

Hex Hox: De Novo Transcriptome Assembly and Embryonic Hox Gene Expression in the Burrowing Mayfly *Hexagenia Limbata* (Ephemeroidea)

Christopher J Gonzalez (✉ cjgonzalez@wildcats.unh.edu)

University of New Hampshire College of Life Sciences and Agriculture <https://orcid.org/0000-0002-6183-4856>

Tobias R Hildebrandt

Plymouth State University Department of Mathematics

Brigid C O'Donnell

Plymouth State University Department of Biological Sciences

Research

Keywords: Body plan evolution, Ephemeroptera, Abdominal appendages, Mayfly gills, Hox genes, Transcriptome, Immunohistochemistry

Posted Date: November 2nd, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-1020900/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: Hox genes are key regulators of appendage development in the insect body plan. The body plan of Mayfly (Ephemeroptera) nymphs differs due to the presence of evolutionarily significant abdominal appendages called gills. Despite mayflies' basal phylogenetic position and novel morphology amongst insects, little is known of their developmental genetics. Here we present an annotated transcriptome for the mayfly *Hexagenia limbata*, with annotated sequences for putative Hox peptides and embryonic expression profiles for the Hox genes *Antp* and *Ubx/abd-A*.

Results: Transcriptomic sequencing of early instar *H. limbata* nymphs yielded a high-quality assembly of 83,795 contigs, of which 22,975 were annotated against *Folsomia candida*, *Nilaparvata lugens*, *Zootermopsis nevadensis* and UniRef90 protein databases. Peptide annotations included eight of the ten canonical Hox genes (*lab*, *pb*, *Dfd*, *Scr*, *Antp*, *Ubx*, *abd-A* and *Abd-B*), most of which contained all functional domains and motifs conserved in insects. Expression patterns of *Antp* and *Ubx/abd-A* in *H. limbata* were visualized from early to late embryogenesis, and are also highly conserved with patterns reported for other non-holometabolous insects.

Conclusions: We present evidence that both *H. limbata* Hox peptide sequences and embryonic expression patterns for *Antp* and *Ubx/abd-A* are extensively conserved with other insects. These findings suggest mayfly *Antp* and *Ubx/abd-A* play similar appendage promoting and repressing roles in the thorax and abdomen, respectively. The identified expression of *Ubx* and *abd-A* in early instar nymphs further suggests that mayfly gill development is not subject to *Ubx* or *abd-A* repression. Previous studies have shown that insect *Ubx* and *abd-A* repress appendages by inhibiting their distal structures, which can permit the development of proximal appendage types. In line with past morphology-based work, we propose that mayfly gills are proximal appendage structures, possibly homologous to the proximal appendage structures of crustaceans.

Background

Arthropods are the most speciose clade of animals on earth, an evolutionary success widely attributed to the evolution and diversification of segmented body plans [1]. Of particular note is the insect body plan, which consists of a head with antennae and gnathal appendages, a thorax with three pairs of walking legs, and an abdomen devoid of appendages except external genitalia [2, 3]. This relatively simple body plan is the basis for a vast range of appendage diversification. One example is the transition of ancestral gnathal appendages to piercing and sucking structures in hemipterans, an elongated proboscis in many lepidopterans, a sponge-like proboscis in many dipterans, and structures adapted for nest construction or defense, as in numerous hymenopterans [4, 5]. Similarly, the thoracic leg segments may be elongated for mobility on the water surface, as in Gerridae hemipterans, or enlarged for jumping, as in many orthopterans. With such immense diversity, insects provide unique opportunities to study the evolutionary mechanisms of body patterning and appendage diversification.

From the perspective of developmental genetics, insect appendage diversity has been explained in part by the Hox genes, a highly conserved gene family first characterized in the fruit fly *Drosophila melanogaster* [6, 7]. The canonical Hox family comprises ten genes organized on a single chromosome, each expressed along the anterior-posterior axis of the embryo in parallel with their chromosomal order [8, 9, 10, 11]. Hox genes are key factors in regulating body patterning and appendage identity, developing unique appendage phenotypes, and shifting appendage morphology [12, 13, 14, 15, 11, 16, 17]. Insect Hox genes are organized in two complexes. The anterior Antennapedia complex is key for specifying the development of antennae, gnathal appendages, and thoracic legs, and consists of the genes *labial (lab)*, *proboscipedia (pb)*, *Hox3*, *Deformed (Dfd)*, *Sex combs reduced (Scr)*, *fushi tarazu (ftz)* and *Antennapedia (Antp)* [11]. The posterior Bithorax complex functions primarily to regulate appendage development in the abdomen, and contains the genes *Ultrabithorax (Ubx)*, *abdominal-A (abd-A)*, and *Abdominal-B (Abd-B)* [11]. All Hox genes contain a DNA binding homeodomain sequence, and can be further distinguished by the presence or absence of several conserved functional regions, such as the SSYF motif, hexapeptide, and UbdA motif [18, 19]. Collectively, these genes are crucial for evolving novelties in the insect body plan, such as the specialized thoracic legs of orthopterans and Gerridae hemipterans, and possibly the abdominal appendages of ephemeropterans [20, 21, 22].

Mayflies (Ephemeroptera) belong to one of the earliest branching clades of winged insects [23, 24], and develop from aquatic nymphs that develop paired abdominal gills on the first seven abdominal segments (Additional files 1: Fig.S1). Gill morphology is incredibly diverse, ranging from small thin threads to flattened, leaf-like lamellae, highly sclerotized plates, and bilamellate, feathery structures. Moreover, gill position on the abdomen may also be dorsal, lateral, or ventral, and many species exhibit shape and size differences in gills along the abdominal segments [25, 26]. Coincident with this morphological diversity, gills serve a variety of functions in food acquisition/water movement, locomotion, oxygen and ion uptake, protection of other gills, and adherence to the substrate (*e.g.* [27, 28, 29]). The unique morphology and functional roles of mayfly gills represent a distinct divergence from the appendage-less abdomen that defines most insects, raising key questions on the origins of the insect body plan and the evolution of novel appendage types, such as wings.

The evolutionary position of mayflies has sparked recent interest in their genetics. The first mayfly transcriptome for *Cloeon viridulum* (Baetidae) was sequenced to study differential gene expression during metamorphosis, and a genome and several transcriptomes of *C. dipterum* were sequenced to assess mayfly lifecycle adaptations and support the development of *C. dipterum* as an emerging model system [30, 31, 32]. However, next-generation sequencing outside the Baetidae family is sparse, as is the availability of data for Hox genes from early clades of winged insects. Furthermore, only two published studies have documented gene expression patterns during mayfly embryogenesis, both focused on the expression of the segment polarity genes *Engrailed* and *wingless* [33, 34]. The developmental genetics of nymphal body patterning, and the potential role of Hox genes, thus remains unknown.

In this study, we assembled an annotated transcriptome for early instar nymphs of the North American burrowing mayfly *Hexagenia limbata* (Ephemeridae; [35]), with detailed annotations of eight putative *H.*

limbata Hox proteins. We also provide the first visualization of Hox gene expression in a mayfly, focusing on *H. limbata Antp* and *Ubx/abd-A* from early to late embryogenesis. Our data shows that Hox peptide sequences contain all expected functional sequences for insects, while *Antp* and *Ubx/abd-A* expression domains are highly conserved with non-holometabolite insects. These results suggest that mayfly gills are proximal appendage structures not regulated by Hox gene regulation, and provide a basis for further exploration of insect appendage evolution via mayflies as an evo-devo model.

Methods

Egg collection and maintenance

Mature *H. limbata* females were collected by black lighting at Sky Pond (New Hampton, Belknap Co., NH) on peak hatch nights in June and July, 2013-2016. Eggs were extracted from captured females by submerging the abdomen into a conical tube containing pond water to stimulate egg laying [36]. Egg production amongst female mayflies tends to scale with overall body size, with each large female *Hexagenia* producing ~8,000 eggs [37, 36, 38].

Collected eggs were washed in a solution of 10% bleach and rinsed thoroughly with aged (24h) distilled water. All eggs were maintained in aged distilled water at room temperature (approximately 25°C) until fixation or reaching approximately 50% development, then stored at 4°C to induce a diapause-like state [39] for future study. Hatched nymphs were reared in glass containers with aged distilled water, while older nymphs were collected directly from pond mud samples and housed in containers filled with pond water and mud. All nymphs were maintained at room temperature and ambient light conditions (approximately 12-hour light-dark cycles).

Immunohistochemistry

Embryos were fixed by modifying an established protocol [34]. Live eggs were first washed thoroughly with PBTw (1X phosphate-buffered saline + 0.1% Tween), soaked for 6 minutes in a 50% bleach solution to remove the chorion, then fixed for 30-50 minutes with agitation in a 6% formaldehyde and PBTw fixative with heptanes at a 2:1 ratio. Following fixation, eggs were washed in PBTw and stored at -20°C in absolute methanol.

Fixed eggs were rinsed in PBTw and stripped of the vitelline membrane by submersion in a waterbath sonicator (Fisher Scientific FS20D) at 42Khz (+/- 6%) for several seconds. Embryos were then soaked in SuperBlock T20 (Thermo Scientific, MA) for 30 minutes at room temperature and incubated overnight at 4°C in either 12ng/μl of Antp 4C3 (DSHB, University of Iowa; deposited by Brower, D.) or 20ng/μl of Ubx/ABDA FP6.87 primary antibody (DSHB, University of Iowa; deposited by White, R.) diluted in SuperBlock T20. Following primary antibody incubation, the embryos were washed with PBTw for one hour (1 wash/10 minutes), then incubated for two hours in a 1:500 dilution of horseradish-peroxidase conjugated goat anti-mouse secondary antibody (Jackson ImmunoResearch, PA) in SuperBlock T20. Embryos were washed again as above in PBTw, equilibrated for 20 min (1 wash/5 minutes) in 1X stable

peroxide buffer (1XHP) (Thermo Scientific, MA), and developed for 10 minutes using 1:10 dilution of metal-enhanced diaminobenzadine substrate (Thermo Scientific, MA) in 1XHP buffer. After developing, embryos were washed in PBTw and stored at -20°C in 80% glycerol. Negative control embryos of all stages were incubated in SuperBlock T20 instead of primary antibody, and showed little sign of non-specific staining (Additional files 1: Fig. S2).

Embryos were imaged on a BX53 Olympus compound microscope using differential interference contrast optics, a Q-Color 5 Olympus camera, and QCapture Suite Plus v.3.1.3.10 (QImaging, Surrey, BC, Canada). Nymphs were imaged on a Leica EZ4 HD stereomicroscope with Leica Acquire v.1.0 (Buffalo Grove, IL, USA). Image values for exposure, contrast, light balance, and color were adjusted in Keynote v.6.6.2 to improve quality (Additional files 1: Fig. S2). Scale bars for all images were calibrated in Image J v.1.46r (Wayne Rasband, NIH, USA).

Nymphal cDNA library preparation

We used multiple nymphs as starting material to represent the full pool of transcripts present in early nymphal development. Of the approximately 100 µl of nymphs used for RNA sequencing, most were first instar, and the rest second instar. Total mRNA was extracted using TRIzol (Ambion), then column purified with RNeasy (Qiagen). Purified RNA was treated with Turbo DNase (Ambion), quantified, and checked for purity with a NanoDrop 2000 (NanoDrop Technologies, Wilmington, DE) before storage at -80°C.

Total mRNA was sent to the Hubbard Center for Genome Studies (HCGS, University of New Hampshire, Durham, NH) and checked for quality and quantity with an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA). An Illumina compatible library was constructed using an Illumina TruSeq RNA Prep Kit V2 with index Set A (RS-122-2101), following the low sample input protocol (Part #15026495 Rev. F). Briefly, 1 µg of total mRNA was used as initial input; mRNA was then purified, fragmented, and primed with random hexamers using poly-T oligo attached magnetic beads. cDNA was reverse-transcribed with SuperScript II Reverse Transcriptase, 3' adenylated, ligated with RNA adapter indices, and PCR-enriched. Finally, the cDNA library was checked for quality with an Agilent 2100 Bioanalyzer, and normalized to 10 nM prior to sequencing.

Transcriptome Assembly & Assessment

Raw paired-end reads were quality-checked using FastQC (v0.11.9; <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) before transcriptomic assembly. Two assemblies were then made, first via the de-novo assembly option in CLC Genomics WorkBench v.6.0.4 (CLCBio, Boston, MA) using raw reads, with scaffolding enabled, a minimum contig length of 200bp, and automatic word and bubble sizes of 24 and 50, respectively. The second was assembled using the Oyster River Protocol (ORP, v2.2.6; MacManes, 2018) and default settings (TPM_FILT=1, STRAND=RF, MEM=150, CPU=24), on an Amazon Web Service EC2 server with 32 vCPUs and 128 GB of RAM. The first step in the ORP pipeline is read error correction via Rcorrector (v1.0.3), followed by Illumina adaptor removal and trimming of reads with Phred quality below 3, using Trimmomatic (v0.38). Trinity (v2.8.4; [40]), Spades55, Spades75 (v3.13.0; [41]) and Transabyss (v2.0.1; [42]) are then used to make four de novo assemblies,

which are merged via a modified version of Orthofinder (v2.2.6; [43]) packaged in OrthoFuser [44]. The CLC and ORP-merged assemblies were assessed via BUSCO (v4.0.6; [45]) with the insecta_odb10 database, and TransRate (v1.0.3; [46]) read mapping.

Transcriptome Annotation

Full transcriptome annotation was done using Diamond (version 0.9.24.125; [47]) on an Amazon Web Service EC2 server (48 vCPUs, 192 GB of RAM, and one 900 NVMe SSD), with the default e-value cutoff of 0.001 against a merged protein database containing sequences from *Folsomia candida* (springtail), *Nilaparvata lugens* (brown planthopper), *Zootermopsis nevadensis* (termite) and UniRef90 (<https://www.uniprot.org/downloads>; [48]). Hox-specific annotations were further supported via a reciprocal blast pipeline, in which *D. melanogaster* protein homologs for *lab*, *pb*, *zen*, *Dfd*, *Scr*, *Ftz*, *Antp*, *Ubx*, *abd-A*, and *Abd-B* were aligned against ORP contigs with tBLASTn [49]. Because the *Hox3/zen* and *ftz* sequences differ notably from many insects in *Drosophila* and have been difficult to identify in early-branching insects, additional blasts with *zen* and *Ftz* homologs for the springtail *Folsomia candida*, brown planthopper *Nilaparvata lugens*, thrip *Frankliniella occidentalis*, locust *Schistocerca gregaria*, and red flour beetle *Tribolium castaneum* were conducted. All contigs from blast outputs with an e-value of 1e-20 or less were translated with TransDecoder (v5.5.0; <https://github.com/TransDecoder/TransDecoder/wiki>), then reciprocally blasted via BLASTp against the NCBI non-redundant protein database (<http://blast.ncbi.nlm.nih.gov/>). To further explore mayfly Hox sequences, the reciprocal blast pipeline was then used with genome-based protein datasets for the mayflies *Ephemera simulans* (*E.s.*) and *Cloeon dipterum* (*C.d.*). All mayfly proteins with consistent hits for a given Hox gene were aligned to insect Hox homologs with the MAFFT L-INS-i algorithm (v7.471; [50, 51]), and annotated for functional domains using Genomic SMART (v8.0; [52]). All accession values for publicly available sequences are provided in Additional files 1: Table S1.

Results

H. limbata read & assembly statistics

Sequenced paired-end raw reads were 151 base pairs in length. Phred-based quality scores for read pairs differed dramatically, with the forward R1 reads having a mean Phred score of 33 (99.95% call accuracy) or greater across all bases, while the reverse R3 reads had mean Phred scores that varied extensively, from 22 (99.37%) or greater initially, but dropping rapidly after base pair 64 down to 2 (36.90%) by base pair 90 (Additional files 1: Figs. S3 and S4). No reads were flagged as poor quality. The ORP and CLC assemblies of *H. limbata* produced a similar number of contigs; however, the ORP assembly produced generally longer contig sequences (Table 1). The two assemblies differed notably in terms of quality, with the ORP producing far higher TransRate and BUSCO completion scores (Table 1). Given the superior quality of the ORP assembly, open reading frame (ORF) annotation and alignments were done exclusively with ORP contigs. The only two exceptions are *H. limbata lab*, which is derived from a CLC contig, and *H. limbata Dfd*, which used the consensus from combining an ORP and CLC contig (Additional files 1: Fig.

S5). The ORP assembly was annotated against *Folsomia candida*, *Nilaparvata lugens*, *Zootermopsis nevadensis* and UniRef90 peptide databases (Table 1; Additional file 2).

Table 1
Assembly and quality metrics for ORP and CLC transcriptomes of *H. limbata* reads.

| Criteria | ORP | CLC |
|-----------------------------|--------------|-------------|
| Number of contigs | 83795 | 93561 |
| Number of contigs with ORFs | 15543 | 16100 |
| Number of annotated contigs | 25134 | NA* |
| Mean contig length | 702 | 529.36 |
| Longest contig | 26091 | 22280 |
| Shortest contig | 201 | 82 |
| TransRate Assembly score | 0.215 | 0.022 |
| TransRate Optimal score | 0.243 | 0.035 |
| BUSCO Complete Score | 1213 (88.8%) | 839 (61.4%) |

The number of contigs with ORFs was provided by TransRate. The TransRate assembly score assesses how accurate and complete an assembly is, while the optimal score is the assembly score obtained after removing all poorly assembled contigs. The complete BUSCO score represents the number of database sequences identified in the assembly, and was calculated using the insect ortholog database Odb10.

* The number of annotated ORFs was determined via Diamond annotation, and was not assessed for CLC contigs.

Hox Peptide Annotations

Assembly annotation, Reciprocal blast hits, and peptide alignments identified eight of the ten canonical Hox genes (*lab*, MZ773605; *pb*, MZ773606; *Dfd*, MZ773607; *Scr*, MZ773608; *Antp*, MZ773609; *Ubx*, MZ773610; *abd-A*, MZ773611; *Abd-B*, MZ773612) in the *H. limbata* assemblies; these were also identified in the *C. dipterum* and *E. danica* peptide datasets. Possible *Hox3* and *Ftz* peptides were identified in *E. danica* only (Additional files 1: Figs. S6 and S7). Excluding hypothetical proteins, all top reciprocal BLASTp hits were for the 8 identified Hox genes, with e-values ranging from 3E-53 to 2E-104 for *H. limbata*, 2E-28 to 1E-96 for *E. danica*, and 4E-52 to 2E-103 for *C. dipterum* (Table 2 and Additional file 1: Tables S2 and S3). Of the eight *H. limbata* Hox sequences, all but *lab*, *Dfd*, and *Abd-B* code for complete protein sequences.

Table 2
Top three reciprocal BLASTp hits for putative *H. limbata* Hox peptides.

| peptide | Species | Abbreviated Description | Accession | E-Value |
|---------|---------------------------------|-------------------------------------|----------------|---------|
| lab | <i>Nilaparvata lugens</i> | Hox-B1a-like | XP_022198853.1 | 4E-59 |
| | <i>Aethina tumida</i> | PREDICTED: Hox-B1b-like | XP_019875930.1 | 3E-53 |
| | <i>Agrius planipennis</i> | Hox-A1-like | XP_018322094.1 | 3E-53 |
| pb | <i>Cephus cinctus</i> | proboscipedia | XP_015600387.1 | 4E-90 |
| | <i>Formica exsecta</i> | proboscipedia | XP_029663337.1 | 1E-85 |
| | <i>Chelonus insularis</i> | proboscipedia | XP_034949502.1 | 3E-83 |
| Dfd | <i>Cimex lectularius</i> | Hox-B4 isoform X2 | XP_024083281.1 | 1E-75 |
| | <i>Cimex lectularius</i> | Hox-B4 isoform X3 | XP_014240109.1 | 3E-75 |
| | <i>Cryptotermes secundus</i> | deformed | XP_023702719.1 | 1E-74 |
| Scr | <i>Diabrotica</i> | Sex combs reduced-like isoform X1 | XP_028132252.1 | 5E-98 |
| | <i>Nicrophorus vespilloides</i> | PREDICTED: Sex combs reduced-like | XP_017768265.1 | 3E-97 |
| | <i>Cryptotermes secundus</i> | Sex combs reduced | XP_023702514.1 | 2E-96 |
| Antp | <i>Onthophagus taurus</i> | antennapedia | XP_022907641.1 | 2E-93 |
| | <i>Nicrophorus</i> | PREDICTED: antennapedia-like | XP_017772124.1 | 2E-91 |
| | <i>Sitophilus oryzae</i> | antennapedia-like | XP_030758247.1 | 2E-90 |
| Ubx | <i>Thrips palmi</i> | ultrabithorax isoform X2 | XP_034234417.1 | 2E-104 |
| | <i>Bemisia tabaci</i> | PREDICTED: ultrabithorax | XP_018908365.1 | 4E-102 |
| | <i>Orchesella cincta</i> | Ultrabithorax | CDI44541.1 | 6E-102 |
| abd-A | <i>Fopius arisanus</i> | PREDICTED: abd-A homolog isoform X1 | XP_011311715.1 | 2E-103 |
| | <i>Microplitis demolitor</i> | PREDICTED: abd-A homolog isoform X1 | XP_008546318.1 | 3E-103 |
| | <i>Vollenhovia emeryi</i> | PREDICTED: abd-A homolog isoform X1 | XP_011870678.1 | 5E-103 |
| Abd-B | <i>Parhyale hawaiiensis</i> | abdominal-B isoform I | AGC12523.1 | 1E-93 |

| peptide | Species | Abbreviated Description | Accession | E-Value |
|---------|-----------------------------|-----------------------------|----------------|---------|
| | <i>Hyalella azteca</i> | PREDICTED: abdominal-B-like | XP_018009927.1 | 1E-91 |
| | <i>Parhyale hawaiiensis</i> | abdominal-B isoform II | AGC12524.1 | 3E-91 |

Top reciprocal BLASTp hits exclude hypothetical proteins. *H. limbata* nucleotide accession values and contig sources are provided in Additional file 1: Table S1.

Most identified *H. limbata* Hox peptides have functional domains and motifs that are highly conserved with those of other hexapods. The homeodomain (HD) is highly conserved, with identities of over 93% compared to homologs from other hexapods (Figs. 1-8). The hexapeptide (Hx) motif was identified in seven of the eight *H. limbata* Hox peptides (Figs. 1-7), and contained a core YKWM (Fig. 1) or YPWM (Figs. 2-7) sequence. The *H. limbata* Abd-B peptide lacks the Hx motif, and instead has only one conserved tryptophan residue, similar to that of other hexapod homologs (Fig. 8). Including this conserved tryptophan, all eight *H. limbata* Hox peptides contained a linker region (LR) between the Hx and HD that was longest in the anterior Hox peptides (Figs. 1 and 2, 69 and 15 residues long respectively), and progressively shorter in most posterior peptides (e.g. Figs. <link rid="fig4">4</link> and 5, 14 and 4 residues long respectively), with three semi-conserved residues in the Abd-B homolog (Fig. 8).

Specific subgroups of Hox proteins contain additional conserved motifs. At the N-terminus is the SSYF motif found in Scr, Antp, Ubx, and abd-A peptides (Figs. 4-7). As the N-terminus arm is missing in our partial *H. limbata* Dfd peptide, it is unclear if *limbata* Dfd contains the SSYF motif as do hexapod homologs (Fig. 3). Ubx and abd-A contain a number of unique signatures identified in *H. limbata*, including the UbdA peptide, the QAQA and Poly-A sequences unique to Ubx, and the TDWM and PFER motifs unique to the abd-A linker region (Figs. 6-7).

Antennapedia and Ubx/abd-A gene expression patterns

Early *H. limbata* embryos do not show Antp expression (Fig. 9a, b). Once the precursors to the gnathal and thoracic segments form, weak expression appears in the three thoracic segments, especially along the sides where the thoracic legs will develop (Fig. 9c, d, *black arrowheads*). As the thoracic limb buds appear, expression becomes prominent at their edges (Fig. 9e-g), with midline expression present but weaker (Fig. 9h, *black arrow*). Expression in post-segmentation embryos becomes stronger in the thoracic midline, and extends through the abdominal segments beginning in A1-A3 (Fig. 9i), then through to A9 (Fig. 9j, k). In late-stage embryos (Fig. 9k), abdominal expression becomes uniform across the thorax and abdomen, but is still absent from the abdominal lateral edges and the A10 segment (Fig. 9k).

Like Antp, Ubx/abd-A expression is not evident in early *H. limbata* embryos (Fig. 10a), including those that have nearly developed the presumptive gnathal and thoracic segments (Fig. 10b). Once segmentation is complete, expression is prominent throughout all abdominal segments except for A10, with weak expression appearing at the posterior edges of T2 and T3 (Fig. 10c, *black arrows*). Thoracic expression

becomes more prominent in post-segmentation embryos (Fig. 10d), with expression beginning to spread in T3 (*black arrowheads*), and intensifying at the T3/A1 border (*black arrow*). In the oldest embryos imaged (Fig. 10e), these expression patterns persist, with strong expression at the T3/A1 border and T2 expression evident along the entire posterior border. In all documented stages, expression is not observed in the developing appendages or in the A10 segment. Unlike Antp expression, lateral expression in the segments is strong, while midline expression in the thorax (Fig. 10d) and abdomen (Fig. 10d, e) is reduced or absent.

Discussion

ORP Assembly Quality is Comparable to Assemblies of Non-model Insects

Our 88.8% ORP BUSCO completion score was similar to or higher than those seen in recently published insect transcriptome data (*e.g.* [53, 54, 55]). Likewise, our ORP TransRate assembly score of 0.215 is within the range of scores currently reported for insect transcriptome data (*e.g.* [56, 57]), and is of higher quality than nearly 50% of transcriptomes deposited in the NCBI TSA database as of 2016 [46]. The TransRate assembly score is significantly impacted by both read quality and read duplication during PCR amplification [46]. Thus, the low quality seen in many of our R3 reads may have depressed the assembly score despite read trimming done by the ORP. These factors also highlight the importance of read trimming in assembly quality, as trimming was only conducted for the ORP assembly.

Mayfly Hox peptides are highly conserved relative to other hexapods

H. limbata Hox peptides contain all functional domains and motifs widely conserved in insect Hox genes, including a homeodomain, linker region, and hexapeptide motif or residue. Several functional regions specific to particular Hox genes are present in the *H. limbata* peptides. These include the presence of an N-terminal SSYF motif in Scr, Antp, Ubx, and abd-A, and its absence in lab and Abd-B; TDWM and PFER motifs in the linker region of abd-A; the C-terminal UbdA peptide in both Ubx and abd-A; and C-terminal QAQA and poly-A sequences in Ubx [19].

Further evidence of high sequence conservation in *H. limbata* Hox genes comes from specific residues within these functional regions. For example, there are four residues unique to Hox homeodomains: a glutamic acid in alpha-helix 1, an arginine and glutamic acid in alpha-helix 2, and a methionine in alpha-helix 3 [19]. These residues were all identified in our *H. limbata* sequences (*e.g.*, residues Glu-136, Arg-148, Glu-150, and Met-171 in Fig. 1; see Figs. 2-8 and [19]). Additional unique homeodomain residues exist that are specific to particular Hox genes. Homeodomains for the Hox genes *lab* and *pb* have the largest number of unique residues, primarily within the N-terminal arm and first and third alpha-helices [19].

Three residues unique to the homeodomain N-terminals of Antp, Ubx, and abd-A (*e.g.*, Gly-22, Gln-24, and Thr-25 in Fig. 5; see also Figs. 6-7) and Abd-B (Lys-12, Lys-13, and Pro-16, Fig. 8) [19], were identified in the corresponding *H. limbata* homologs. The SSYF and hexapeptide motifs are likewise present in the putative *H. limbata* Hox peptides.

A number of residues in the linker regions of lab, pb, Dfd, and Scr peptides are also conserved, though these vary more than the homeodomain and hexapeptide regions. For example, many animal lab linker regions share a VKRXXPKTXKXE sequence [19], which in *H. limbata* is represented by VKRXXPKP (Fig. 1, residues 7-14; with the conserved threonine replaced by proline). The rest of the linker sequence varies in most aligned hexapods, a phenomenon that is also seen in the conserved XKKXXXK sequence for pb (Fig. 2, residues 7-13), and the KVHL sequence in Dfd and Scr (Figs. 3, 4, residues 11-14 and 9-12 respectively; [19]).

Antp and Ubx/abd-A embryonic expression is highly conserved amongst insects

Expression of both *Antp* and *Ubx/abd-A* during *H. limbata* embryogenesis closely resembles that of other insects, particularly non-holometabolous species. In the case of *H. limbata* *Antp*, we documented expression primarily through the embryonic thorax and abdominal midline. During segmentation in *D. melanogaster*, *Antp* expression occurs from the posterior of the labial segment to the abdominal segments, with the strongest expression in the thorax. During germ band retraction, expression remains strongest in the thorax, while abdominal expression is limited to the midline [58, 59, 60]. Most studies of *Antp* in holometabolous (*Apis mellifera*, [61]) and non-holometabolous species (*Schistocerca americana*, [59]; *Gryllus bimaculatus*, [62]) reveal an anterior expression boundary in the posterior labial segment, as we observed in *H. limbata*. Similarly, *Antp* expression in *H. limbata* occurs throughout the thorax and midline abdominal segments, and closely matches that of other holometabolous and non-holometabolous insects, though some species show lateral staining in the abdominal tracheal pits [61]. The reduced midline thoracic staining and stronger proximal staining of *H. limbata* thoracic limb buds is also observed in orthopterans [59, 62], providing further evidence that *Antp* expression is highly conserved between *H. limbata* and other insects, particularly non-holometabolous species.

Similar to *Antp* expression, *Ubx/abd-A* expression is highly conserved between *H. limbata* and other insects despite the distinct differences in development between many holometabolous and non-holometabolous species. *Ubx/abd-A* expression was strongest at the T3/A1 border and along the lateral portions of the A1-A8 abdominal segments, with weaker expression from A8-A10. In *D. melanogaster*, *Ubx* and *abd-A* show largely overlapping and complementary expression profiles. *D. melanogaster* *Ubx* is expressed before segmentation in the presumptive T3 and A1-A7 segments, particularly at the T3/A1 border and within the anterior of each segment; this pattern persists after complete segmentation, with additional expression along the abdominal midline and weakly in A8 [63, 64, 65]. After the development of all body segments, *D. melanogaster* *abd-A* expression is seen nearly simultaneously within A1-A7, most strongly within the posterior of each; like *Ubx*, it later extends to the abdominal midline and into A8 [65,

66, 67]. *Ubx* and *abd-A* expression is similar in the honeybee *Apis mellifera*, but begins in A1-A4 before spreading through A1-A7 and does not extend to the abdominal midline [61], a highly conserved pattern seen in both *H. limbata* and many other insects. In the orthopteran *Gryllus bimaculatus*, *Ubx* is first expressed in the posterior growth zone and in the presumptive T3, with expression after segmentation strongest on the T3/A1 border [62]. *Ubx* expression in the apterygote *Thermobia domestica* is similar, but also extends anteriorly around the T2 and T3 limb buds during germ band extension, similar to *H. limbata* lateral staining in the T2 and T3 segments [68]. The extension of *Ubx* and *abd-A* lateral expression through the developing abdomen until A10, followed by a post-segmentation weakening of expression from A8-A10, is widely conserved between *H. limbata* and other insect taxa [68, 62].

The role of Hox genes in nymphal body patterning

The body plan of mayfly nymphs diverges from that of most other insects in possessing unique abdominal appendages. Developmental regulation of thoracic legs and abdominal appendages in insects is controlled by the Hox genes *Antp*, *Ubx*, and *abd-A*, each of which contain functional regions well conserved between *H. limbata* and many insects. These include the SSYF motif necessary for the transcriptional activation of downstream target genes [69], and the hexapeptide motif, which contributes to extradenticle protein binding [70]. The length of each Hox gene linker region is likewise widely conserved between *H. limbata* and other insects, and facilitates proper gene function [71, 19]. However, as little is known regarding the functional significance of most linker region residues [19], the potential impact of both *H. limbata* specific and gene-specific differences in linker region sequences remains unknown. Conserved N-terminal residues of *Antp*, *Ubx*, and *abd-A* were also identified in *H. limbata*, and play a major role in specifying DNA binding affinity [72, 19]. Several functional regions are specific to *Ubx* and *abd-A*, including the *UbdA* peptide, QAQA, and poly-A sequences; all three have been demonstrated in *Drosophila* to repress the gene *Distal-less* (*Dll*) [73, 74, 75], and are present in the identified *H. limbata* homologs. *H. limbata* *abd-A* also contain the TDWM and PFER linker region motifs, which regulate extradenticle binding and *wingless* transcription, respectively [76, 77].

Collectively, these shared functional regions between studied insects and *H. limbata* suggest similar functional roles for *Antp*, *Ubx*, and *abd-A* at the phenotypic level. *Antp* promotes leg development in the thorax [78], and is abdominally expressed as part of the developing central nervous system in both insects [61, 79, 58, 59, 62] and crustaceans [80]. In *D. melanogaster*, *Ubx* and *abd-A* prevent abdominal limbs from developing through their inhibition of *Dll* [81], which specifies the distal portion of developing appendages [82, 83]. Appendage inhibition by *Ubx* and *abd-A* is also seen in lepidopterans, which require both the repression of *Ubx* and *abd-A*, and the expression of *Antp* and *Dll* in abdominal limb primordia for the larval prolegs to develop [84, 85, 86]. In coelopterans and orthopterans, *Ubx* serves as an appendage modifier that is co-expressed with *Dll* in the A1 segment, resulting in pleuropod development during embryogenesis [87, 88, 62]. Some insect taxa with abdominal structures do not show novel changes in *Ubx* and *abd-A* expression or function. For example, *Ubx* and *abd-A* expression do not appear to play a role in the development of styli on the A7-A9 segments in firebrats, possibly because styli may not be homologous to true appendages [89, 68]. In sawfly embryos, abdominal prolegs develop despite both the

abdominal expression of *Ubx* and *abd-A* and the lack of abdominal *Dll* expression, suggesting that sawfly prolegs consist exclusively of morphologically proximal structures [86, 90].

Novel *Ubx* and *abd-A* expression profiles are not observed in *H. limbata* embryos; in fact, *H. limbata Antp*, *Ubx*, and *abd-A* expression all closely resemble the expression patterns of several other insects with abdominal appendages, including the orthopteran *G. bimaculatus*, the firebrat *T. domestica*, and hymenopteran sawfly larvae [68, 86, 62, 90]. Notably, the gill-bearing abdominal segments of *H. limbata* fall within the range of both *Ubx* and *abd-A* expression, indicating they experience limb repression during embryogenesis.

While it remains unknown if *H. limbata Dll* is abdominally expressed, it is possible that mayfly gills are exclusively proximal structures, akin to the prolegs of sawfly larvae, and thus not subject to *Dll* inhibition by *Ubx* or *abd-A*. Such a hypothesis is further supported by previous work that posits apterygote styli and mayfly gills as proximal appendicular structures possibly related to the proximal appendicular structures of crustacean epipodites [3, 91].

Conclusions

Our investigation of *H. limbata* Hox peptides identified eight of the ten canonical Hox genes, and revealed the presence of key functional regions highly conserved with other Hox insect homologs. The expression of embryonic *Antp* also is highly conserved with that of other insects, becoming apparent in the developing thoracic limb buds and subsequently spreading throughout the thoracic and abdominal midline. Similarly, embryonic *Ubx/abd-A* expression closely matches what is observed in other studied insects, particularly non-holometabolan species. The extensive conservation of both *H. limbata* Hox sequences and expression profiles with those of other insects suggests that the unique body plan of mayfly nymphs may not be directly regulated by Hox genes during embryogenesis. Based on previous morphological studies and our current findings, we hypothesize that mayfly abdominal gills are proximal morphological structures. As the gills of *H. limbata* develop in the second nymphal instar [92, 93], Additional files 1: Fig. S1) rather than during or immediately after embryogenesis, studies assessing the expression of *Ubx/abd-A* and *Dll* during first instars, and their overall functional roles during *H. limbata* development, will rigorously test this hypothesis and provide further clarity on the unique evolution of mayflies.

Declarations

Acknowledgements

We thank Dave Angelini (Colby College, ME) and Chris Chabot (Plymouth State University, NH) for constructive feedback on this project. We also would like to thank Adam Stuckert, Jordan S. Ramsdell, and Sydney Birch for their guidance on bioinformatic protocols, as well as

Kelly Thomas and Krystalynne Morris for their assistance with transcriptomic sequencing (all at University of New Hampshire, NH). Additional thanks to Brandon Huckins (Plymouth State University, NH) for collecting many of the mayflies used in expression studies. The antibodies Ubx/ABDA FP6.87, developed by R. White at the University of Cambridge, and Antp 4C3, developed by D. Brower (University of Arizona), were obtained from the Developmental Studies Hybridoma Bank, created by the NICHD of the NIH and maintained at The University of Iowa, Department of Biology, Iowa City, IA 52242. Final thanks to Dave Plachetzki, Jessica Bolker (both at University of New Hampshire, NH) and Taylor Hornig for providing invaluable feedback on the manuscript.

Funding

The *H. limbata* transcriptome was sequenced thanks to funding from two sources at Plymouth State University: a Research Advisory Council (RAC) grant and a Biological Sciences Department graduate student grant, both to CJG.

Author's Contributions

CJG conceived of, analyzed, and drafted this manuscript's contents in fulfillment of their master's degree. BO'D was CJG's primary graduate advisor. TH contributed to ORP transcriptome assembly and annotation. All authors read and approved the final manuscript.

Corresponding Author

Correspondence to Christopher Gonzalez.

Competing Interests

The authors declare no competing interests.

Consent for Publication

Not applicable.

Ethics approval and consent to participate

Studies of insects do not require ethics approval or consent to participate.

Availability of data and materials

H. limbata annotated Hox sequences have been deposited in NCBI's GenBank database under the accession numbers MZ2773605-MZ2773612, while the source sequencing reads are available in NCBI's SRA database via the accession number SRX6489924. ORP and CLC contig assemblies are available in the following Zenodo repository:<https://doi.org/10.5281/zenodo.5308383>. All other data generated or analyzed during this study are included or referenced in this published article and its supplementary information files.

References

1. Giribet G, Edgecombe GD. Reevaluating the arthropod tree of life. *Annual Review of Entomology*. 2012;57(1):167–86.
2. Snodgrass, R. E. General Organization and Development. In: *Principles of Insect Morphology*. New York: McGraw-Hill Book; 1935a. p.14-47.
3. Snodgrass, R. E. The Abdomen. In: *Principles of Insect Morphology*. New York: McGraw-Hill Book; 1935b. p. 246-79.
4. Snodgrass, R. E. The Organs of Ingestion. In: *Principles of Insect Morphology*. New York: McGraw-Hill Book; 1935c. p. 280-346.
5. Angelini DR, Kaufman TC. Insect appendages and comparative ontogenetics. *Developmental Biology*. 2005;286(1):57–77.
6. Bridges CB, Morgan TH. The third-chromosome group of mutant characters of *Drosophila melanogaster*. Carnegie Institution of Washington, publication, no. 327; 1923. p. 1–251.
7. Lewis EB. A gene complex controlling segmentation in *Drosophila*. *Nature*. 1978;276(5688):565–70.
8. Lewis EB. Genes and developmental pathways. *American Zoologist*. 1963 Feb 1:33-56.
9. Sánchez-Herrero E, Vernós I, Marco R, Morata G. Genetic organization of *Drosophila bithorax* complex. *Nature*. 1985 Jan;313(5998):108-13.
10. Kaufman TC, Seeger MA, Olsen G. Molecular and genetic organization of the *Antennapedia* gene complex of *Drosophila melanogaster*. *Advances in genetics*. 1990 Jan 1;27:309-62.
11. Hughes CL, Kaufman TC. Hox genes and the evolution of the arthropod body plan 1. *Evolution & development*. 2002 Nov;4(6):459-99.
12. Averof M, Akam M. Hox genes and the diversification of insect and crustacean body plans. *Nature*. 1995 Aug;376(6539):420-3.
13. Carroll SB. Homeotic genes and the evolution of arthropods and chordates. *Nature*. 1995 Aug;376(6540):479-85.
14. Gellon G, McGinnis W. Shaping animal body plans in development and evolution by modulation of Hox expression patterns. *BioEssays*. 1998 Feb;20(2):116-25.
15. Popadic AL, Abzhanov AR, Rusch DO, Kaufman TC. Understanding the genetic basis of morphological evolution: the role of homeotic genes in the diversification of the arthropod bauplan. *International Journal of Developmental Biology*. 1998 Jan 1;42:453-61.
16. Mahfooz NS, Li H, Popadić A. Differential expression patterns of the hox gene are associated with differential growth of insect hind legs. *Proceedings of the National Academy of Sciences*. 2004 Apr 6;101(14):4877-82.
17. Konopova B, Akam M. The Hox genes *Ultrabithorax* and *abdominal-A* specify three different types of abdominal appendage in the springtail *Orchesella cincta* (Collembola). *EvoDevo*. 2014 Dec;5(1):1-4.
18. Mann RS. The specificity of homeotic gene function. *Bioessays*. 1995 Oct;17(10):855-63.

19. Merabet S, Hudry B, Saadaoui M, Graba Y. Classification of sequence signatures: a guide to Hox protein function. *Bioessays*. 2009 May;31(5):500-11.
20. Tojo KO, Machida RY. Techniques in embryological studies of mayflies (Insecta: Ephemeroptera). *InProc 10th Int Conf Ephem 14th Int Sym Pleco 2003*.
21. Mahfooz N, Turchyn N, Mihajlovic M, Hrycaj S, Popadić A. Ubx regulates differential enlargement and diversification of insect hind legs. *PLoS One*. 2007 Sep 12;2(9):e866.
22. Khila A, Abouheif E, Rowe L. Evolution of a novel appendage ground plan in water striders is driven by changes in the Hox gene *Ultrabithorax*. *PLoS genetics*. 2009 Jul 31;5(7):e1000583.
23. Trautwein MD, Wiegmann BM, Beutel R, Kjer KM, Yeates DK. Advances in insect phylogeny at the dawn of the postgenomic era. *Annual review of entomology*. 2012 Jan 7;57:449-68.
24. Misof B, Liu S, Meusemann K, Peters RS, Donath A, Mayer C, Frandsen PB, Ware J, Flouri T, Beutel RG, Niehuis O. Phylogenomics resolves the timing and pattern of insect evolution. *Science*. 2014 Nov 7;346(6210):763-7.
25. Riek EF. The classification of the Ephemeroptera. In: *Proceedings of the First International Conference on Ephemeroptera*. Leiden: EJ Brill; 1973. p. 160-178.
26. Sartori, M., & Brittain, J. E. Order Ephemeroptera. In: *Thorp and Covich's Freshwater Invertebrates*. 4th Edition. London: Elsevier; 2015. p. 873-891.
27. Morgan AH, Grierson MC. The functions of the gills in burrowing may flies (*Hexagenia recurvata*). *Physiological Zoology*. 1932 Apr 1;5(2):230-45.
28. Wingfield CA. The function of the gills of mayfly nymphs from different habitats. *Journal of Experimental Biology*. 1939 Jul;16(3):363-73.
29. Changfa Z, Leyi Z, Kaiya Z. Morphological diversity of mayfly nymphs and its adaptive derivation. *Chinese Journal of Zoology*. 2003 Jan 1;38(6):81-5.
30. Si Q, Luo JY, Hu Z, Zhang W, Zhou CF. De novo transcriptome of the mayfly *Cloeon viridulum* and transcriptional signatures of *Prometabola*. *PloS one*. 2017 Jun 21;12(6):e0179083.
31. Almudi I, Martín-Blanco CA, García-Fernandez IM, López-Catalina A, Davie K, Aerts S, Casares F. Establishment of the mayfly *Cloeon dipterum* as a new model system to investigate insect evolution. *Evodevo*. 2019 Dec;10(1):1-0.
32. Almudi I, Vizueta J, Wyatt CD, de Mendoza A, Marlétaz F, Firbas PN, Feuda R, Masi-ero G, Medina P, Alcaina-Caro A, Cruz F. Genomic adaptations to aquatic and aerial life in mayflies and the origin of insect wings. *Nature communications*. 2020 May 26;11(1):1-1.
33. Niwa N, Akimoto-Kato A, Niimi T, Tojo K, Machida R, Hayashi S. Evolutionary origin of the insect wing via integration of two developmental modules. *Evolution & development*. 2010 Mar;12(2):168-76.
34. O'Donnell BC, Jockusch EL. The expression of wingless and *Engrailed* in developing embryos of the mayfly *Ephoron leukon* (Ephemeroptera: Polymitarcyidae). *Development genes and evolution*. 2010 Jun;220(1):11-24.

35. McCafferty WP. The burrowing mayflies of the United States (Ephemeroptera: Ephemeridae). *Transactions of the American Entomological Society* (1890-). 1975 Sep 1;101(3):447-504.
36. Fremling CR. Methods for mass-rearing *Hexagenia* mayflies (Ephemeroptera: Ephemeridae). *Transactions of the American Fisheries Society*. 1967 Oct 1;96(4):407-10.
37. Needham JG. Burrowing mayflies of our larger lakes and streams. US Government Printing Office; 1920.
38. Brittain JE. Biology of mayflies. *Annual review of entomology*. 1982 Jan;27(1):119-47.
39. Giberson DJ, Rosenberg DM. Egg development in *Hexagenia limbata* (Ephemeroptera: Ephemeridae) from Southern Indian Lake, Manitoba: temperature effects and diapause. *Journal of the North American Benthological Society*. 1992 Jun 1;11(2):194-203.
40. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature biotechnology*. 2011 Jul;29(7):644-52.
41. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Pribelski AD, Pyshkin AV. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of computational biology*. 2012 May 1;19(5):455-77.
42. Robertson G, Schein J, Chiu R, Corbett R, Field M, Jackman SD, Mungall K, Lee S, Okada HM, Qian JQ, Griffith M. De novo assembly and analysis of RNA-seq data. *Nature methods*. 2010 Nov;7(11):909-12.
43. Emms DM, Kelly S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome biology*. 2015 Dec;16(1):1-4.
44. MacManes MD. The Oyster River Protocol: a multi-assembler and kmer approach for de novo transcriptome assembly. *PeerJ*. 2018 Aug 3;6:e5428.
45. Seppey M, Manni M, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness. In: *Gene prediction*. Humana, New York, NY; 2019. p. 227-245.
46. Smith-Unna R, Boursnell C, Patro R, Hibberd JM, Kelly S. TransRate: reference-free quality assessment of de novo transcriptome assemblies. *Genome research*. 2016 Aug 1;26(8):1134-44.
47. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. *Nature methods*. 2015 Jan;12(1):59-60.
48. Suzek BE, Wang Y, Huang H, McGarvey PB, Wu CH, UniProt Consortium. UniRef clusters: a comprehensive and scalable alternative for improving sequence similarity searches. *Bioinformatics*. 2015 Mar 15;31(6):926-32.
49. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *Journal of molecular biology*. 1990 Oct 5;215(3):403-10.
50. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution*. 2013 Jan 16;30(4):772-80.

51. Katoh K, Rozewicki J, Yamada KD. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Briefings in bioinformatics*. 2019 Jul;20(4):1160-6.
52. Letunic I, Bork P. 20 years of the SMART protein domain annotation resource. *Nucleic acids research*. 2018 Jan 4;46(D1):D493-6.
53. Kim JY, Lim HY, Shin SE, Cha HK, Seo JH, Kim SK, Park SH, Son GH. Comprehensive transcriptome analysis of *Sarcophaga peregrina*, a forensically important fly species. *Scientific data*. 2018 Nov 6;5(1):1-8.
54. Singh S, Gupta M, Pandher S, Kaur G, Goel N, Rathore P. Using de novo transcriptome assembly and analysis to study RNAi in *Phenacoccus solenopsis* Tinsley (Hemiptera: Pseudococcidae). *Scientific reports*. 2019 Sep 23;9(1):1-6.
55. Yuan H, Zhang X, Zhao L, Chang H, Yang C, Qiu Z, Huang Y. Characterization and analysis of full-length transcriptomes from two grasshoppers, *Gomphocerus licenti* and *Mongolotettix japonicus*. *Scientific reports*. 2020 Aug 26;10(1):1-2.
56. Tassone EE, Cowden CC, Castle SJ. De novo transcriptome assemblies of four xylem sap-feeding insects. *GigaScience*. 2017 Mar;6(3):giw007.
57. Shah A, Hoffman JI, Schielzeth H. Transcriptome assembly for a colour-polymorphic grasshopper (*Gomphocerus sibiricus*) with a very large genome size. *BMC genomics*. 2019 Dec;20(1):1-0.
58. Carroll SB, Laymon RA, McCutcheon MA, Riley PD, Scott MP. The localization and regulation of Antennapedia protein expression in *Drosophila* embryos. *Cell*. 1986 Oct 10;47(1):113-22.
59. Hayward DC, Patel NH, Rehm EJ, Goodman CS, Ball EE. Sequence and Expression of Grasshopper Antennapedia: Comparison to *Drosophila*. *Developmental biology*. 1995 Dec 1;172(2):452-65.
60. Levine M, Hafen E, Garber RL, Gehring WJ. Spatial distribution of Antennapedia transcripts during *Drosophila* development. *The EMBO journal*. 1983 Nov;2(11):2037-46.
61. Walldorf U, Binner P, Fleig R. Hox genes in the honey bee *Apis mellifera*. *Development genes and evolution*. 2000 Sep 1;210(10):483-92.
62. Zhang H, Shinmyo Y, Mito T, Miyawaki K, Sarashina I, Ohuchi H, Noji S. Expression patterns of the homeotic genes *Scr*, *Antp*, *Ubx*, and *abd-A* during embryogenesis of the cricket *Gryllus bimaculatus*. *Gene expression patterns*. 2005 Apr 1;5(4):491-502.
63. White RA, Wilcox M. Distribution of Ultrabithorax proteins in *Drosophila*. *The EMBO journal*. 1985 Aug;4(8):2035-43.
64. Akam ME, Martinez-Arias A. The distribution of Ultrabithorax transcripts in *Drosophila* embryos. *The EMBO journal*. 1985 Jul;4(7):1689-700.
65. Karch F, Bender W, Weiffenbach B. *abdA* expression in *Drosophila* embryos. *Genes & development*. 1990 Sep 1;4(9):1573-87.
66. Macías A, Casanova JO, Morata G. Expression and regulation of the *abd-A* gene of *Drosophila*. *Development*. 1990 Dec 1;110(4):1197-207.

67. Tear G, Akam MI, Martinez-Arias AL. Isolation of an abdominal-A gene from the locust *Schistocerca gregaria* and its expression during early embryogenesis. *Development*. 1990 Nov 1;110(3):915-25.
68. Peterson MD, Rogers BT, Popadić A, Kaufman TC. The embryonic expression pattern of labial, posterior homeotic complex genes and the teashirt homologue in an apterygote insect. *Development genes and evolution*. 1999 Feb 1;209(2):77-90.
69. Tour E, Hittinger CT, McGinnis W. Evolutionarily conserved domains required for activation and repression functions of the *Drosophila* Hox protein Ultrabithorax. *Development*. 2005;132(23):5271–81.
70. Chang CP, Shen WF, Rozenfeld S, Lawrence HJ, Largman C, Cleary ML. Pbx proteins display hexapeptide-dependent cooperative DNA binding with a subset of Hox proteins. *Genes & development*. 1995 Mar 15;9(6):663-74.
71. In der Rieden PM, Mainguy G, Woltering JM, Durston AJ. Homeodomain to hexapeptide or PBC-interaction-domain distance: size apparently matters. *Trends in Genetics*. 2004 Feb 1;20(2):76-9.
72. Ekker SC, Jackson DG, Von Kessler DP, Sun BI, Young KE, Beachy PA. The degree of variation in DNA sequence recognition among four *Drosophila* homeotic proteins. *The EMBO journal*. 1994 Aug;13(15):3551-60.
73. Ronshaugen M, McGinnis N, McGinnis W. Hox protein mutation and macroevolution of the insect body plan. *Nature*. 2002 Feb;415(6874):914-7.
74. Galant R, Carroll SB. Evolution of a transcriptional repression domain in an insect Hox protein. *Nature*. 2002 Feb;415(6874):910-3.
75. Merabet S, Saadaoui M, Sambrani N, Hudry B, Pradel J, Affolter M, Graba Y. A unique Extradenticle recruitment mode in the *Drosophila* Hox protein Ultrabithorax. *Proceedings of the National Academy of Sciences*. 2007 Oct 23;104(43):16946-51.
76. Merabet S, Kambris Z, Capovilla M, Bérenger H, Pradel J, Graba Y. The hexapeptide and linker regions of the AbdA Hox protein regulate its activating and repressive functions. *Developmental cell*. 2003 May 1;4(5):761-8.
77. Lelli KM, Noro B, Mann RS. Variable motif utilization in homeotic selector (Hox)–cofactor complex formation controls specificity. *Proceedings of the National Academy of Sciences*. 2011 Dec 27;108(52):21122-7.
78. Struhl G. Genes controlling segmental specification in the *Drosophila* thorax. *Proceedings of the National Academy of Sciences*. 1982 Dec 1;79(23):7380-4.
79. Nagata T, Suzuki Y, Ueno K, Kokubo H, Xu X, Hui CC, Hara W, Fukuta M. Developmental expression of the *Bombyx* Antennapedia homologue and homeotic changes in the Nc mutant. *Genes to cells*. 1996 Jun;1(6):555-68.
80. Abzhanov A, Kaufman TC. Crustacean (malacostracan) Hox genes and the evolution of the arthropod trunk. *Development*. 2000 Jun 1;127(11):2239-49.
81. Vachon G, Cohen B, Pfeifle C, McGuffin ME, Botas J, Cohen SM. Homeotic genes of the Bithorax complex repress limb development in the abdomen of the *Drosophila* embryo through the target

- gene *Distal-less*. *Cell*. 1992 Oct 30;71(3):437-50.
82. Cohen SM, Jürgens G. Proximal–distal pattern formation in *Drosophila*: cell autonomous requirement for *Distal-less* gene activity in limb development. *The EMBO journal*. 1989 Jul;8(7):2045-55.
83. Lecuit T, Cohen SM. Proximal–distal axis formation in the *Drosophila* leg. *Nature*. 1997 Jul;388(6638):139-45.
84. Warren RW, Nagy L, Selegue J, Gates J, Carroll S. Evolution of homeotic gene regulation and function in flies and butterflies. *Nature*. 1994 Dec;372(6505):458-61.
85. Zheng Z, Khoo A, Fambrough Jr D, Garza L, Booker R. Homeotic gene expression in the wild-type and a homeotic mutant of the moth *Manduca sexta*. *Development genes and evolution*. 1999 Jul 1;209(8):460-72.
86. Suzuki Y, Palopoli MF. Evolution of insect abdominal appendages: are prolegs homologous or convergent traits?. *Development genes and evolution*. 2001 Oct;211(10):486-92.
87. Palopoli MF, Patel NH. Evolution of the interaction between Hox genes and a downstream target. *Current Biology*. 1998 May 7;8(10):587-90.
88. Lewis DL, DeCamillis M, Bennett RL. Distinct roles of the homeotic genes *Ubx* and *abd-A* in beetle embryonic abdominal appendage development. *Proceedings of the National Academy of Sciences*. 2000 Apr 25;97(9):4504-9.
89. Woodland JT. The styli of the firebrat, *Thermobia domestica* (Packard). In: *Proceedings of the Pennsylvania Academy of Science*. Penn State University Press; 1952. Jan 1;26:27-32.
90. Oka K, Yoshiyama N, Tojo K, Machida R, Hatakeyama M. Characterization of abdominal appendages in the sawfly, *Athalia rosae* (Hymenoptera), by morphological and gene expression analyses. *Development genes and evolution*. 2010 Jun 1;220(1-2):53-9.
91. Averof M, Cohen SM. Evolutionary origin of insect wings from ancestral gills. *Nature*. 1997 Feb;385(6617):627-30.
92. Wiebe AH. The first three larval stages of *Hexagenia bilineata* Say. *Ohio Journal of Science*. 1926 Sept;26(5):267-275.
93. Neave F. A study of the May flies (*Hexagenia*) of Lake Winnipeg. *Contributions to Canadian Biology and Fisheries*. 1932 Jan 1;7(1):177-201.

Figures

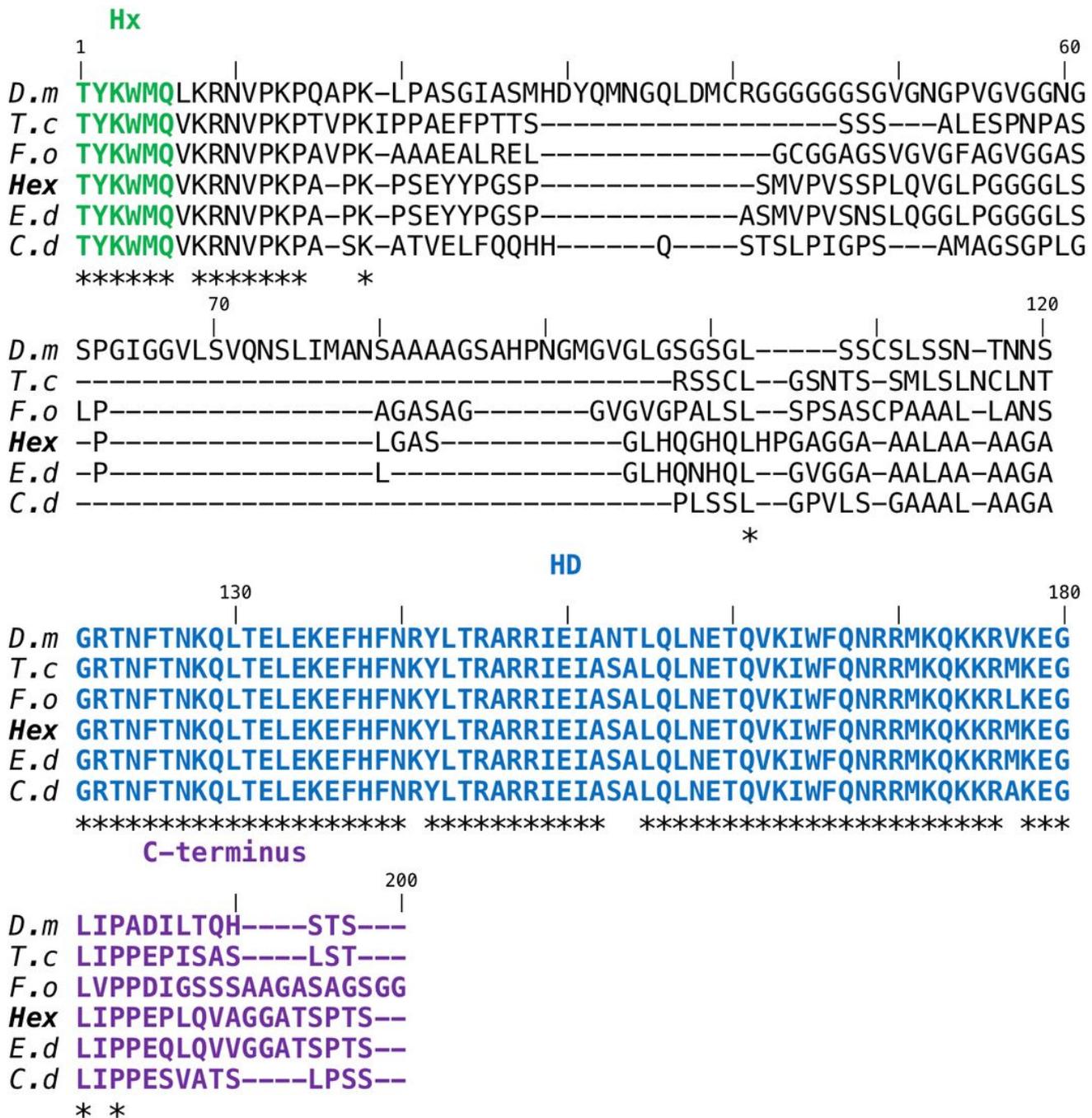


Figure 1

Amino acid alignment of putative *H. limbata* (*Hex*) labial with homologs from five additional hex-apod species. Sequence annotations marked in bold, as follows: Hx, Hexapeptide motif (green); HD, Homeodomain (blue); C-terminus (purple); linker region is in black and not boldfaced. Dashes indicate gaps in the sequence, asterisks label identical residues. Species abbreviations are as follows: *D.m.*, *Drosophila melanogaster*; *T.c.*, *Tribolium castaneum*; *F.o.*, *Frankliniella occidentalis*; *E.d.*, *Ephemera danica*; *C.d.*, *Cloeon dipterum*. All sequence accession numbers are reported in Additional file 1: Table S1.

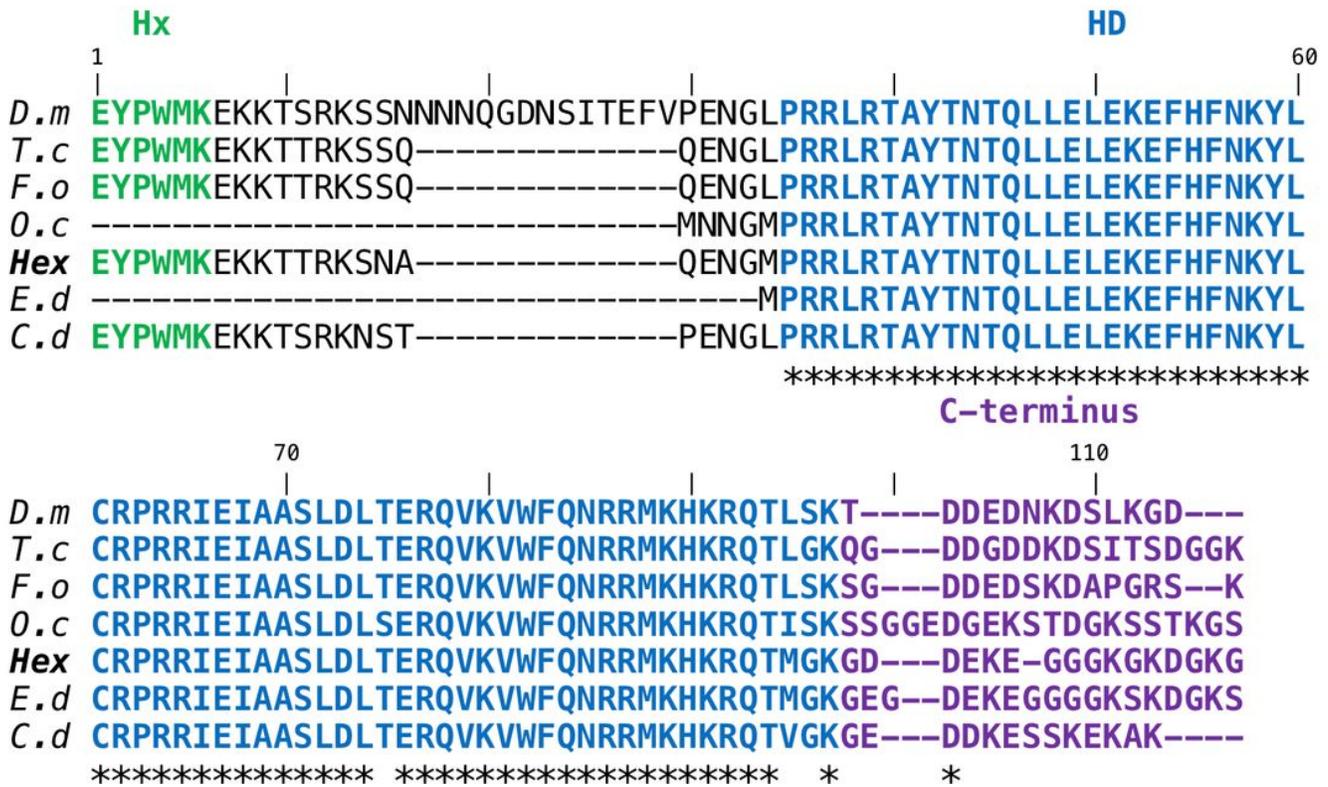


Figure 2

Amino acid alignment of putative *H. limbata* (*Hex*) Proboscopedia with homologs from six additional hexapod species. Sequence annotations marked in bold, as follows: Hx, Hexapeptide motif (green); HD, Homeodomain (blue); C-terminus (purple); linker region is in black and not boldfaced. Dashes indicate gaps in the sequence, asterisks label identical residues. Species abbreviations are as follows: *D.m.*, *Drosophila melanogaster*; *T.c.*, *Tribolium castaneum*; *F.o.*, *Frankliniella occidentalis*; *O.c.*, *Orchesella cincta*; *E.d.*, *Ephemera danica*; *C.d.*, *Cloeon dipterum*. All sequence accession numbers are reported in Additional file 1: Table S1.

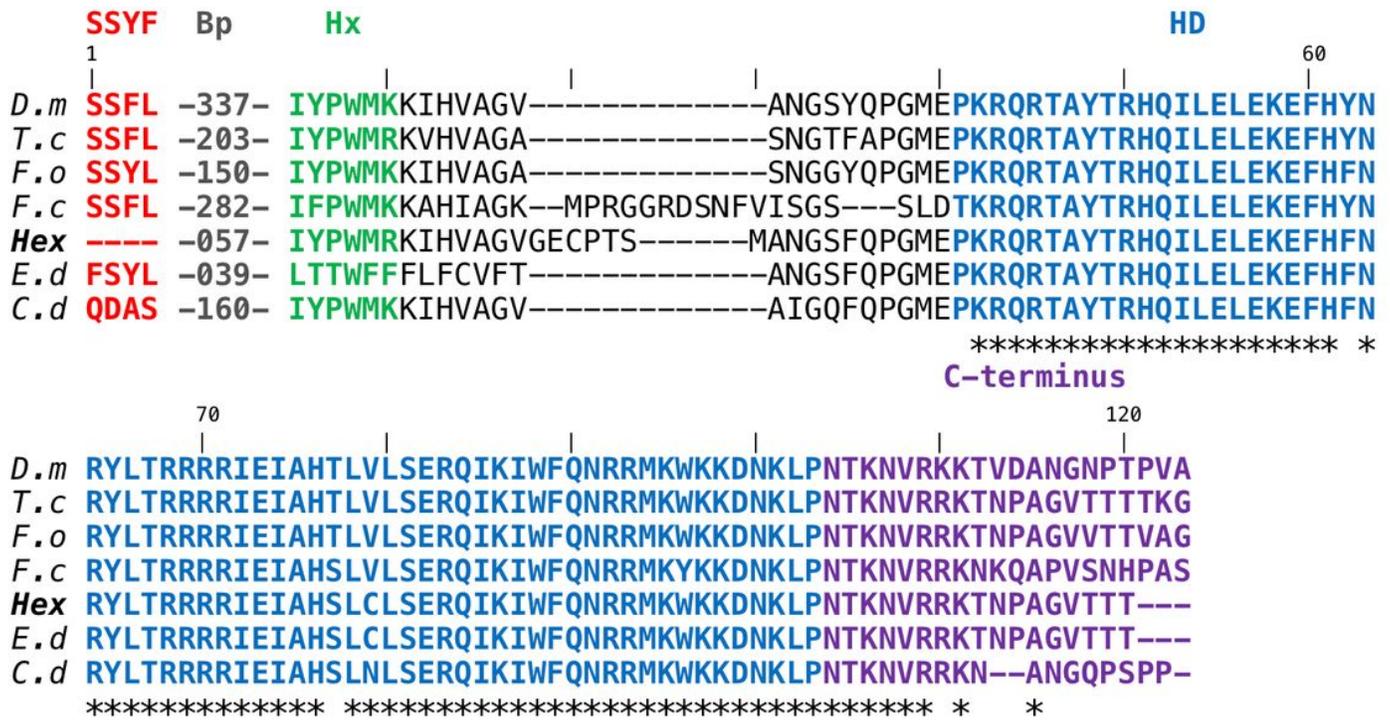


Figure 3

Amino acid alignment of putative *H. limbata* (Hex) Deformed with homologs from six additional hexapod species. Sequence annotations marked in bold, as follows: SSYF motif (red); Bp, count of trimmed base pairs (in grey); Hx, Hexapeptide motif (green); HD, Homeodomain (blue); C-terminus (purple); linker region is in black and not boldfaced. Dashes indicate gaps in the sequence, asterisks label identical residues. Species abbreviations are as follows: *D.m.*, *Drosophila melanogaster*; *T.c.*, *Tribolium castaneum*; *F.o.*, *Frankliniella occidentalis*; *F.c.*, *Folsomia candida*; *E.d.*, *Ephemera dan-ica*; *C.d.*, *Cloeon dipterum*. All sequence accession numbers are reported in Additional file 1: Table S1.

castaneum; F.o., Frankliniella occidentalis; F.c., Folsomia candida; E.d., Ephemera danica; C.d., Cloeon dipterum. All sequence accession numbers are reported in Additional file 1: Table S1.

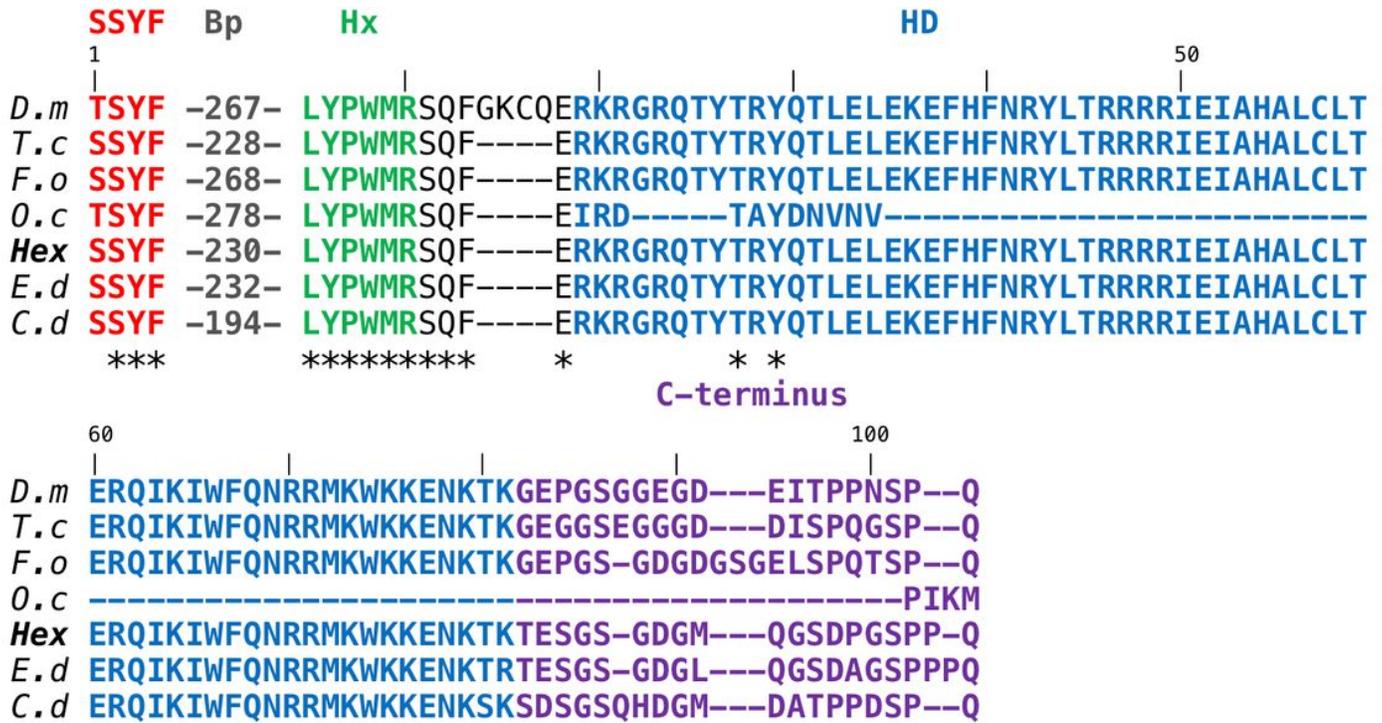


Figure 5

Amino acid alignment of putative *H. limbata* (Hex) Antennapedia with homologs from six additional hexapod species. Sequence annotations marked in bold, as follows: SSYF motif (red); Bp, count of trimmed base pairs (grey); Hx, Hexapeptide motif (green); HD, Homeodomain (blue); C-terminus (purple); linker region is in black and not boldfaced. Dashes indicate gaps in the sequence, asterisks label identical residues. Species abbreviations are as follows: D.m., *Drosophila melanogaster*; T.c., *Tribolium castaneum*; F.o., *Frankliniella occidentalis*; O.c., *Orchesella cincta*; E.d., *Ephemera danica*; C.d., *Cloeon dipterum*. All sequence accession numbers are reported in Additional file 1: Table S1.

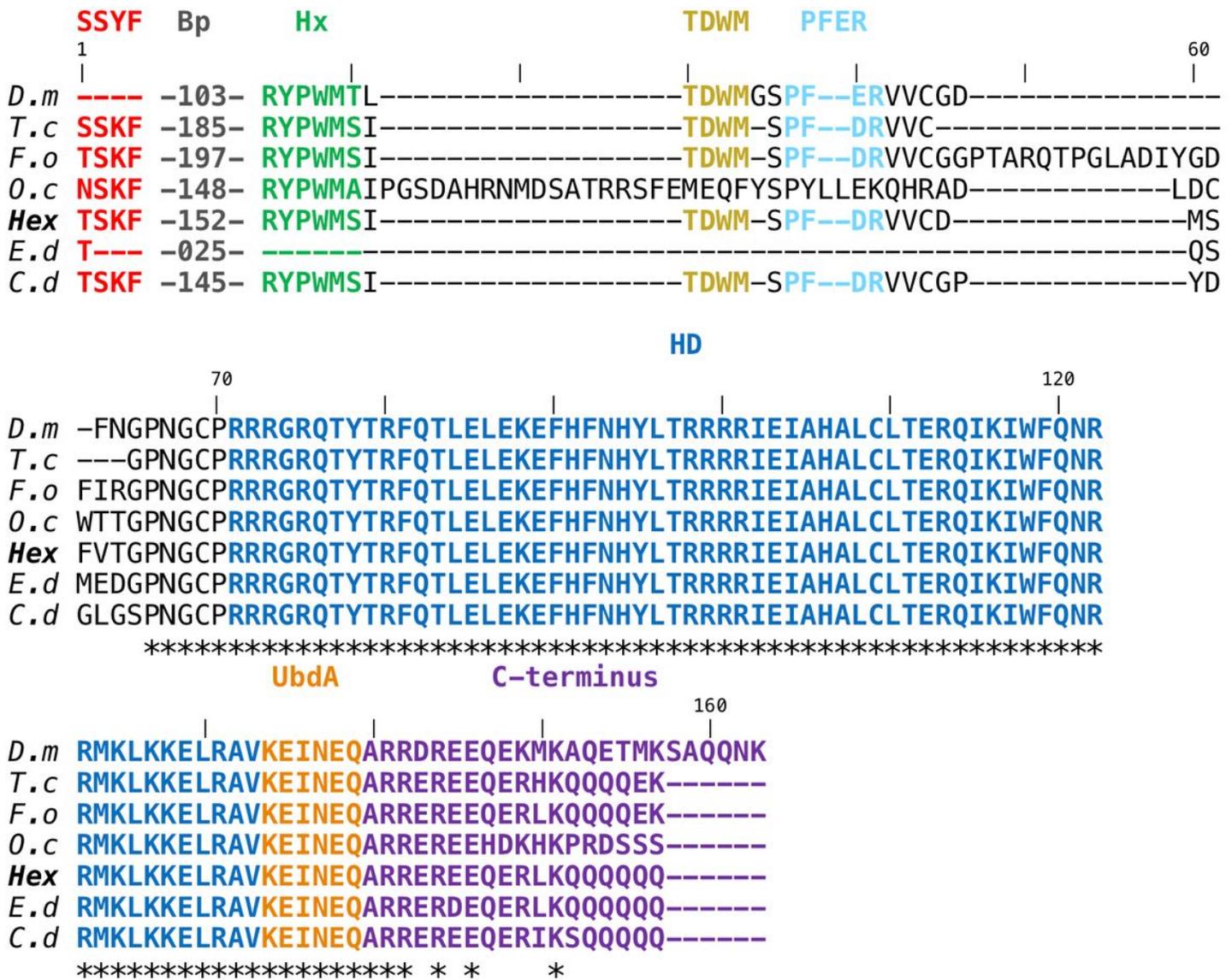


Figure 7

Amino acid alignment of putative *H. limbata* abdominal-A with homologs from six additional hex-apod species. Sequence annotations marked in bold, as follows: SSYF motif (red); Bp, count of trimmed base pairs (grey); Hx, Hexapeptide motif (green); PFER motif (light blue); HD, Homeodo-main (blue); UbdA motif (orange); C-terminus (purple); linker region is in black and not boldfaced. Dashes indicate gaps in the sequence, asterisks label identical residues. Species abbreviations are as follows: *D.m.*, *Drosophila melanogaster*; *T.c.*, *Tribolium castaneum*; *F.o.*, *Frankliniella occidentalis*; *O.c.*, *Orchesella cincta*; *E.d.*, *Ephemera danica*; *C.d.*, *Cloeon dipterum*. All sequence accession numbers are reported in Additional file 1: Table S1.

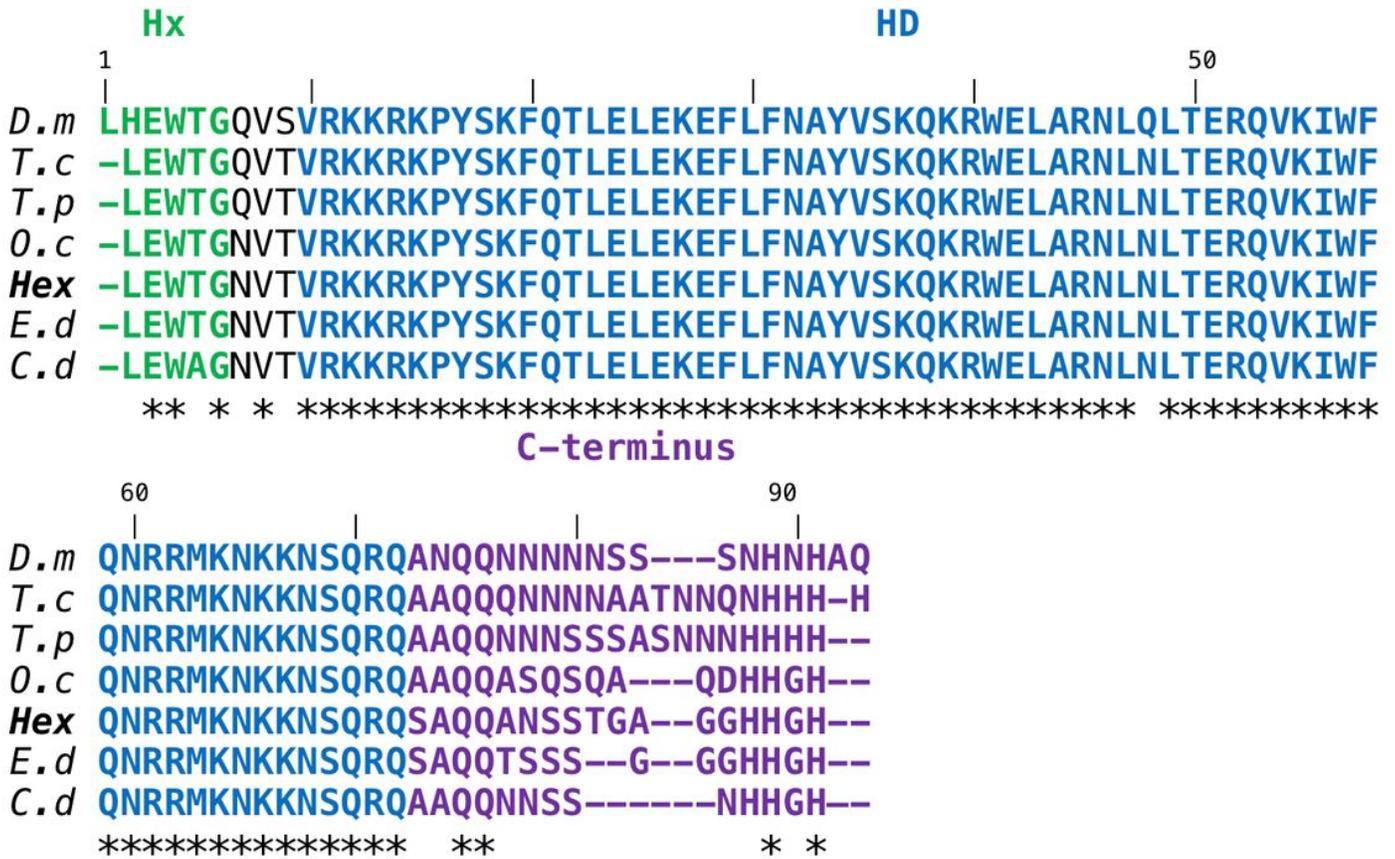


Figure 8

Amino acid alignment of putative *H. limbata* (*Hex*) Abdominal-B with homologs from six additional hexapod species. Sequence annotations marked in bold, as follows: Hx Region, Hexapeptide region (green); HD, Homeodomain (blue); C-terminus (purple). Dashes indicate gaps in the sequence, asterisks label identical residues. Species abbreviations are as follows: *D.m.*, *Drosophila melanogaster*; *T.c.*, *Tribolium castaneum*; *T.p.*, *Thrips palmi*; *O.c.*, *Orchesella cincta*; *E.d.*, *Ephemera danica*; *C.d.*, *Cloeon dipterum*. All sequence accession numbers are reported in Additional file 1: Table S1.

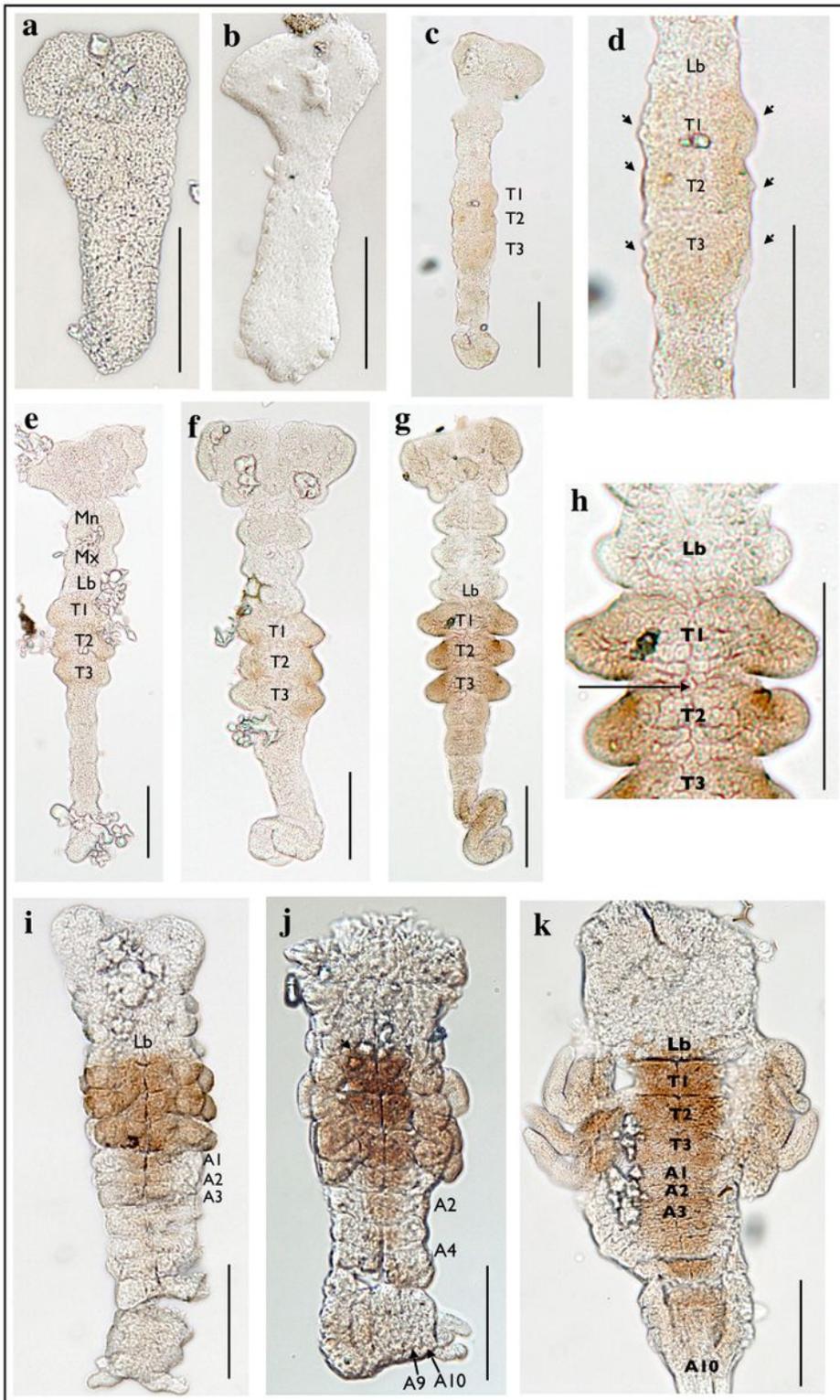


Figure 9

Horseradish peroxidase staining of *H. limbata* embryos using the Antp (4C3) antibody. Early embryos do not show any expression (a, b). As the embryo adds segments to the posterior, expression is present in the three thoracic segments, where thoracic limb buds will eventually protrude (c, d). As the limb buds elongate, expression extends into the buds (f), eventually appearing as patches of strong expression at the anterior of T2 and T3 (arrowheads in g, h), while expression at the midline of each segment is less

strong (g, h). At this time, expression also begins to faintly appear on the posterior edge of the labial segment (g, h). After segmentation concludes, strong expression in the thorax and thoracic limbs is present, weak expression is still evident at the posterior edge of labial segment, and expression is seen in the center of the A1-A3 segments (i, j). In the oldest embryos we imaged, expression in the abdomen becomes stronger extends up to the A10 segment, but is absent from the lateral edges (k). Expression at this stage extends from the posterior of the labial segment through the A10 segment (k). Mn, mandible; Mx, maxilla; Lb, labium; T1-T3, thoracic segments; A1-A10, abdominal segments. Ventral view, anterior to the top in all panels. Scale bars 0.10mm. Image magnifications are 200X for a-b and i-k, 100X for c and e-g.

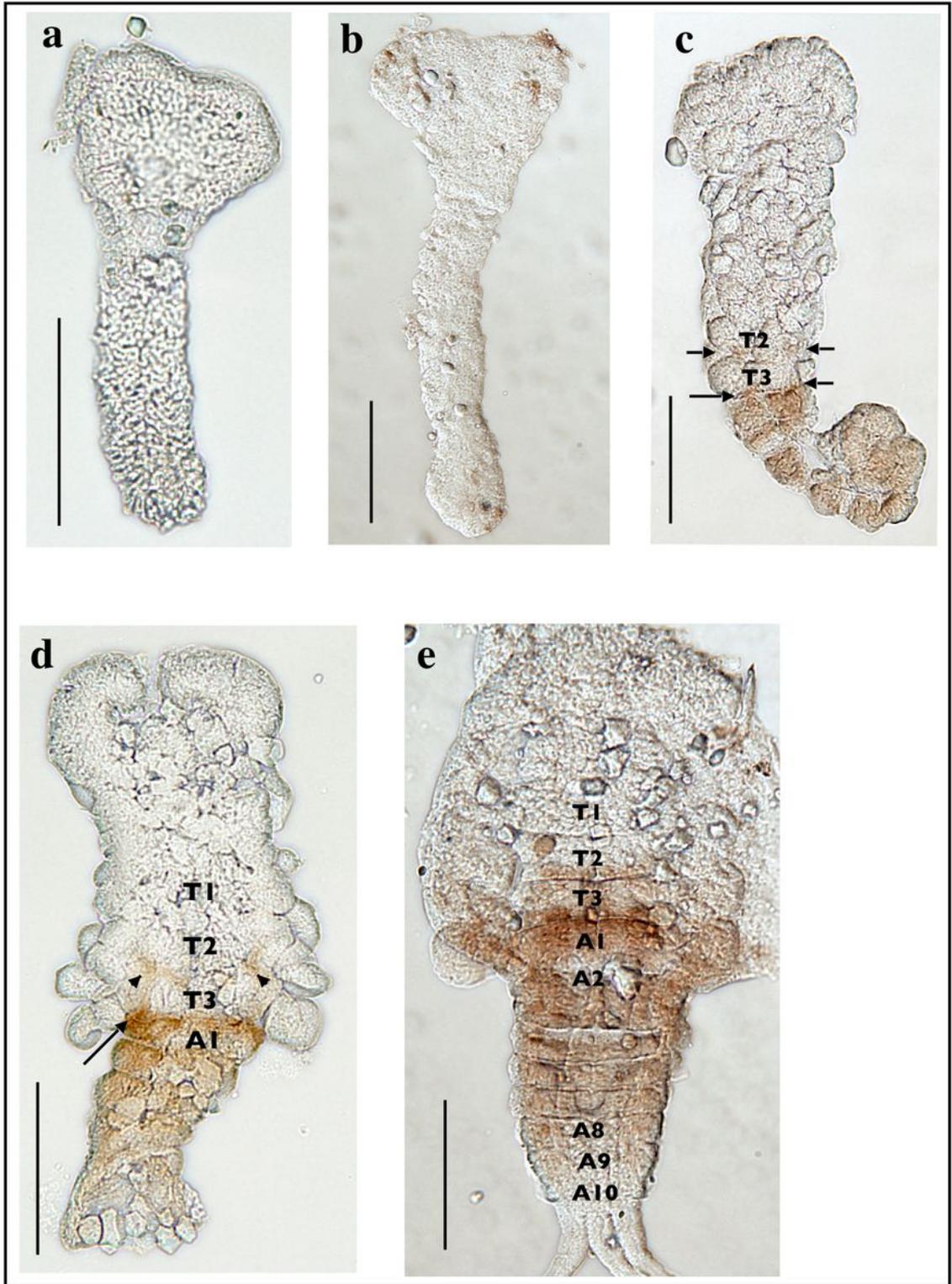


Figure 10

Horseradish peroxidase staining of *H. limbata* embryos using the Ubx/ABD-A (FP6.87) antibody. Early in embryogenesis, expression was not visible (a, b). The first apparent expression was observed as small, centrally located patches in the posterior of the T2 and T3 segments (arrowheads in c). Abdominal expression was stronger and extended through all abdominal segments until A10, with the strongest staining found in the A1 segment (arrow in c). Thoracic staining was weak but apparent at the

posteriolateral border of T2 (arrows in d), with stronger expression in the lateral regions as compared to the midline. In addition, expression was strong at the posterior border of T3 and the anterior border of A1 (d). In the latest stage embryos we imaged (e), expression was present along the posterior border of T2 and T3 (e), with the strongest expression at the posterior of T3 and anterior of A1, but also extended through the more posterior abdominal segments, with a reduction in intensity (e). T1-T3, thoracic segments; A1-A10, abdominal segments. Ventral view, anterior to the top for all images. Scale bars 0.10mm. Magnifications are 200X for all images.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Additionalfile1.pdf](#)