neurotransmitter receptors and their gene expression in response to feeding in Ixodes scapularis (Ixodidae) vs. Ornithodoros turicata (Argasidae). *Insect Mol Biol* **25**: 72-92.

Eyun SI, Soh HY, Posavi M, Munro JB, Hughes DST, Murali SC, Qu J, Dugan S, Lee SL, Chao H, Dinh H, Han Y et al. 2017. Evolutionary History of Chemosensory-Related Gene Families across the Arthropoda. *Mol Biol Evol* **34**: 1838-1862.

Faddeeva-Vakhrusheva A, Derks MF, Anvar SY, Agamennone V, Suring W, Smit S, van Straalen NM, Roelofs D. 2016. Gene Family Evolution Reflects Adaptation to Soil Environmental Stressors in the Genome of the Collembolan Orchesella cincta. *Genome Biol Evol* **8**: 2106-2117.

Fernández R, Edgecombe GD, Giribet G. 2016. Exploring Phylogenetic Relationships within Myriapoda and the Effects of Matrix Composition and Occupancy on Phylogenomic Reconstruction. *Syst Biol* **65**: 871-889.

Fontana JR, Posakony JW. 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.

Francois CM, Duret L, Simon L, Mermillod-Blondin F, Malard F, Konecny-Dupré L, Planel R, Penel S, Douady CJ, Lefébure T. 2016. No Evidence That Nitrogen Limitation Influences the Elemental Composition of Isopod Transcriptomes and Proteomes. *Mol Biol Evol* **33**: 2605-2620.

Fu X, Li J, Tian Y, Quan W, Zhang S, Liu Q, Liang F, Zhu X, Zhang L, Wang D, Hu J. 2017. Long-read sequence assembly of the firefly Pyrocoelia pectoralis genome. *Gigascience* **6**: 1-7.

Fu Y, Yang Y, Zhang H, Farley G, Wang J, Quarles KA, Weng Z, Zamore PD. 2018. The genome of the Hi5 germ cell line from Trichoplusia ni, an agricultural pest and novel model for small RNA biology. *Elife* **7**: e31628.

Fuzita FJ, Pinkse MW, Patane JS, Juliano MA, Verhaert PD, Lopes AR. 2015. Biochemical, transcriptomic and proteomic analyses of digestion in the scorpion Tityus serrulatus: insights into function and evolution of digestion in an ancient arthropod. *PLoS One* **10**: e0123841.

Gompert Z, Comeault AA, Farkas TE, Feder JL, Parchman TL, Buerkle CA, Nosil P. 2014. Experimental evidence for ecological selection on genome variation in the wild. *Ecol Lett* **17**: 369-379.

Gulia-Nuss M, Nuss AB, Meyer JM, Sonenshine DE, Roe RM, Waterhouse RM, Sattelle DB, de la Fuente J, Ribeiro JM, Megy K, Thimmapuram J, Miller JR et al. 2016. Genomic insights into the Ixodes scapularis tick vector of Lyme disease. *Nat Commun* **7**: 10507.

Harrison MC, Jongepier E, Robertson HM, Arning N, Bitard-Feildel T, Chao H, Childers CP, Dinh H, Doddapaneni H, Dugan S, Gowin J, Greiner C et al. 2018. Hemimetabolous genomes reveal molecular basis of termite eusociality. *Nat Ecol Evol* **2**: 557-566.

Hase H, Koukai M, Hamanaka Y, Goto SG, Tachibana S, Shiga S. 2017. Transcriptome analysis of the brain under diapause and nondiapause conditions in the blowfly *Protophormia terraenovae*. *Physiol Entomol* **42**: 282-289.

Holland PW, Marlétaz F, Maeso I, Dunwell TL, Paps J. 2017. New genes from old: asymmetric divergence of gene duplicates and the evolution of development. *Philos Trans R Soc Lond B Biol Sci* **372**: 20150480.

Holt RA, Subramanian GM, Halpern A, Sutton GG, Charlab R, Nusskern DR, Wincker P, Clark AG, Ribeiro JM, Wides R, Salzberg SL, Loftus B et al. 2002. The genome sequence of the malaria mosquito *Anopheles gambiae*. *Science* **298**: 129-149.

Hosono K, Sasaki T, Minoshima S, Shimizu N. 2004. Identification and characterization of a novel gene family YPEL in a wide spectrum of eukaryotic species. *Gene* **340**: 31-43.
Hoy MA, Waterhouse RM, Wu K, Estep AS, Ioannidis P, Palmer WJ, Pomerantz AF, Simão FA, Thomas J, Jiggins FM, Murphy TD, Pritham EJ et al. 2016. Genome Sequencing of the Phytoseiid Predatory Mite Metaseiulus occidentalis Reveals Completely Atomized Hox Genes and Superdynamic Intron Evolution. *Genome Biol Evol* **8**: 1762-1775.

Hunt BJ, Özkaya Ö, Davies NJ, Gaten E, Seear P, Kyriacou CP, Tarling G, Rosato E. 2017. The Euphausia superba transcriptome database, SuperbaSE: An online, open resource for researchers. *Ecol Evol* **7**: 6060-6077.

Husnik F, McCutcheon JP. 2016. Repeated replacement of an intrabacterial symbiont in the tripartite nested mealybug symbiosis. *Proc Natl Acad Sci U S A* **113**: E5416-24.
i5K C. 2013. The i5K Initiative: advancing arthropod genomics for knowledge, human health, agriculture, and the environment. *J Hered* **104**: 595-600.

Ikeda KT, Hirose Y, Hiraoka K, Noro E, Fujishima K, Tomita M, Kanai A. 2015. Identification, expression, and molecular evolution of microRNAs in the "living fossil" Triops cancriformis (tadpole shrimp). *RNA* **21**: 230-242.

Ioannidis P, Simao FA, Waterhouse RM, Manni M, Seppey M, Robertson HM, Misof B, Niehuis O, Zdobnov EM. 2017. Genomic Features of the Damselfly Calopteryx splendens Representing a Sister Clade to Most Insect Orders. *Genome Biol Evol* **9**: 415-430.

Jones DT, Cozzetto D. 2015. DISOPRED3: precise disordered region predictions with annotated protein-binding activity. *Bioinformatics* **31**: 857-863.

Kelley JL, Peyton JT, Fiston-Lavier AS, Teets NM, Yee MC, Johnston JS, Bustamante CD, Lee RE, Denlinger DL. 2014. Compact genome of the Antarctic midge is likely an adaptation to an extreme environment. *Nat Commun* **5**: 4611.

Kenny NJ, Chan KW, Nong W, Qu Z, Maeso I, Yip HY, Chan TF, Kwan HS, Holland PW, Chu KH, Hui JH. 2016. Ancestral whole-genome duplication in the marine chelicerate horseshoe crabs. *Heredity (Edinb)* **116**: 190-199.

Kirkness EF, Haas BJ, Sun W, Braig HR, Perotti MA, Clark JM, Lee SH, Robertson HM, Kennedy RC, Elhaik E, Gerlach D, Kriventseva EV et al. 2010. Genome sequences of the human body louse and its primary endosymbiont provide insights into the permanent parasitic lifestyle. *Proc Natl Acad Sci U S A* **107**: 12168-12173.

Kjer KM, Simon C, Yavorskaya M, Beutel RG. 2016. Progress, pitfalls and parallel universes: a history of insect phylogenetics. *J R Soc Interface* **13**: 20160363.

Kocot KM, Struck TH, Merkel J, Waits DS, Todt C, Brannock PM, Weese DA, Cannon JT, Moroz LL, Lieb B, Halanych KM. 2017. Phylogenomics of Lophotrochozoa with Consideration of Systematic Error. *Syst Biol* **66**: 256-282.

Kono N, Nakamura H, Ito Y, Tomita M, Arakawa K. 2016. Evaluation of the impact of RNA preservation methods of spiders for de novo transcriptome assembly. *Mol Ecol Resour* **16**: 662-672.

Lee BY, Kim HS, Hwang DS, Won EJ, Choi BS, Choi IY, Park HG, Rhee JS, Lee JS. 2015. Whole transcriptome analysis of the monogonont rotifer Brachionus koreanus provides molecular resources for developing biomarkers of carbohydrate metabolism. *Comp Biochem Physiol Part D Genomics Proteomics* **14**: 33-41.

Lefébure T, Morvan C, Malard F, François C, Konecny-Dupré L, Guéguen L, Weiss-Gayet M, Seguin-Orlando A, Ermini L, Sarkissian C, Charrier NP, Eme D et al. 2017. Less effective selection leads to larger genomes. *Genome Res* **27**: 1016-1028.

Lemer S, Kawauchi GY, Andrade SC, González VL, Boyle MJ, Giribet G. 2015. Re-evaluating the phylogeny of Sipuncula through transcriptomics. *Mol Phylogenet Evol* **83**: 174-183.

Li S, Zhu S, Jia Q, Yuan D, Ren C, Li K, Liu S, Cui Y, Zhao H, Cao Y, Fang G, Li D et al. 2018. The genomic and functional landscapes of developmental plasticity in the American cockroach. *Nat Commun* **9**: 1008.

Li ZQ, Zhang S, Ma Y, Luo JY, Wang CY, Lv LM, Dong SL, Cui JJ. 2013. First Transcriptome and Digital Gene Expression Analysis in Neuroptera with an Emphasis on Chemoreception Genes in Chrysopa pallens (Rambur). *PLoS One* **8**: e67151.

Lou F, Gao T, Cai S, Han Z. 2018. De novo assembly and annotation of the whole transcriptome of Oratosquilla oratoria. *Mar Genomics* **38**: 17-20.

Luo YJ, Kanda M, Koyanagi R, Hisata K, Akiyama T, Sakamoto H, Sakamoto T, Satoh N. 2018. Nemertean and phoronid genomes reveal lophotrochozoan evolution and the origin of bilaterian heads. *Nat Ecol Evol* **2**: 141-151.

Luo YJ, Takeuchi T, Koyanagi R, Yamada L, Kanda M, Khalturina M, Fujie M, Yamasaki S, Endo K, Satoh N. 2015. The Lingula genome provides insights into brachiopod evolution and the origin of phosphate biomineralization. *Nat Commun* **6**: 8301.

Lynch M. 2007. The frailty of adaptive hypotheses for the origins of organismal complexity. *Proc Natl Acad Sci USA* **104 Suppl 1**: 8597-8604.

Maier D. 2019. The evolution of transcriptional repressors in the Notch signaling pathway: a computational analysis. *Hereditas* **156**: 5.

Maier D, Chen AX, Preiss A, Ketelhut M. 2008. The tiny Hairless protein from Apis mellifera: a potent antagonist of Notch signaling in Drosophila melanogaster. *BMC Evol Biol* **8**: 175.

Maier D, Stumm G, Kuhn K, Preiss A. 1992. Hairless, a Drosophila gene involved in neural development, encodes a novel, serine rich protein. *Mech Dev* **38**: 143-156.

Marck C. 1988. 'DNA Strider': a 'C' program for the fast analysis of DNA and protein sequences on the Apple Macintosh family of computers. *Nucleic Acids Res* **16**: 1829-1836.

Mathers TC, Chen Y, Kaithakottil G, Legeai F, Mugford ST, Baa-Puyoulet P, Bretaudeau A, Clavijo B, Colella S, Collin O, Dalmay T, Derrien T et al. 2017. Rapid transcriptional plasticity of duplicated gene clusters enables a clonally reproducing aphid to colonise diverse plant species. *Genome Biol* **18**: 27.

Millard CJ, Fairall L, Schwabe JW. 2014. Towards an understanding of the structure and function of MTA1. *Cancer Metastasis Rev* **33**: 857-867.

Millard CJ, Watson PJ, Celardo I, Gordiyenko Y, Cowley SM, Robinson CV, Fairall L, Schwabe JW. 2013. Class I HDACs share a common mechanism of regulation by inositol phosphates. *Mol Cell* **51**: 57-67.

Miller JR, Koren S, Dilley KA, Harkins DM, Stockwell TB, Shabman RS, Sutton GG. 2018. A draft genome sequence for the Ixodes scapularis cell line, ISE6. *F1000Res* **7**: 297.

Miller SW, Rebeiz M, Atanasov JE, Posakony JW. 2014. Neural precursor-specific expression of multiple *Drosophila* genes is driven by dual enhancer modules with overlapping function. *Proc Natl Acad Sci USA* **111**: 17194-17199.

Misof B, Liu S, Meusemann K, Peters RS, Donath A, Mayer C, Frandsen PB, Ware J, Flouri T, Beutel RG, Niehuis O, Petersen M et al. 2014. Phylogenomics resolves the timing and pattern of insect evolution. *Science* **346**: 763-767.

Nibu Y, Zhang H, Levine M. 1998. Interaction of short-range repressors with Drosophila CtBP in the embryo. *Science* **280**: 101-104.

Niehuis O, Hartig G, Grath S, Pohl H, Lehmann J, Tafer H, Donath A, Krauss V, Eisenhardt C, Hertel J, Petersen M, Mayer C et al. 2012. Genomic and morphological evidence converge to resolve the enigma of Strepsiptera. *Curr Biol* **22**: 1309-1313.

Nossa CW, Havlak P, Yue JX, Lv J, Vincent KY, Brockmann HJ, Putnam NH. 2014. Joint assembly and genetic mapping of the Atlantic horseshoe crab genome reveals ancient whole genome duplication. *Gigascience* **3**: 9.

Oxley PR, Ji L, Fetter-Pruneda I, McKenzie SK, Li C, Hu H, Zhang G, Kronauer DJ. 2014. The genome of the clonal raider ant Cerapachys biroi. *Curr Biol* **24**: 451-458.

Papanicolaou A, Schetelig MF, Arensburger P, Atkinson PW, Benoit JB, Bourtzis K, Castañera P, Cavanaugh JP, Chao H, Childers C, Curril I, Dinh H et al. 2016. The whole genome sequence of the Mediterranean fruit fly, Ceratitis capitata (Wiedemann), reveals insights into the biology and adaptive evolution of a highly invasive pest species. *Genome Biol* **17**: 192.

Poynton HC, Hasenbein S, Benoit JB, Sepulveda MS, Poelchau MF, Hughes DST, Murali SC, Chen S, Glastad KM, Goodisman MAD, Werren JH, Vineis JH et al. 2018. The Toxicogenome of Hyalella azteca: A Model for Sediment Ecotoxicology and Evolutionary Toxicology. *Environ Sci Technol* **52**: 6009-6022.

Rebeiz M, Posakony JW. 2004. GenePalette: A universal software tool for genome sequence visualization and analysis. *Dev Biol* **271**: 431-438.

Regier JC, Shultz JW, Zwick A, Hussey A, Ball B, Wetzer R, Martin JW, Cunningham CW. 2010. Arthropod relationships revealed by phylogenomic analysis of nuclear protein-coding sequences. *Nature* **463**: 1079-1083.

Rehan SM, Glastad KM, Lawson SP, Hunt BG. 2016. The Genome and Methylome of a Subsocial Small Carpenter Bee, Ceratina calcarata. *Genome Biol Evol* **8**: 1401-1410.
Rider SD, Morgan MS, Arlian LG. 2015. Draft genome of the scabies mite. *Parasit Vectors* **8**: 585.

Rosenfeld JA, Reeves D, Brugler MR, Narechania A, Simon S, Durrett R, Foox J, Shianna K, Schatz MC, Gandara J, Afshinnekoo E, Lam ET et al. 2016. Genome assembly and geospatial phylogenomics of the bed bug Cimex lectularius. *Nat Commun* **7**: 10164.

Roxström-Lindquist K, Faye I. 2001. The Drosophila gene Yippee reveals a novel family of putative zinc binding proteins highly conserved among eukaryotes. *Insect Mol Biol* **10**: 77-86.

Sahoo PK, Kar B, Mohapatra A, Mohanty J. 2013. *De novo* whole transcriptome analysis of the fish louse, *Argulus siamensis*: first molecular insights into characterization of Toll downstream signalling molecules of crustaceans. *Exp Parasitol* **135**: 629-641.

Sanggaard KW, Bechsgaard JS, Fang X, Duan J, Dyrlund TF, Gupta V, Jiang X, Cheng L, Fan D, Feng Y, Han L, Huang Z et al. 2014. Spider genomes provide insight into composition and evolution of venom and silk. *Nat Commun* **5**: 3765.

Sasaki M, Akiyama-Oda Y, Oda H. 2017. Evolutionary origin of type IV classical cadherins in arthropods. *BMC Evol Biol* **17**: 142.

Schwager EE, Sharma PP, Clarke T, Leite DJ, Wierschin T, Pechmann M, Akiyama-Oda Y, Esposito L, Bechsgaard J, Bilde T, Buffry AD, Chao H et al. 2017. The house spider

genome reveals an ancient whole-genome duplication during arachnid evolution. *BMC Biol* **15**: 62.

Shen J, Cong Q, Kinch LN, Borek D, Otwinowski Z, Grishin NV. 2016. Complete genome of Pieris rapae, a resilient alien, a cabbage pest, and a source of anti-cancer proteins. *F1000Res* **5**: 2631.

Sikkink KL, Kobiela ME, Snell-Rood EC. 2017. Genomic adaptation to agricultural environments: cabbage white butterflies (Pieris rapae) as a case study. *BMC Genomics* **18**: 412.

Simakov O, Kawashima T, Marlétaz F, Jenkins J, Koyanagi R, Mitros T, Hisata K, Bredeson J, Shoguchi E, Gyoja F, Yue JX, Chen YC et al. 2015. Hemichordate genomes and deuterostome origins. *Nature* **527**: 459-465.

Simakov O, Marletaz F, Cho SJ, Edsinger-Gonzales E, Havlak P, Hellsten U, Kuo DH, Larsson T, Lv J, Arendt D, Savage R, Osoegawa K et al. 2013. Insights into bilaterian evolution from three spiralian genomes. *Nature* **493**: 526-531.

Simon S, Blanke A, Meusemann K. 2018. Reanalyzing the Palaeoptera problem - The origin of insect flight remains obscure. *Arthropod Struct Dev* **47**: 328-338.

Smith AF, Posakony JW, Rebeiz M. 2017. Automated tools for comparative sequence analysis of genic regions using the GenePalette application. *Dev Biol* **429**: 158-164.

Song H, Hasson P, Paroush Z, Courey AJ. 2004. Groucho oligomerization is required for repression in vivo. *Mol Cell Biol* **24**: 4341-4350.

Techer MA, Rane RV, Grau ML, Roberts JMK, Sullivan ST, Liachko I, Childers AK, Evans JD, Mikheyev AS. 2019. Genomic analyses of sibling honey bee ectoparasitic mite species show divergent strategies of adaptation. *bioRxiv* **doi: 10.1101/512988**.

Tribolium GSC, Richards S, Gibbs RA, Weinstock GM, Brown SJ, Denell R, Beeman RW, Gibbs R, Beeman RW, Brown SJ, Bucher G, Friedrich M et al. 2008. The genome of the model beetle and pest Tribolium castaneum. *Nature* **452**: 949-955.

True JR, Haag ES. 2001. Developmental system drift and flexibility in evolutionary trajectories. *Evol Dev* **3**: 109-119.

von Reumont BM, Campbell LI, Richter S, Hering L, Sykes D, Hetmank J, Jenner RA, Bleidorn C. 2014. A Polychaete's powerful punch: venom gland transcriptomics of Glycera reveals a complex cocktail of toxin homologs. *Genome Biol Evol* **6**: 2406-2423.

Werren JH, Richards S, Desjardins CA, Niehuis O, Gadau J, Colbourne JK, Werren JH, Richards S, Desjardins CA, Niehuis O, Gadau J, Colbourne JK et al. 2010. Functional and evolutionary insights from the genomes of three parasitoid *Nasonia* species. *Science* **327**: 343-348.

Whelan NV, Kocot KM, Santos SR, Halanych KM. 2014. Nemertean toxin genes revealed through transcriptome sequencing. *Genome Biol Evol* **6**: 3314-3325.

Wu C, Jordan MD, Newcomb RD, Gemmell NJ, Bank S, Meusemann K, Dearden PK, Duncan EJ, Grosser S, Rutherford K, Gardner PP, Crowhurst RN et al. 2017a. Analysis of the genome of the New Zealand giant collembolan (Holacanthella duospinosa) sheds light on hexapod evolution. *BMC Genomics* **18**: 795.

Wu C, Twort VG, Crowhurst RN, Newcomb RD, Buckley TR. 2017b. Assembling large genomes: analysis of the stick insect (Clitarchus hookeri) genome reveals a high repeat content and sex-biased genes associated with reproduction. *BMC Genomics* **18**: 884.

Yuan Z, Praxenthaler H, Tabaja N, Torella R, Preiss A, Maier D, Kovall RA. 2016. Structure and Function of the Su(H)-Hairless Repressor Complex, the Major Antagonist of Notch Signaling in Drosophila melanogaster. *PLoS Biol* **14**: e1002509.

Zacharioudaki E, Falo Sanjuan J, Bray S. 2019. Mi-2/NuRD complex protects stem cell progeny from mitogenic Notch signaling. *Elife* **8**: e41637.

Zhang G, Fang X, Guo X, Li L, Luo R, Xu F, Yang P, Zhang L, Wang X, Qi H, Xiong Z, Que H et al. 2012. The oyster genome reveals stress adaptation and complexity of shell formation. *Nature* **490**: 49-54.

Zhao C, Escalante LN, Chen H, Benatti TR, Qu J, Chellapilla S, Waterhouse RM, Wheeler D, Andersson MN, Bao R, Batterton M, Behura SK et al. 2015. A massive expansion of effector genes underlies gall-formation in the wheat pest Mayetiola destructor. *Curr Biol* **25**: 613-620.

Table 1. Co-repressor recruitment motifs in protostome Su(H) proteins

Species	CtBP motif	Gro motif	Source
Ecdyonurus insignis	YPDNH**PVDLS**SPRPH	APMIP**GSLTPPDKMNGE**HPHHG	1
Calopteryx splendens	YTDNH**PVDLS**SPRPP	HHMIP**GSLTPPDKMNGE**HPAMH	2
Atelura formicaria	YPDNH**PVDLS**SPRPQ	PHMIP**GSLTPPDKMNGE**HPHHS	3
Machilis hrabei	YPDNH**PVDLS**SPRPH	PHMLP**GSLTPPDKMNGE**HPHHG	4
Catajapyx aquilonaris	STANN**PVDLS**SPRGS	APMIP**GSLTPPDKVNGE**HHSHH	5
Holacanthella duospinosa	VPNSN**PVDLS**NPSPS	SNFVP**GSLSPPERMNGN**DPSLL	6
Pollicipes pollicipes	YPDNH**PVDLS**SPRPE	GPLIA**GSLTPPDKLGAE**LGLHA	7
Hyalella azteca	SLGHR**PVDLS**QAPSP	AAMLA**GSLTPPDKLNSD**PQQQQ	8
Eurytemora affinis	SETSA**PVDLS**APRPN	YGMLP**GSLTPPDKLNGD**HCSPG	9
Triops cancriformis	HPEAR**PVDLS**SSRLL	YHSSS**LTLTPPDKVNVD**GSNSQ	10
Argulus siamensis	YPENN**PVDLS**NSRTG	SPMIP**GSLTPPDKMNGE**HHPGH	11
Strigamia maritima	FADNH**PVDLS**NSHRG	SHMIA**GSLTPPDKVNGE**HGHQL	12
Metaseiulus occidentalis	GADRK**PLDMS**AAHRS		13
Ixodes scapularis	QAAGA**PVDMS**SHPAR		14
***Parasteatoda tepidariorum* 1**	VIDSH**PVDLS**SPKPS		15
***Parasteatoda tepidariorum* 2**	RYEGR**PVDLS**SPRPN		16
***Limulus polyphemus* 1**	PYDGH**PVDLS**NQRPD		17
***Limulus polyphemus* 2**	TYESH**PVDLS**NQRPD		18
Centruroides sculpturatus	GYESS**PVDLS**SHRSV	MQLIS**GSMTSHDKVNGD**QHSLG	19
Euperipatoides kanangrensis	NSYDN**PVDLS**SHRSS	QQILP**GSLGPSDKVNGD**LVSLA	20
Naineris dendritica	DPNGH**PVDL**SHSRHI	PHMIH**GSLTPPDRVNGE**PGSGL	21
Glycera dibranchiata	APYDD**PVDLS**SRHPA	GHMIP**GTLTPPDKLNGD**HHHHH	22
Nephasoma pellucidum	AGYET**PVDLS**SPRPC	SHLIP**GSLTPPDKINGE**GITTS	23
Owenia sp.	QPYEN**PVDLS**RRHIK	AHLIP**GSLTPPDKINGD**MVTMA	24
Octopus bimaculoides	NGFDN**PMDLS**NGKVV	HLMPA**GSLTPPDKISGD**SISMA	25
Crassostrea gigas	GGYEN**PMDLS**SNKPG	SHIVA**GSLTPPEKINGD**PGAMA	26
Lottia gigantea	AGVEN**PVDLS**NGRIS	SHLFT**GSLTPPEKPNGD**LVPMS	27
Notospermus geniculatus	VQYDN**PIDLS**NRLEG	NHMIP**GSLTPPDKVNGD**MVPLP	28
Malacobdella grossa	LHYDN**PLDLT**NRLDE	GSGIA**GSMTPPDGGKGN**DLDLQ	29
Lingula anatina	GGYEN**PMDLS**RRTEM	AHMIP**GNLTPPDKVNGE**MVPMA	30
Phoronis australis	QHDNR**PMDLS**SRGQH	SHLIA**GSLTPPDKVNGD**VVSMA	31
Procotyla fluviatilis*	ETLFE**PLDLR**SPIGV		32
Brachionus koreanus	AKDET**PIDLS**SKKSK		33
Xenoturbella bocki	KRYSA**PLNLT**VHDKC	DVRVL**GRLTPPDKQHVN**NDVGA	34

Shown are alignments of short linear amino acid motifs (bold) in the N-terminal region of Su(H) proteins that mediate direct recruitment of the co-repressors CtBP and Gro. Numbers in the column at right refer to the source of the corresponding sequence data, with accession numbers and publication citations indicated (see list below).

1 GCCL01029953.1 (Simon et al. 2018)
2 LYUA01002621.1 (Ioannidis et al. 2017)
3 GAYJ02050375.1 (Misof et al. 2014)

4 Scaffold 1 (i5K 2013)
5 JYFJ02000853.1 (i5K 2013)
6 NIPM01000059.1 (Wu et al. 2017a)
7 GGJN01104381.1 (unpublished)
8 NW_017238139.1 (i5K 2013)
9 NW_019396104.1 (i5K 2013)
10 BAYF01001879.1 (Ikeda et al. 2015)
11 JW959185.1 (Sahoo et al. 2013)
12 JH430541.1 (Chipman et al. 2014)
13 NW_003805473.1 (Hoy et al. 2016)
14 NW_002722632.1 (Gulia-Nuss et al. 2016)
15 NW_018383625.1 (Schwager et al. 2017)
16 NW_018370942.1 (Schwager et al. 2017)
17 NW_013671976.1 (Battelle et al. 2016)
18 NW_013676581.1 (Battelle et al. 2016)
19 NW_019384406.1 (Schwager et al. 2017)
20 LN881712.1 (unpublished)
21 (Andrade et al. 2015)
22 GASB01032477.1 (von Reumont et al. 2014)
23 (Lemer et al. 2015)
24 (Andrade et al. 2015)
25 NW_014678436.1 (Albertin et al. 2015)
26 NW_011936122.1 (Zhang et al. 2012)
27 NW_008708401.1 (Simakov et al. 2013)
28 GFRY01035878.1 (Luo et al. 2018)
29 (Whelan et al. 2014)
30 GDJY01029776.1 (Luo et al. 2015)
31 GFSC01078935.1 (Luo et al. 2018)
32 GAKZ01044347.1 (unpublished)
33 GBXV02009219.1 (Lee et al. 2015)
34 (Brauchle et al. 2018)

Figure legends

Figure 1. Hairless mediates indirect recruitment of co-repressor proteins to Su(H). (**A**) Diagram denoting locations of conserved domains and motifs within Hairless, and illustrating extreme size differences of the protein in different species. Shown is Hairless from the carpenter bee *Ceratina calcarata* and the blowfly *Protophormia terraenovae* (Hase et al. 2017), with scale and protein sizes indicated. (**B**) Summary of Hairless's mode of action as an adaptor protein that recruits the global co-repressors C-terminal Binding Protein (CtBP) and Groucho (Gro) to Suppressor of Hairless [Su(H)], the transducing transcription factor for the Notch (N) cell-cell signaling pathway (Barolo et al. 2002). In the absence of signaling through the Notch receptor (left), Su(H) acts as a repressor of Notch target genes, despite the presence of transcriptional activator proteins (orange oval). Upon activation of the Notch receptor (middle), Su(H), in a complex with the receptor's intracellular domain (NICD) and the co-activator Mastermind (Mam), functions to activate transcription of pathway target genes in cooperation with other transcriptional activators. In the absence of Hairless and hence in the absence of Su(H)'s repressive activity (right), the partner transcription factors are often sufficient to activate expression of target genes in a signal-independent manner (Barolo and Posakony 2002).

Figure 1-figure supplement 1. Graph showing predicted disordered regions in *Drosophila melanogaster* Hairless, generated by DISOPRED3 (Buchan et al. 2013; Jones and Cozzetto 2015). Locations of the Gro-binding motif (purple arrow) and CtBP-binding motif (red arrow) are indicated. Note that most of the region between these two motifs is strongly predicted to be disordered.

Figure 2. Phylogenetic distribution of Hairless and related S-CAP proteins**.** (**A**) Based on extensive BLAST searches of available genome and transcriptome assemblies, orthologs of the canonical *Hairless* gene are found only in the Pancrustacea (blue bar), while orthologs of a gene that encodes the related S-CAP protein are found in the Myriapods (mS-CAP, red bar) and Chelicerates (cS-CAP, pink bar). We suggest S-CAP as a suitable umbrella nomenclature for this gene family (black bracket). Tree adapted from Figure 2 of Regier *et al.* (Regier et al. 2010). (**B**) Consistent with the presence of Hairless as an adaptor protein, Su(H) in most insect orders (the Neoptera clade) has lost the ancestral short linear motifs that mediate direct recruitment of the CtBP and Gro co-repressor proteins (red bar). However, in the Crustacea, Collembola, Diplura, and a subset of Insecta, the ancestral recruitment motifs have been retained in Su(H), despite the presence of Hairless (see Table 1 and Supplementary file 3). Tree adapted from Misof *et al.* (Misof et al. 2014) and Kjer *et al.* (Kjer et al. 2016).

Figure 3. Direct binding of co-repressor proteins by Su(H) from the centipede *Strigamia maritima*. (**A**) The PVDLS motif in the N-terminal region of Su(H) from the centipede *Strigamia maritima* directly binds *Drosophila* CtBP. A His-tagged 116-aa segment of the *Strigamia* Su(H) protein, bearing a PVDLS recruitment motif for CtBP, binds strongly to GST-dCtBP (WT, lane 2). Mutation of the motif to alanines (AAAAA) abolishes this interaction (PVDLSmut, lane 1). (**B**) The conserved GSLTPPDKV motif in the N-terminal region of *Strigamia* Su(H) directly binds *Drosophila* Gro. His-tagged E(spl)mβ-HLH protein, which bears a C-terminal WRPW motif that recruits Gro, is used as a binding control. Wild-type (WT) HLHmβ binds GST-Gro (lane 2), while a truncated

version of the protein lacking the WRPW motif (lane 1) fails to bind. A synthetic version of HLHmβ in which the WRPW motif has been replaced by the wild-type GSLTPPDKV motif also binds GST-Gro efficiently (lane 4), while a mutant version in which GSLTPPDKV is replaced by alanines (AAAAAAAAA) shows extremely weak binding (lane 3). No binding of any of the His-tagged proteins to GST alone is observed, even with substantially greater amounts of GST compared to GST-Gro.

Figure 4. S-CAP proteins in Myriapods and Chelicerates contain a Hairless-like domain that binds Su(H)**.** (**A**) Alignment of the Suppressor of Hairless Binding Domain (SBD) in *Drosophila melanogaster* (Dmel) Hairless with the related motif in the S-CAP proteins from a representative set of Myriapods and Chelicerates. Numbers flanking each sequence segment represent amino acid positions within the protein. The contiguous SBD motif is highlighted in bold. Pairwise amino acid sequence identities within the motifs are indicated by vertical lines; conservative substitutions are indicated by + signs. Amino acids in Hairless that have been shown to make direct contact with Su(H) [including the non-contiguous tryptophan (W) residue] (Yuan et al. 2016) are highlighted in red. Hydrophobic residues nearly always found immediately adjacent to the W are underlined. Species names are as follows: Smar (*Strigamia maritima*) [see also (Maier 2019)]; Hsub (*Hydroschendyla submarina*) (Fernández et al. 2016); Agir (*Anopsobius giribeti*) (Fernández et al. 2016); Ptep (*Parasteatoda tepidariorum*); Hruf (*Hypochthonius rufulus*) (Bast et al. 2016); Ssca (*Sarcoptes scabiei*); Vjac (*Varroa jacobsoni*) (Techer et al. 2019); Iper (*Ixodes persulcatus*); Mmar (*Mesobuthus martensii*) (Cao et al. 2013); Lpol (*Limulus polyphemus*). (**B**) SBD motifs in both Hairless and S-CAP proteins (red) are encoded in two exons with the same splice junction (indicated by /; see grey highlight).

Pairwise amino acid sequence identities within the motifs are indicated by vertical lines; conservative substitutions are indicated by + signs. Species names as in A, except for Isca (*Ixodes scapularis*). (**C**) Spider S-CAP protein binds directly to *Drosophila* Su(H) in vitro. In all panels, lanes 1-4 represent the indicated His-tagged segments of wild-type *Drosophila* (Dmel) Hairless (lane 1); Dmel Hairless bearing alanine substitutions for each of five SBD residues shown to contact Su(H) (lane 2); wild-type S-CAP from the spider *Parasteatoda tepidariorum* (Ptep) (lane 3); Ptep S-CAP bearing the same five alanine substitutions (lane 4). Input levels of these His-tagged proteins for each experiment are shown in the respective "input" panels. Remaining two panels show the results of pulldown assays using Sepharose beads bearing only GST (left side) or GST-Su(H) (right side). Left: No binding of the His-tagged proteins to GST alone is observed. Right: Wild-type Dmel Hairless binds efficiently to GST-Su(H) (lane 1); this interaction is severely reduced by the introduction of the five alanine substitutions (lane 2). Wild-type Ptep S-CAP likewise binds to GST-Su(H) (lane 3), while no binding is observed with the alanine-substitution mutant (lane 4); the same result is obtained even when the amount of input Ptep S-CAPs (wild-type and mutant) is increased by a factor of 10 (lanes 3 and 4, 10X). Amounts of GST and GST-Su(H) on the beads are shown in the Coomassie stains below the corresponding pulldown lanes.

Figure 5. Sequence characteristics of S-CAP proteins in Myriapods and Chelicerates. (**A**) Diagrams of representative examples of Myriapod and Chelicerate S-CAP proteins, denoting locations of SBD motifs and CtBP recruitment motifs. Scale and protein sizes are indicated. (**B**) Chelicerate, but not Myriapod, S-CAP proteins share N-terminal ELM2 and SANT domains with an MTA zinc-finger protein from the same species. Scale and protein sizes are indicated. (**C**) Shared ELM2 and SANT domains in

Chelicerate MTA and S-CAP proteins are encoded in two exons with the same splice junction (indicated by /; red arrows). Pairwise amino acid sequence identities within the motifs are indicated by vertical lines; conservative substitutions are indicated by + signs. Species names as in Figure 4A.

Figure 5-figure supplement 1. Alignments of sequence regions shared by representative S-CAP proteins from (**A**) Chelicerates and (**B**) Myriapods. Amino acids at the ends of the aligned sequences are numbered. Pairwise amino acid sequence identities are indicated by vertical lines; conservative substitutions are indicated by + signs. ELM2 and SANT domains found in Chelicerate S-CAPs (A) are highlighted in colors corresponding to those in Figure 5B. Species names are as follows: (A) Hruf (*Hypochthonius rufulus*), Lpol (*Limulus polyphemus*), Otur (*Ornithodoros turicata*) (Egekwu et al. 2016), Tser (*Tityus serrulatus*) (Fuzita et al. 2015), Lhes (*Latrodectus hesperus*) (Clarke et al. 2015); (B) Smar (*Strigamia maritima*), Agir (*Anopsobius giribeti).*

Figure 6. Genes encoding both Hairless and S-CAP proteins are frequently located immediately adjacent to an *MTA* gene. Separate scale for each diagram is shown at the left. Three examples are shown for *Hairless*: the carpenter bee *Ceratina calcarata* (Ccal), the wheat aphid *Schizaphis graminum* (Sgra) (QEWZ01001380.1), and the lady beetle *Harmonia axyridis* (Haxy). Note that microsynteny is often preserved even when gene locations and relative orientations are changed. One example each is shown for *S-CAP* in Myriapods [the centipede *Strigamia maritima* (Smar)] and Chelicerates [the house spider *Parasteatoda tepidariorum* (Ptep)]. See also Supplementary file 1 and Supplementary file 4.

Figure 7. Speculative possible sources for key elements of the Hairless and S-CAP proteins. (**A**) A C-terminal segment of the highly conserved Yippee-like protein (Roxström-Lindquist and Faye 2001; Hosono et al. 2004) is closely related to the C-terminal half of Hairless and S-CAP SBDs. Upper diagram is a sequence alignment of the entire Yippee-like proteins from *Drosophila melanogaster* (Dm) and *Homo sapiens* (Hs). Aligned below are contiguous SBD motifs from *Drosophila* Hairless and five Myriapod and Chelicerate S-CAPs; their C-terminal halves are shown in bold. Two leucine (L) residues shown to make direct contact with Su(H) (Yuan et al. 2016) are highlighted in red. Amino acid sequence identities are indicated by vertical lines; conservative substitutions are indicated by + signs. Other species names as in Figure 4A. (**B**) As shown in the gene diagram at the bottom, the CtBP recruitment motif in Hairless is encoded by a very small exon located at the extreme 3′ end of the gene [example is from the Oriental fruit fly *Bactrocera dorsalis* (Bdor; JFBF01000273.1); scale indicated]. A pre-existing gene encoding a protein that utilizes the same PLNLS recruitment motif is a possible source of this exon. Example shown is a portion of the *senseless* gene from the red flour beetle *Tribolium castaneum* (Tribolium et al. 2008). Senseless directly recruits the CtBP co-repressor via the PLNLS motif (Miller et al. 2014). This portion of the protein is encoded in exon 2; splice junction is indicated by a red /. Aligned beneath it is the last exon of the Bdor *Hairless* gene, illustrating its splice junction in the same frame as *senseless* exon 2.

Supplementary file legends

Supplementary file 1. Representative catalog of Hairless proteins in the Pancrustacea, selected from a curated collection of approximately 400 full-length sequences. Annotated sequences of the entries in this list are provided in Supplementary file 2. In the "MTA microsynteny?" column, + and – in parentheses denote the relative orientation of the *Hairless* and *MTA* genes in the genome.

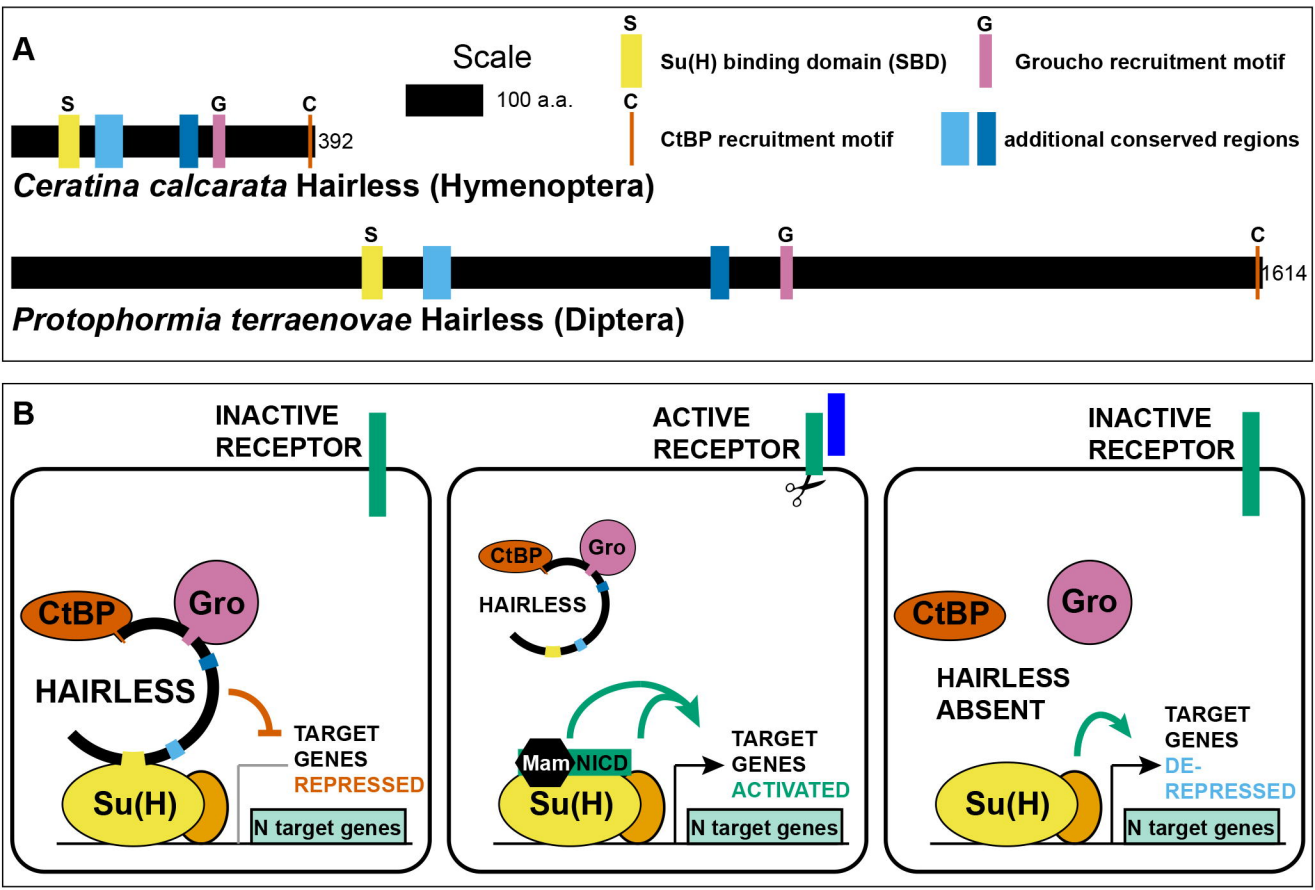
Supplementary file 2. Annotated FASTA file of sequences of the Hairless proteins included in Supplementary file 1. Characteristic conserved domains and motifs are colored as in Figure 1A.

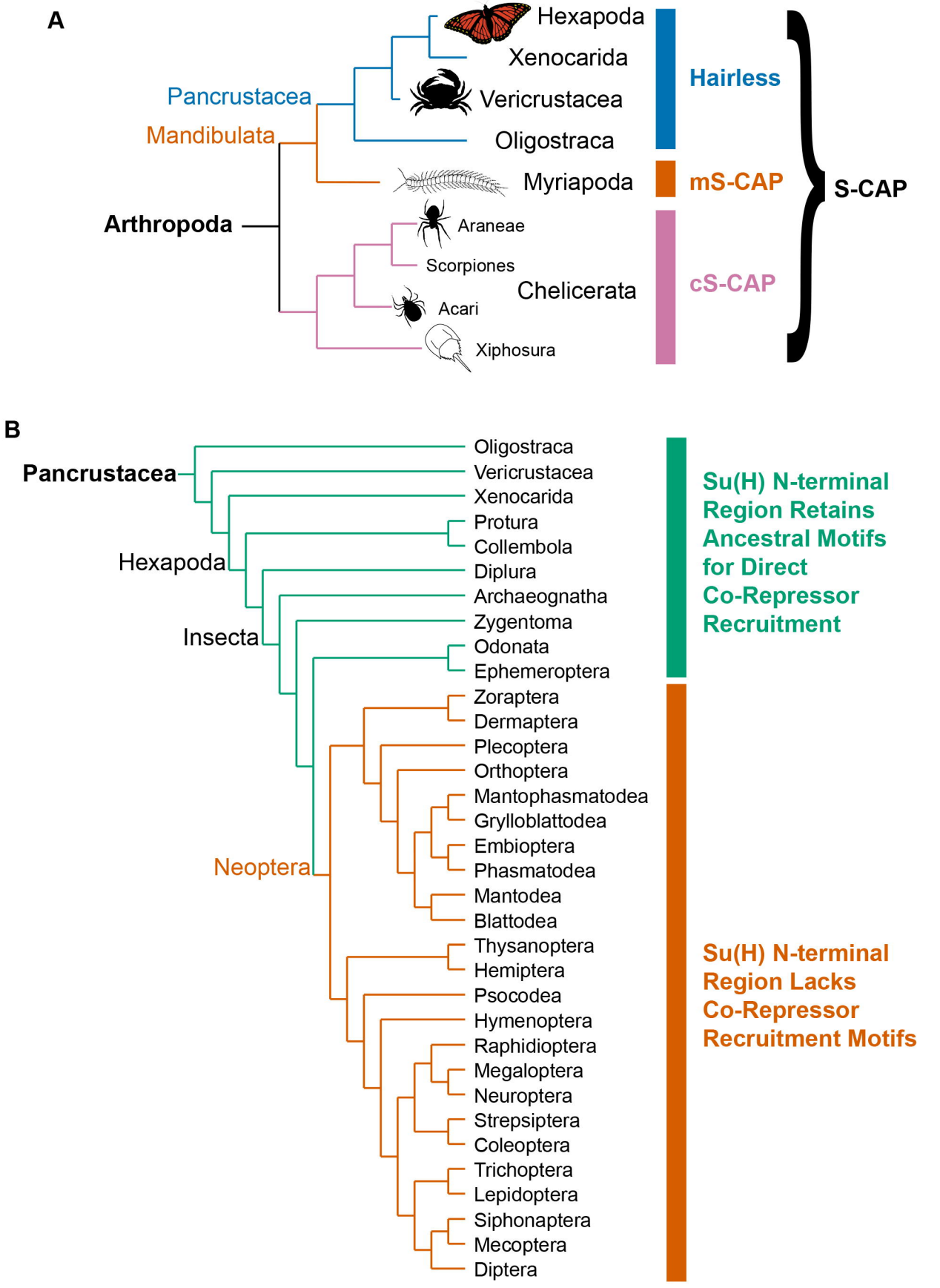
Supplementary file 3. Annotated FASTA file of sequences of the Su(H) proteins included in Table 1; species for which a full-length Su(H) sequence is not available are omitted here. Colors denote conserved sequence features. Motifs for direct recruitment of the CtBP and Gro co-repressor proteins (aligned in Table 1) are shown in red and green, respectively. Large region highlighted in orange is the highly conserved body of Su(H), extending from "LTREAM" to "YTPEP".

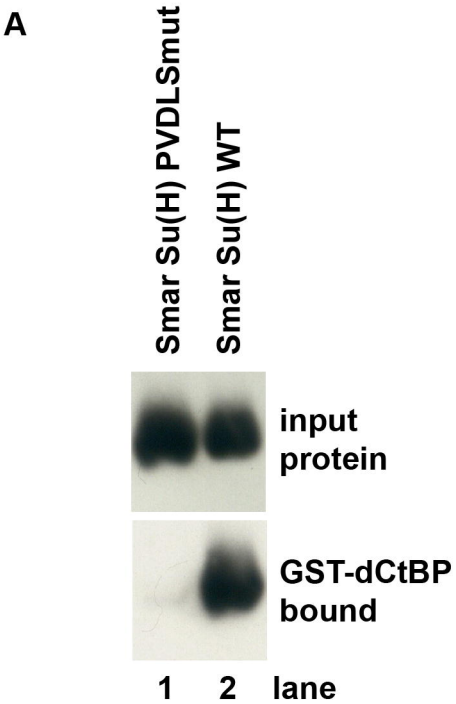
Supplementary file 4. Representative catalog of S-CAP proteins in the Myriapoda and Chelicerata, selected from a curated collection of approximately 50 full-length sequences. Annotated sequences of the entries in this list are provided in Supplementary file 5. In the "MTA microsynteny?" column, + and – in parentheses denote the relative orientation of the *S-CAP* and *MTA* genes in the genome. The "H SBD splice?" column indicates whether the Su(H)-binding domain (SBD) in the listed

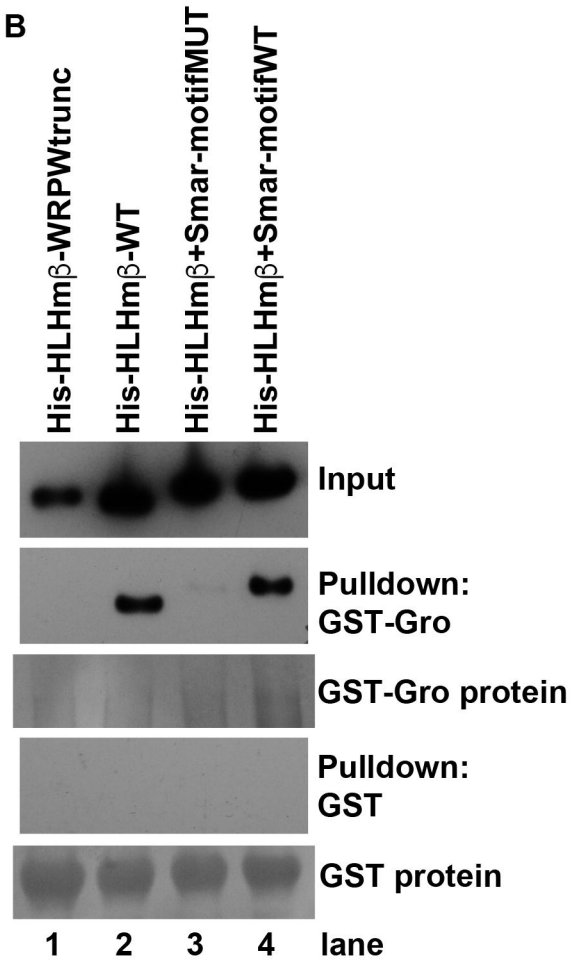
protein is encoded by two exons with the same splice junction as in *Hairless*. If not, the alternative exon structure is indicated.

Supplementary file 5. Annotated FASTA file of sequences of the S-CAP proteins included in Supplementary file 4. Characteristic domains and motifs are colored as follows: Su(H)-binding domain (SBD), orange; CtBP recruitment motifs, red.


A
Scale
100 a.a.
S
Su(H) binding domain (SBD)
G
Groucho recruitment motif
C
CtBP recruitment motif
additional conserved regions
S
G
C
392
Ceratina calcarata Hairless (Hymenoptera)
S
G
C
1614
Protophormia terraenovae Hairless (Diptera)
B
INACTIVE RECEPTOR
CtBP
Gro
HAIRLESS
TARGET GENES REPRESSED
Su(H)
N target genes
ACTIVE RECEPTOR
CtBP
Gro
HAIRLESS
Mam
NICD
Su(H)
TARGET GENES ACTIVATED
N target genes
INACTIVE RECEPTOR
CtBP
Gro
HAIRLESS ABSENT
TARGET GENES DE-REPRESSED
Su(H)
N target genes


A
Hexapoda
Xenocarida
Vericrustacea
Oligostraca
Pancrustacea
Mandibulata
Myriapoda
Arthropoda
Araneae
Scorpiones
Chelicerata
Acari
Xiphosura
Hairless
mS-CAP
cS-CAP
S-CAP
B
Pancrustacea
Oligostraca
Vericrustacea
Xenocarida
Protura
Collembola
Hexapoda
Diplura
Archaeognatha
Zygentoma
Insecta
Odonata
Ephemeroptera
Zoraptera
Dermaptera
Plecoptera
Orthoptera
Mantophasmatodea
Grylloblattodea
Embioptera
Neoptera
Phasmatodea
Mantodea
Blattodea
Thysanoptera
Hemiptera
Psocodea
Hymenoptera
Raphidioptera
Megaloptera
Neuroptera
Strepsiptera
Coleoptera
Trichoptera
Lepidoptera
Siphonaptera
Mecoptera
Diptera
Su(H) N-terminal Region Retains Ancestral Motifs for Direct Co-Repressor Recruitment
Su(H) N-terminal Region Lacks Co-Repressor Recruitment Motifs


A
Smar Su(H) PVDLSmut
Smar Su(H) WT
input protein
GST-dCtBP bound
1
2
lane


B
His-HLHmβ-WRPWtrunc
His-HLHmβ-WT
His-HLHmβ+Smar-motifMUT
His-HLHmβ+Smar-motifWT
Input
Pulldown: GST-Gro
GST-Gro protein
Pulldown: GST
GST protein
1
2
3
4
lane

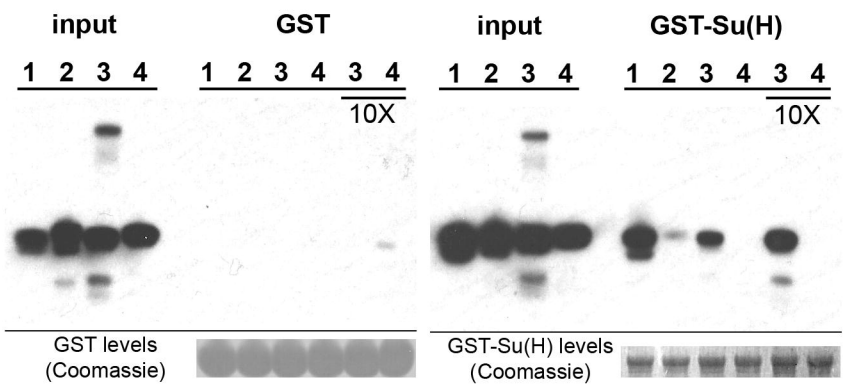
A

```
Dmel Hairless   223 HGNNSWGGYGGRLQFFKDGKFILELARSKDGDKSGWVSVTRKTFRPPSAA 272
                              |+|  |+ ||+|||||
Smar S-CAP       77 RTGSSENDTNGKLSFYLDGRFILELEHKPGGKRRNGGWIQTKGITWMPRP 126
                              ||| |||||||+|||
Hsub S-CAP       68 EENVGNSNLNGKLRFYLDGRFVLELRHTSDELTHAGKWIPSEEKYKGDNQ 117
                              |+|||+|||||||||
Agir S-CAP      248 SSSIVQQESSGRLRFFLDGRFVLELTHAGDNTRNTSGWVQSKGKIWTPSS 297
                              | |+||| ||+|| |
Ptep S-CAP      265 PQDVPATSTVGSLKFFLGGRLVLKLNAQQDGGSGNKCQWVQSNDLPKHSN 314
                              |||||+  | |||||
Hruf S-CAP      336 QESSGSISTVGSLKFYKDGQLVLKLNAQQEDTGQKCLWIESQDTPKLPRP 385
                              ||| +|  |+|||+|
Ssca S-CAP      501 NEADSSKTNLGSLNLYLHGELVLRLNAQQQDSGQKWVESNEIQSSFKDDA 550
                              ||| +|| |+|+|+|
Vjac S-CAP      288 GAGGGAAPGQGSLKFYLGGQLILKLNHQAHQERKSWVEAPDNPAQVTSPA 337
                              |||||+||| |+|||
Iper S-CAPa     294 EESPASVAPVGSLKFFLGGRLVLKLSAQEGGAWVEAQDTPRLGRPRQPPP 343
                              ||| ||||| |||||
Mmar S-CAP      275 SNNESTSFNLGSLSFFLGGQLVLKLNAQQENGPDNQNCQWVLSQDTPKLP 324
                              ||| |++ | |||||
Lpol S-CAPa     283 ATLDLETSPLGSLQFYFHGKLVLKLNAQQHANTNGHQCQWVPSVDTPKVP 332
```

B

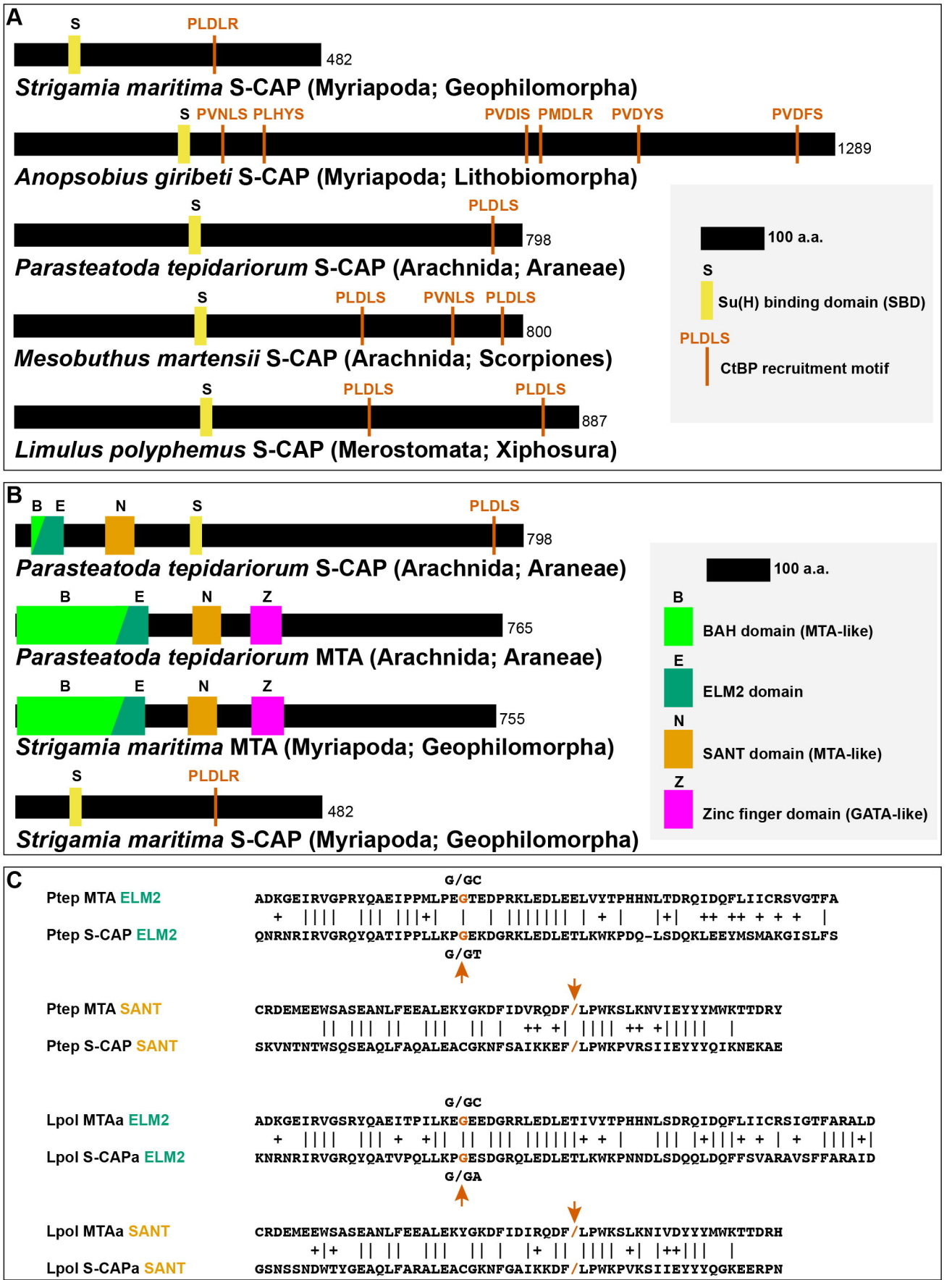
```
Dmel Hairless   GGTGGTCGTTTGCAGTTCTTTAAAG/ATGGCAAATTCATATTGGAACTGGCGCGGTCC
                G  G  R  L  Q  F  F  K  D   G  K  F  I  L  E  L  A  R  S
                   |  +  |     |  +     |   |  +  |  |  |  |  |
Smar S-CAP      AATGGCAAACTTAGTTTTTATTTGG/ACGGCCGATTCATCTTGGAATTGGAACACAAA
                N  G  K  L  S  F  Y  L  D   G  R  F  I  L  E  L  E  H  K
                   |     |     |  +  |     |   |  +  +  |     |
Ptep S-CAP      GTTGGAAGCTTGAAATTTTTTTTAG/GAGGTCGTTTAGTACTTAAATTAAATGCGCAA
                V  G  S  L  K  F  F  L  G   G  R  L  V  L  K  L  N  A  Q
                +  |  |  |     |  +  +     |  +  |  |  |  |  |  |  |  |
Lpol S-CAP      CTTGGAAGCCTTCAGTTCTATTTCC/ATGGAAAACTGGTTCTGAAACTTAATGCCCAA
                L  G  S  L  Q  F  Y  F  H   G  K  L  V  L  K  L  N  A  Q
                |  |  |  |     |  +  +     |     |  |  |  |  |  |  |  |
Mmar S-CAP      CTTGGAAGTTTAAGTTTTTTCCTTG/GAGGCCAGTTGGTACTGAAATTAAATGCCCAA
                L  G  S  L  S  F  F  L  G   G  Q  L  V  L  K  L  N  A  Q
                +  |  |  |     |  |  |  |   |     |  |  |  |  |     |  |
Isca S-CAP      GTGGGCAGCCTCAAGTTCTTCCTCG/GCGGCCGCCTGGTGTTGAAGTTGAGCGCCCAG
                V  G  S  L  K  F  F  L  G   G  R  L  V  L  K  L  S  A  Q
```

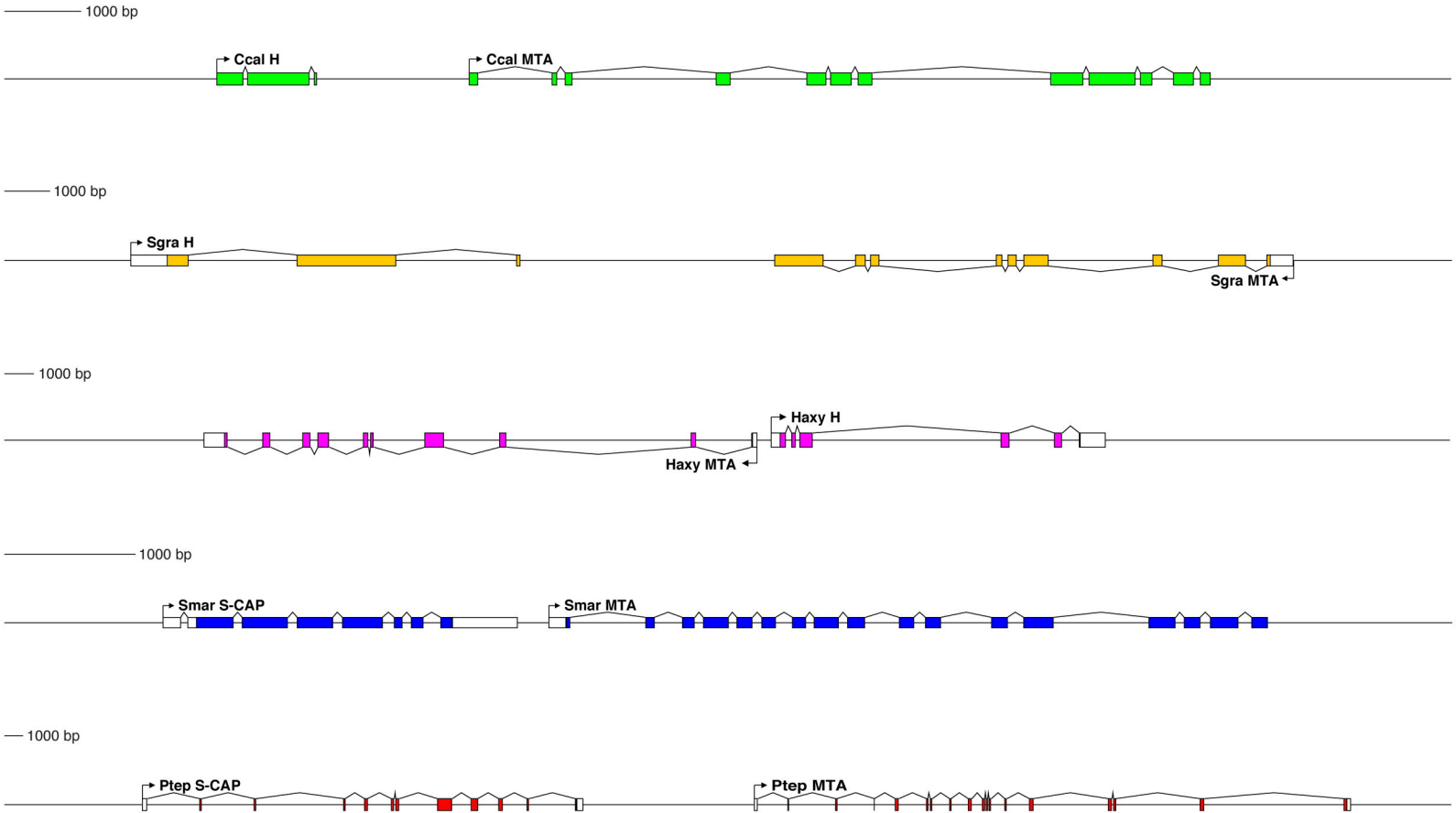
C

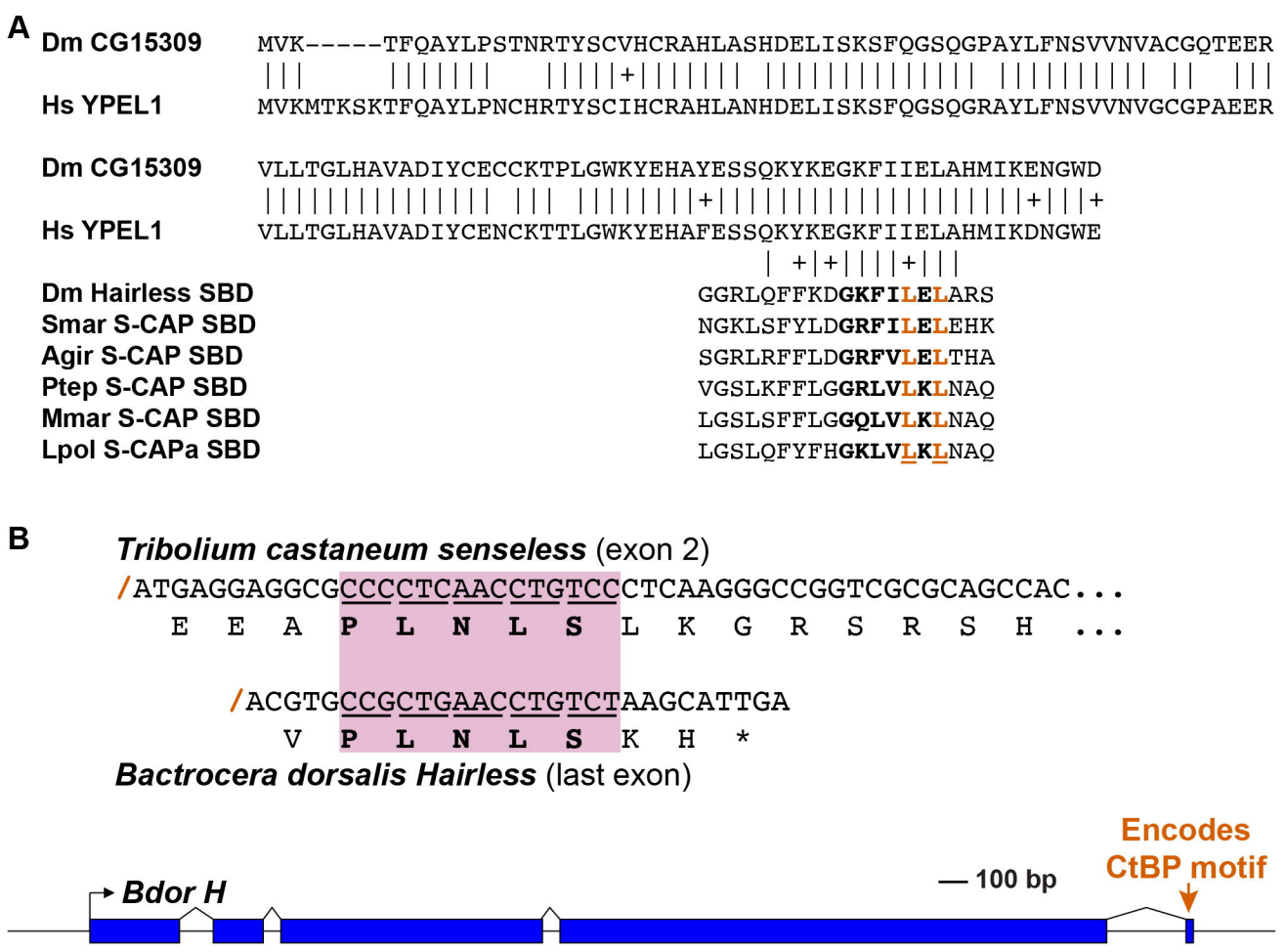


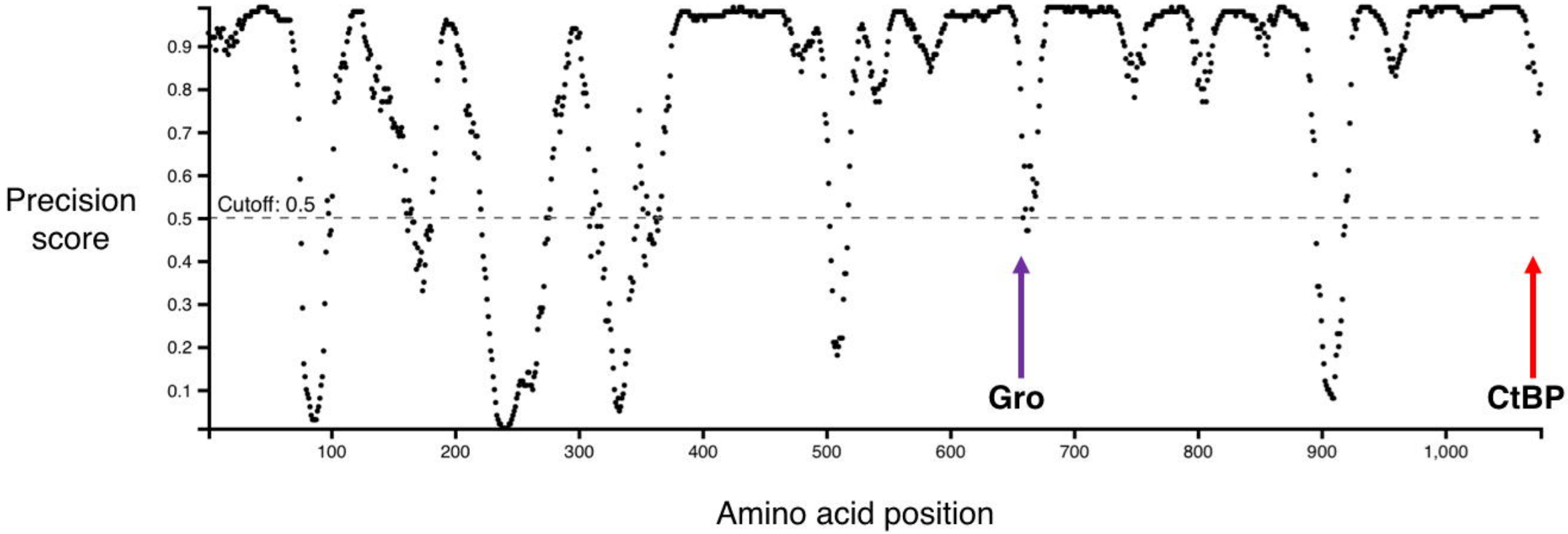
HIS-tagged constructs:

1) Dmel Hairless 192-389
2) Dmel Hairless 192-389 5Amut
3) Ptep S-CAP 233-432
4) Ptep S-CAP 233-432 5Amut


A
S
PLDLR
482
Strigamia maritima S-CAP (Myriapoda; Geophilomorpha)
S PVNLS PLHYS PVDIS PMDLR PVDYS PVDFS
1289
Anopsobius giribeti S-CAP (Myriapoda; Lithobiomorpha)
S
PLDLS
798
Parasteatoda tepidariorum S-CAP (Arachnida; Araneae)
S
PLDLS PVNLS PLDLS
800
Mesobuthus martensii S-CAP (Arachnida; Scorpiones)
S
PLDLS PLDLS
887
Limulus polyphemus S-CAP (Merostomata; Xiphosura)
100 a.a.
S
Su(H) binding domain (SBD)
PLDLS
CtBP recruitment motif
B
B E N S
PLDLS
798
Parasteatoda tepidariorum S-CAP (Arachnida; Araneae)
B E N Z
765
Parasteatoda tepidariorum MTA (Arachnida; Araneae)
B E N Z
755
Strigamia maritima MTA (Myriapoda; Geophilomorpha)
S
PLDLR
482
Strigamia maritima S-CAP (Myriapoda; Geophilomorpha)
100 a.a.
B
BAH domain (MTA-like)
E
ELM2 domain
N
SANT domain (MTA-like)
Z
Zinc finger domain (GATA-like)
C
G/GC
Ptep MTA ELM2 ADKGEIRVGPRYQAEIPPMLPEGTEDPRKLEDLEELVYTPHHNLTDRQIDQFLIICRSVGTFA
Ptep S-CAP ELM2 QNRNRIRVGRQYQATIPPLLKPGEKDGRKLEDLETLKWKPDQ-LSDQKLEEYMSMAKGISLFS
G/GT
Ptep MTA SANT CRDEMEEWSASEANLFEEALEKYGKDFIDVRQDF/LPWKSLKNVIEYYYMWKTTDRY
Ptep S-CAP SANT SKVNTNTWSQSEAQLFAQALEACGKNFSAIKKEF/LPWKPVRSIIEYYYQIKNEKAE
G/GC
Lpol MTAa ELM2 ADKGEIRVGSRYQAEITPILKEGEEDGRRLEDLETIVYTPHHNLSDRQIDQFLIICRSIGTFARALD
Lpol S-CAPa ELM2 KNRNRIRVGRQYQATVPQLLKPGESDGRQLEDLETLKWKPNNDLSDQQLDQFFSVARAVSFFARAID
G/GA
Lpol MTAa SANT CRDEMEEWSASEANLFEEALEKYGKDFIDIRQDF/LPWKSLKNIVDYYYMWKTTDRH
Lpol S-CAPa SANT GSNSSNDWTYGEAQLFARALEACGKNFGAIKKDF/LPWKPVKSIIEYYYQGKEERPN


1000 bp
Ccal H
Ccal MTA
1000 bp
Sgra H
Sgra MTA
1000 bp
Haxy H
Haxy MTA
1000 bp
Smar S-CAP
Smar MTA
1000 bp
Ptep S-CAP
Ptep MTA


A
Dm CG15309
MVK-----TFQAYLPSTNRTYSCVHCRAHLASHDELISKSFQGSQGPAYLFNSVVNVACGQTEER
Hs YPEL1
MVKMTKSKTFQAYLPNCHRTYSCIHCRAHLANHDELISKSFQGSQGRAYLFNSVVNVGCGPAEER
Dm CG15309
VLLTGLHAVADIYCECCKTPLGWKYEHAYESSQKYKEGKFIIELAHMIKENGWD
Hs YPEL1
VLLTGLHAVADIYCENCKTTLGWKYEHAFESSQKYKEGKFIIELAHMIKDNGWE
Dm Hairless SBD
GGRLQFFKDGKFILELARS
Smar S-CAP SBD
NGKLSFYLDGRFILELEHK
Agir S-CAP SBD
SGRLRFFLDGRFVLELTHA
Ptep S-CAP SBD
VGSLKFFLGGRLVLKLNAQ
Mmar S-CAP SBD
LGSLSFFLGGQLVLKLNAQ
Lpol S-CAPa SBD
LGSLQFYFHGKLVLKLNAQ
B
Tribolium castaneum senseless (exon 2)
/ATGAGGAGGCGCCCCTCAACCTGTCCCTCAAGGGCCGGTCGCGCAGCCAC...
E E A P L N L S L K G R S R S H ...
/ACGTGCCGCTGAACCTGTCTAAGCATTGA
V P L N L S K H *
Bactrocera dorsalis Hairless (last exon)
Bdor H
100 bp
Encodes CtBP motif


Precision score
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
Cutoff: 0.5
Gro
CtBP
100
200
300
400
500
600
700
800
900
1,000
Amino acid position

A

Alignment of Chelicerate cS-CAP Conserved Domains

```
Hruf S-CAP ELM2     21 PNRNRIRIGRQYQATVPPLLKPGESDGRKLEELETLRWKPDNELSDQQLDQYLSVAK-AVSLFA 83
                       |||||+|||||||||| |||||||||| ||+||||+||| |+||||||||++|||+ |||+||
Lpol S-CAPa ELM2    20 KNRNRIRVGRQYQATVPQLLKPGESDGRQLEDLETLKWKPNNDLSDQQLDQFFSVAR-AVSFFA 82
                       ||||||||||||||| |||||||||| |||||||+|+|    |||| +| ++|+|+ |||+||
Otur S-CAP ELM2     19 NNRNRIRVGRQYQATVPPLLKPGESDGRRLEDLETLRWRPES-LSDQSIDEYLSMAK-AVSLFA 80
                       ||||||||||||||||||||||+|||  ||||||+|+|+  |||| || |||||  ||||||
Tser S-CAP ELM2     25 RNRNRIRVGRQYQATVPPLLKPGETDGRCCEDLETLKWKPDNDLSDQQIDQYLSMASRAVSLFA 88
                       ||||||||||||+|||||||| ||| |||||||||||++ |||| ++ |+|||   +|||
Lhes S-CAP ELM2     21 QNRNRIRVGRQYQATIPPLLKPGEKDGRKCEDLETLKWKPEQ-LSDQKLEEYMSMAK-GISLFS 82

Hruf S-CAP          133 SASVGDHHIQSALKGLSDFVTSHHPGHHDAGCRVSDP 169
                        |    |  ||+||+||+++|||||| |||| | +|
Lpol S-CAPa          91 SDKESDKCIQTALRGLTEYVTSHHPCHHDANCQLSQG 127
                        |+  || |+||||+||+++||||| ||||| ||+
Otur S-CAP           89 SEEASDGCLQTALKGLSDFVTSHHGCHHDAGCQVGPS 125
                        +    |  |||||+|||||||||| |||||||||
Tser S-CAP           97 TGNGPDKTLQTALRGLSDFVTSHHPCHHDAGCQVPSP 133
                        |    | +||+|++||++|| |||| ||| ||
Lhes S-CAP           90 TSEKSDNSLQSAIKGLTEFVASHHPYHHDDGCRQVPK 126
```

Additional Conserved Region

```
Hruf S-CAP SANT     185 SGLITGEWTPNEMKLFTRALDVCGKNFGAIKKEFLPWKSVKSIIEHYYLGIDKLLS 240
                            + +||   |  || |||+ ||||||||||+||||| |||||| || | +
Lpol S-CAPa SANT    135 GSNSSNDWTYGEAQLFARALEACGKNFGAIKKDFLPWKPVKSIIEYYYQGKEERPN 190
                        |      |||  || ||||||+ ||||| |||||||||||||+|||||||| |+
Otur S-CAP SANT     123 GPSLPSDWTPNEASLFARALDECGKNFSAIKKDFLPWKPVKSLIEYYYQGKLEKQD 178
                          |+   ||| || ||||||+ ||||| ||||+|||||||+|+||+||+|| || |
Tser S-CAP SANT     142 KHSIKNGWTPTEAHLFARALEACGKNFGAIKKEFLPWKPVRSIIEFYYNGKEEKYD 197
                          | |+| || ||| ||| ||||| |||||||||||||||+||+ | || +
Lhes S-CAP SANT     135 PKISTNVWSPFEAQLFAQALESCGKNFCAIKKEFLPWKPVRSIIEYYYQIKNEKNE 190
```

B

Alignment of Myriapod mS-CAP Conserved Regions

```
Smar S-CAP          262 PDNASLPLPEWYLDPSKSQIPVNEMIGNPRPYLPAAPTH 300
                            |||||||||||||  +|| |||| ||||||
Agir S-CAP          534 ASSSNLPLPEWYLDPSKAEVPVSEMIGYPRPYLPPSTDS 572

Smar S-CAP           361 PSAVQPWPCVSCPVVAYAPTVVNSYYYCCCFLPGCRSNCLCKTEGKQLTS 410
                            + |||   |||||| | +|||||||||||||||   +  |||+ +
Agir S-CAP          1009 SIGLAPWPYPPCPVVAYGPALVNSYYYCCCFLPGCRFHYMYNTEGRPVIA 1058

Smar S-CAP           454 DGCSHSMLKDLLLMPSSE 471
                            | |||||||| ||
Agir S-CAP          1139 SATSGSMLKDLLLKPSRT 1156
```