The functional evolution of collembolan Ubx on the regulation of abdominal appendage formation

Yan Liang[1, 2, 3, *], Yun-Xia Luan[1, 4, *]

[1] Key Laboratory of Insect Developmental and Evolutionary Biology, CAS Center for Excellence in Molecular Plant Sciences, Shanghai Institute of Plant Physiology and Ecology, Chinese Academy of Sciences, Shanghai 200032, China

[2] University of Chinese Academy of Sciences, Beijing 100049, China

[3] Current address: Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge CB10 1SA, United Kingdom

[4] Current address: Guangdong Provincial Key Laboratory of Insect Developmental Biology and Applied Technology, Guangzhou Key Laboratory of Insect Development Regulation and Application Research, Institute of Insect Science and Technology, School of Life Sciences, South China Normal University, Guangzhou 510631, China

[*]CORRESPONDENCE TO: yl10@sanger.ac.uk; yxluan@scnu.ac.cn

Alternative Running Title: Analysis of developmental transcriptomes of *Folsomia candida* provides insights into the formation of collembolan appendages and the functional evolution of Ubx

Keywords: Collembola, Arthropoda, appendage formation, RNA-seq, *Ubx*, transcriptional regulation, functional evolution

Abstract

Folsomia candida is a tiny soil-living arthropod belonging to the Collembola, which represents one of the basal hexapods and serves as an outgroup lineage to Insecta. It possesses three abdominal appendages: the ventral tube located in the first abdominal segment (A1), the retinaculum in A3 and the furca in A4, which are morphologically distinctive to the thorax limbs. Collembolan Ubx and AbdA are responsible for specifying their abdominal appendages; however, they are unable to repress the expression of the appendage marker gene *Dll*. To date, the genes involved in appendage formation during collembolan embryogenesis and the mechanisms by which Ubx and AbdA regulate *Dll* transcription and subsequently regulate appendage development remain unknown. In this study, we analysed the developmental transcriptomes of *F. candida* and identified 36 candidate genes that may be involved in appendage formation, including *Ubx* (*FcUbx*). Furthermore, via electrophoretic mobility shift assays and dual luciferase assays, we found that the binding and repression capacity of FcUbx on *Drosophila Dll* resembles those of the longest isoform of *Drosophila* Ubx (DmUbx_Ib), while the regulatory mechanism of the C-terminus of FcUbx on *Dll* repression is similar to that of crustacean *Artemia franciscana* Ubx (AfUbx), demonstrating that the function of collembolan Ubx is intermediate between that of Insecta and Crustacea. Overall, our study provides novel insights into appendage formation and reconstruction of the functional evolutionary trajectory of Ubx on abdominal appendages from crustaceans to insects.

Introduction

Extant arthropods are traditionally classified into four major groups: chelicerates, myriapods, crustaceans, and hexapods (including proturans, collembolans, diplurans, and insects) (Budd & Telford, 2009; Giribet et al., 2001; Hughes & Kaufman, 2002b). Throughout their evolution, arthropods gradually lose abdominal appendages, particularly insects, which entirely lose these appendages in their adult stages (Jockusch & Smith, 2015; Matsuda, 2017). The establishment of a body plan is regulated by a cascade of orchestrated transcription factors during embryogenesis, including maternal effect genes, pair-rule genes, gap genes and Hox genes (Peel et al., 2005). Hox genes encode transcription factors characterized by a helix-turn-helix DNA-binding homeodomain and play pivotal roles in the identification and appendage formation of each segment along the body axis of arthropods (Angelini & Kaufman, 2005a, 2005b; Budd & Telford, 2009; Hughes & Kaufman, 2002a). Typically, Hox genes are organized in a cluster on the chromosome and exhibit both spatial and temporal collinearity in their expression patterns (Durston et al., 2011). Within the group of pancrustaceans (including crustaceans and hexapods), *Dll* (*Distal-less*) serves as a marker gene for the appendage primordium (Abzhanov & Kaufman, 2000; Cohen et al., 1991; Jockusch et al., 2004; Panganiban et al., 1997) and acts as a downstream target of Ubx (*Ultrabithorax*) (Buffry et al., 2023; Cohen, 1990; Galant & Carroll, 2002; Gebelein et al., 2002; Palopoli & Patel, 1998). In *Drosophila melanogaster*, the Ubx gene generates six isoforms (*Ib*, *Ia*, *IIb*, *IIa*, *IVb*, *IVa*) via alternative splicing events, with variation in the architecture of the linker region between the YPWM motif and homeodomain (HD domain) (Geyer et al., 2015; Passner et al., 1999; Reed et al., 2010). Specifically, the *Ib* isoform (DmUbx_Ib) has the longest linker region, and the *IVa* isoform (DmUbx_IVa) does not contain the linker. Functional assays demonstrated that the longest linker is needed for the suppression of *Dll* expression (Gebelein et al., 2002); in addition, the repressive domain(s) located at its C-terminus (Galant & Carroll, 2002; Gebelein et al., 2002; Hughes & Kaufman, 2002b; Ronshaugen et al., 2002). These structures facilitate

51 its ability to bind to the enhancer of *Dll* (Dll304) and consequently suppress the
52 expression of the *Dll* gene (Cohen et al., 1993; Gebelein et al., 2002). In contrast, in
53 crustaceans such as *Artemia franciscana*, Ubx (AfUbx) loses its capacity to repress *Dll*
54 expression (Ronshaugen et al., 2002). This functional shift can be attributed to the
55 presence of phosphorylation sites within the C-terminus of AfUbx, which impede its
56 ability to suppress limb development (Galant & Carroll, 2002; Ronshaugen et al., 2002).
57 To date, the evolutionary processes and the molecular mechanism that give rise to the
58 loss of abdominal appendages in adult hexapods remain elusive.
59 Collembola (springtails) is a group of basal hexapods, with phylogenetic positions
60 bridging by aquatic crustaceans and terrestrial insects (Gao et al., 2008; Luan et al.,
61 2005; Timmermans et al., 2008). Collembolans bear three distinct types of abdominal
62 appendages: the ventral tube in abdominal segment 1 (A1), the retinaculum in A3, and
63 the furca in A4 (Tully & Potapov, 2015). A previous study demonstrated that the
64 collembolan Ubx specifies A1 and AbdA determines A4, and they both regulate the
65 morphological formation of A2 and A3 (Konopova & Akam, 2014). Intriguingly,
66 collembolan Ubx, AbdA and *Dll* are concurrently expressed in the collembolan
67 abdomen during the embryonic stage, indicating that collembolan Ubx and AbdA
68 cannot repress *Dll* expression in abdominal segments (Palopoli & Patel, 1998).
69 However, the molecular mechanisms of how collembolan Ubx and AbdA interact with
70 *Dll* and thus regulate appendage formation are unclear.
71 To address this question, we conducted an in-depth analysis of the developmental
72 transcriptomes of Collembola, *F. candida* (Denmark strain), covering embryos and
73 juveniles to adults. Through this analysis, we identified the genes potentially involved
74 in appendage formation during embryogenesis, with *Ubx* being a notable inclusion.
75 Furthermore, we investigated the transcriptional regulatory functions of Ubx, and our
76 findings revealed that the binding capacity and repression activity of *F. candida* Ubx
77 (FcUbx) on *D. melanogaster Dll* are analogous to those in *D. melanogaster* Ubx
78 (DmUbx_Ib) (Galant & Carroll, 2002; Gebelein et al., 2002; Ronshaugen et al., 2002).

79 The regulatory mode and site(s) in the C-terminus of FcUbx resemble those of Ubx in
80 the crustacean *A. franciscana* (AfUbx) (Galant & Carroll, 2002; Ronshaugen et al.,
81 2002). Based on these data, we depicted a molecular functional evolutionary model of
82 Ubx from crustaceans to basal hexapods to insects in the broad scope of panarthropods
83 and proposed that the collembolan Ubx might exert its repression function in distinct
84 abdominal segments in a context-specific manner.
85
86 **Results**
87 **Development and time-course transcriptomes of *F. candida***
88 The embryos of *F. candida* (Denmark strain) took approximately 10 days to reach the
89 juvenile stage when incubated at 21 °C (Fountain & Hopkin, 2005). Referring to the
90 observation of the embryonic developmental stages in *F. candida* (Shanghai strain)
91 (Gao et al., 2006), the samples selected for sequencing are illustrated in Figure 1A. The
92 embryos in the 0-0.5 day (E_0.5d) predominantly ranged from the 4-cell stage to the
93 blastula stage. By 1.5 days, most embryos had progressed in the gastrula stage. In 2.5
94 days of development, the prophase of the tissue differentiation stage commenced,
95 characterized by the segmentation and formation of appendage primordia. At 3.5 days,
96 the appendages began to grow. During the period of 4.5 to 6.5 days, the mid-tissue
97 development stage progressed, accompanied by the growth and maturation of
98 appendages. From 7.5 to 8.5 days, in the anaphase of tissue differentiation, the
99 appendages were fully formed, and the animals were preparing for hatching. Finally, at
100 9.5 days, the animals were actively moving within the chorion (eggshell), and some
101 individuals had already hatched. Typically, it took these juveniles one month to reach
102 sexual maturity as adults (Figure 1A, 1C).
103 The analysis of developmental transcriptomes recaptured the expression of a total of
104 25,803 genes throughout the developmental process (Figure 1B, Supplementary Data
105 1). Transcriptomic analyses elucidated that the relationships of all the developmental
106 samples were consistent with the observation of development (Figure 1A, 1D, 1E, 1F).

107 In general, all the post-hatching samples (juveniles and adults) clustered together,
108 demonstrating that the hatching event acted as a critical developmental transitional
109 event (Figure 1D, 1E, 1F). The embryonic samples were grouped separately: in the
110 early stages E_0.5d and E_1.5d, no appendages were observed, and these two stages
111 were grouped; the embryonic stages E_2.5d and E_3.5d, particularly marked by the
112 segmentation and formation of appendage primordia, were clustered; and the period
113 spanning from E_4.5d to E_9.5d, deemed the development of appendages, formed a
114 distinct group. Specifically, E_4.5d to E_5.5d represent the mid-phase of appendage
115 formation, and E_6.5d to E_9.5d were identified as the maturation of appendage
116 formation (Figure 1D, 1E, 1F).
117
118 **Mining of genes related to appendage development in Collembola**
119 As a group of basal hexapods, Collembola bears three pairs of thoracic legs and
120 appendages in the abdomen (Figure 1C). These abdominal appendages, however,
121 exhibit distinct morphological characteristics from each other (Figure 1C). To identify
122 genes associated with the establishment of appendages, we conducted a Short Time-
123 series Expression Miner analysis (STEM analysis) (Ernst & Bar-Joseph, 2006) to
124 extract relevant genes from the developmental transcriptomes. Based on the embryonic
125 observation (Figure 1A) and transcriptomic analyses (Figure 1D, 1E, 1F), we selected
126 four stages for our analysis: E_1.5d, corresponding to the blastula stage without any
127 appendage structure; E_3.5d, representing the early stage of appendage development,
128 marked by the emergence of the appendage primordia; E_5.5d, indicating mid-
129 appendage development; and E_7.5d, the late stage of appendage development.
130 The number of expressed genes (RPKM > 0) in these four stages was 15,131, 18,693,
131 19,114 and 20,841 genes, respectively (Figure 2A, Supplementary Data 1, 2). The
132 STEM analysis categorized these genes into 50 profiles, among which 12 clusters were
133 significantly statistically enriched (permutation test, $p < 0.001$). Per the morphological
134 observations (Figure 1A), the emergence of appendages is characterized by gradual

135 development, with the transition of gene expression levels from low to high (Figure
136 2B). Consequently, Cluster 42, which reflected this trend, was selected for further
137 investigation (Figure 2C). Cluster 42 contained a total of 865 genes and 2,318
138 transcripts, of which 1,229 were annotated by Blast2GO (Supplementary Data 3). These
139 genes were annotated as 10 biological processes of GO terms (Figure 2D) and involved
140 in 32 KEGG pathways (Supplementary Data 3). In particular, 66 transcripts were
141 associated with the developmental process (Figure 2D, Supplementary Data 3), of
142 which 36 transcripts were annotated (Supplementary Data 3). These genes constitute a
143 set of conventional hierarchical developmental genes, including the body patterning
144 gene *Notch* (Kurata et al., 2000; O'Day, 2006); segmentation genes such as *Exd, Hth,*
145 and *Engrailed* (Fujioka et al., 2012; Kobayashi et al., 2003; Peel et al., 2005); and the
146 crucial body plan "selector genes" (Weatherbee & Carroll, 1999), such as the Hox genes
147 *Ubx*, *Antp*, and *Scr*. Interestingly, this cluster also comprised genes relevant to muscle
148 and actin function, *e.g.*, *MYO1E* (an unconventional myosin-like protein) related to
149 myosin (McIntosh & Ostap, 2016), *Lhx1* (LIM homeobox protein) associated with
150 fibroblasts (Costello et al., 2015), and *FBN2* (fibrillin-2-like protein) linked to fibrosis
151 (Mead et al., 2022). In summary, STEM analysis identified Cluster 42, encompassing
152 genes involved in the establishment of appendages. Moreover, the simultaneous
153 expression pattern of the candidate genes strongly implies their potential contribution
154 to developmental processes associated with appendages.
155
156 ***Ubx* showed dominance at the appendage maturation stage in *F. candida***
157 Hox genes encode transcription factors of a helix-turn-helix homeodomain, regulating
158 pattern formation and playing a fundamental role in the developmental control of
159 metazoans (Ruddle et al., 1994). In general, Hox clusters are characterized by two
160 criteria (Akam et al., 1994; Cook et al., 2001; Pace et al., 2016; Pearson et al., 2005):
161 first, they are serially positioned in a cluster on the chromosome, determining the
162 anterior-posterior axis of the body plan in a spatial collinearity manner; second, their

163 expression exhibits temporal collinearity, in which the genes at the 3'/anterior end of
164 the cluster transcriptionally activate first, followed by a progressive initiation of genes
165 towards the 5'/posterior end of the cluster (Monteiro & Ferrier, 2006).
166 According to previously reported genome assemblies (Faddeeva-Vakhrusheva et al.,
167 2017; Luan et al., 2023), the architecture of the Hox cluster in *F. candida* exhibits
168 genomic rearrangement, with *Scr*, *Antp* and *Ubx* inserted into the anterior region of the
169 complex. This organization showed that the Hox genes lack spatial collinearity (Figure
170 3A) (Faddeeva-Vakhrusheva et al., 2017).
171 A previous study indicated that temporal collinearity of the Hox gene is not found in *D.*
172 *melanogaster* (Gaunt, 2015). To investigate the presence of temporal collinearity in *F.*
173 *candida*, we first retrieved the expression profile data of *D. melanogaster* from FlyBase
174 (http://flybase.org) and then compared the expression profiles between these two
175 species. Our expression analysis of Hox genes in *D. melanogaster* revealed a lack of
176 temporal collinearity (Figure 3B), in agreement with the previous study (Gaunt, 2015).
177 Similarly, the Hox genes in *F. candida* did not display temporal collinearity (Figure 3D).
178 For instance, the *pb* gene was transcribed at E_3.5d, occurring later than the initiation
179 of expression of its posterior genes *Dfd* and *Scr*, and the most posterior gene *AbdB* was
180 highly expressed at the early stage of E_4.5d. In *D. melanogaster*, *AbdA* exhibited
181 abundance throughout embryogenesis; notably, in *F. candida*, the rearranged genes *Scr*,
182 *Antp*, and *Ubx* demonstrated dominance, with *Ubx* specifically displaying the highest
183 expression levels at the appendage maturation stage E_7.5d. This observation suggests
184 a robust expression signal, implying a pivotal role for Ubx in the developmental process.
185

186 **The sequence evolution of arthropod Ubx**

187 The regulation of appendage formation and proximal-distal axis (PD axis) patterning
188 has been widely known in arthropods, with key genes such as *Ubx*, *AbdA*, *Exd*, *Hth*,
189 and *Dll* playing crucial roles (Angelini & Kaufman, 2005b). In *D. melanogaster*, Ubx
190 generates six isoforms (*Ib*, *Ia*, *IIb*, *IIa*, *IVb*, *IVa*) through alternative splicing, and the

191 isoforms exhibit variations due to the presence of an optional linker region positioned
192 between the YPWM motif and homeodomain (HD domain) (Geyer et al., 2015; Reed
193 et al., 2010). It has been validated by functional experiments that both the longest linker
194 region and C-terminus of *Drosophila* Ubx (DmUbx_Ib) (Figure 4A) are responsible for
195 *Dll* repression activity (Gebelein et al., 2002; Gebelein et al., 2004; Ronshaugen et al.,
196 2002). However, the mechanisms by which these genes regulate appendage
197 development in basal hexapods remain unknown. To address this question, we cloned
198 and validated the genes *Ubx*, *AbdA*, *Exd*, *Hth*, and *Dll* from *F. candida* (Denmark strain
199 and Shanghai strain). Importantly, we obtained two complete isoforms of *Ubx* from *F.*
200 *candida*, which are produced by alternative splicing and vary at the linker region
201 (FcUbx, isoform 1, with a linker, FcU1; isoform 2, without a linker, FcU2) (Figure 4,
202 Supplementary Data 4). Moreover, to obtain more sequences from basal hexapods, we
203 cloned *Ubx* from the proturan *Sinentomon erythranum* (SeUbx) and dipluran
204 *Campodea augens* (CaUbx). To discern the evolutionary trajectory of sequence features,
205 we performed alignment of Ubx sequences from diverse representatives of
206 panarthropods, including Insecta, Diplura, Collembola, Protura, Crustacea, Myriapoda,
207 Chelicerate, and Onychophora (Figure 4, Supplementary Data 4).
208 As illustrated in the multiple sequence alignment of Ubx across panarthropods (Figure
209 4, Supplementary Data 4), *D. melanogaster* (DmUbx_Ib) exhibits the longest linker
210 region, consisting of more than 40 amino acids. Nevertheless, the Ubx from proturan
211 (SeUbx), collembolan (FcUbx1, OcUbx), and chelicerate (AlUbx) species displays
212 markedly shorter linker regions, typically composed of only 3-5 amino acids. In certain
213 species, this linker region may be entirely absent. Remarkably, the sequences of the C-
214 terminus display evolutionary patterns. Crustaceans, known for possessing abdominal
215 appendages, their Ubx (AfUbx, DmUbx) consistently maintain potential Ser/Thr (S/T)
216 phosphorylation sites (Ronshaugen et al., 2002) (Figure 4). In contrast, the insects,
217 characterized by the loss of abdominal appendages in adult stages, feature a QAQA
218 domain and poly-Ala stretches instead of phosphorylation sites (Galant & Carroll, 2002;

219 Ronshaugen et al., 2002) in their Ubx (DmUbx_Ib, DmUbx_IVa, DdUbx, MdUbx,
220 AmUbx and TcUbx) (Figure 4). Interestingly, in the three groups of basal hexapods,
221 proturans maintain short, cylindrical appendages on the first three abdominal segments
222 (Headrick & Gordh, 2009), diplurans possess a pair of cerci at the last abdominal
223 segments (Headrick & Gordh, 2009), and their Ubx (proturan SeUbx and dipluran
224 CaUbx) display QAQA domains, poly-Ala stretches and S/T sites. Meanwhile, in
225 Collembola, which bears three types of abdominal appendages (Figure 1C), the
226 collembolan Ubx (FcUbx and OcUbx) shows the QAQA domains and a putative
227 phosphorylated site Ser (S) (in FcUbx) but lacks poly-Ala stretches (Figure 4). These
228 sequence variations in the linker and C-terminus regions indicate that these functional
229 domains might play a substantial role in the evolution and regulation of arthropod
230 abdominal appendage formation.
231
232 **Collembolan Ubx can bind to Hox/Exd/Hth DNA binding motifs**
233 In pancrustaceans, Ubx suppresses appendage formation via the repression of the
234 downstream target gene Distall-less (*Dll*), the marker gene of the appendage
235 primordium (Cohen, 1990; Galant & Carroll, 2002; Gebelein et al., 2002; Palopoli &
236 Patel, 1998; Ronshaugen et al., 2002). Generally, Ubx implements this regulation
237 through two key aspects: DNA binding and transcriptional repression (Galant & Carroll,
238 2002; Grenier & Carroll, 2000; Palopoli & Patel, 1998). In Collembola, Ubx and AbdA
239 are concurrent with *Dll* in the abdomen (abdominal segments 1, 3 and 4), indicating
240 that Ubx and AbdA are unable to suppress *Dll* expression (Konopova & Akam, 2014;
241 Palopoli & Patel, 1998). However, the mechanism by which the collembolan Ubx
242 interacts with *Dll* remains unclear.
243 To decipher this underpinning mechanism, we initially conducted the electrical mobile
244 shift assays (EMSAs) to examine the binding capacity of collembolan Ubx to the *Dll*
245 regulatory element of *D. melanogaster* (DMXR, the repression element on Dll304)
246 (Gebelein et al., 2004) (Figure 5A). The proteins of two isoforms of collembolan Ubx

247 (FcUbx1, FcUbx2) can readily bind to DMXR, demonstrating that FcUbxs exhibit
248 binding capability and that their linker region does not affect this binding ability (Figure
249 5B). Furthermore, the polycomb dimer Exd/Hth stimulates Ubx binding to DNA and
250 forms Ubx/Exd/Hth trimeric protein complexes (Figure 5B), indicating that the
251 regulatory complex of collembolan Ubx/Exd/Hth on DNA binding is the same as that
252 of the *Drosophila* complex (Mann et al., 2009).
253 To investigate whether the collembolan Ubx/Exd/Hth could bind to its DNA, we
254 scrutinized approximately 4000 bp of the intergenic genomic region upstream from the
255 first exon of collembolan *Dll* (Supplementary Data 6) to inspect the Hox/Exd/Hth
256 binding sites. Our screening identified a DNA consensus containing the Hox (5'-TATA-
257 3') (Berger et al., 2008; Ekker et al., 1991; Noyes et al., 2008) and Exd (5'-TGAT-3')
258 (Ryoo et al., 1999; Slattery et al., 2011) binding motifs, and we referred to this element
259 as the Putative FcDll Element (PFE) (Figure 5A, Supplementary Data 6). The EMSAs
260 show that FcUbx1 and FcUbx2 cooperatively bind with Exd and Hth on this DNA
261 element (Figure 5C), providing additional evidence that collembolan Ubx is proficient
262 in binding DNA defined as the Hox/Exd binding motif. Nonetheless, owing to the
263 absence of practical genetics tools in collembolans, the validation of whether this DNA
264 element serves as the regulatory region of *Dll* in collembolans remains unattainable.
265 We conclude that the collembolan regulatory trimeric complex Ubx/Exd/Hth can
266 effectively bind to the Hox/Exd/Hth DNA binding motifs, and the lack of *Dll* repression
267 of collembolan Ubx does not appear to be related to any deficiencies in DNA binding
268 capacity. This implies that the mechanisms underlying the absence of repression may
269 stem from other aspects of the regulation progress.
270
271 **The C-terminus of FcUbx contains both repression and regulatory domains**
272 Next, to evaluate the transcriptional activity of Collembolan Ubx on *Dll*, we carried out
273 dual luciferase assays in *Drosophila* S2 cells. This involved assessing the repression
274 capacity of various proteins (Figure 6A, Supplementary Data 5), including the complete,

275 truncated, and chimeric forms of both collembolan and *Drosophila* Ubx proteins on the
276 expression of the firefly luciferase reporter gene under the regulation of the DMXR
277 element (Figure 5A, 6A, Supplementary Data 7).
278 Although previous research reported that collembolan Ubx cannot repress *Dll* (Palopoli
279 & Patel, 1998), surprisingly, our results revealed that the two isoforms of collembolan
280 Ubx (FcU1 and FcU2) were capable of inhibiting the expression of the firefly luciferase
281 reporter gene (Figure 6B). This repression capability is reminiscent of *Drosophila* Ubx
282 (DmUbx_Ib); thus, the linker region is not required for the repression function of
283 collembolan Ubx. In contrast, the longest linker is indispensable for the repression
284 function of *Drosophila* Ubx, as the non-linker isoform of *Drosophila* Ubx (DmUbx_IVa)
285 and the chimeric protein Dm/Fc_L (Linker of *Drosophila* Ubx is replaced by a short
286 linker GQSYL from *F. candida*) significantly lost its repression capability (Figure 6B).
287 Gebelein et al. revealed that the truncated form of the *Drosophila* Ubx C-terminus is
288 partially able to repress *Dll* (Gebelein et al., 2002). In contrast, our repression assay
289 showed that the truncation of FcU1 and FcU2, which lack a C-terminus (FcU1△C and
290 FcU2△C), significantly decreased their repression ability (Figure 6B), indicating that
291 the C-terminus may contain a potential transcriptional inhibitory domain (Figure 4).
292 Unexpectedly, the chimeric protein Dm/Fc_C (C-terminus of *Drosophila* Ubx is
293 replaced by a short C-terminus AKADCKSVY from *F. candida*) exhibited a substantial
294 loss of repression capacity when compared to DmUbx_Ib, suggesting that the C-
295 terminus of collembolan Ubx (AKADCKSVY) probably contains potential regulatory
296 or modification site(s) that are capable of regulating the repression function of
297 DmUbx_Ib (Figure 4, 6B).
298 Collectively, our results indicate that the linker of FcUbx is unnecessary for the
299 repression function, while the C-terminus may exhibit both a repression domain
300 (QAQA domain) and a functional regulatory site (S) (Figure 4). These findings provide
301 functional evidence on the mechanisms of Ubx-mediated gene repression.
302

303 **Putative phosphorylation of Collembolan Ubx**

304 Indeed, genes with similar expression patterns usually participate in the same biological
305 process; thus, the mining of coexpressed genes can help to predict the unknown
306 functions or biological processes in which a candidate gene may be involved (Ala et al.,
307 2008; Bar-Joseph, 2004; Bhar et al., 2013; van Dam et al., 2018). To further infer the
308 potential function or regulation of collembolan Ubx, we performed Spearman
309 correlation ($r > 0.95$, $p < 0.01$) analysis throughout embryogenesis (E_0.5d to E_9.5d)
310 and identified a total of 113 genes with 287 transcripts coexpressed with *Ubx*, including
311 the Hox gene *AbdA*. Blast2GO annotation revealed that these genes are involved in
312 diverse biological processes, encompassing DNA template-dependent transcriptional
313 regulation, animal organ morphogenesis, morphogenesis of the epithelium, and pattern
314 specification processes, indicating that these genes widely participate in morphogenesis
315 and organogenesis during collembolan embryogenesis (Supplementary Data 8).
316 Intriguingly, phosphorylation kinase receptors (*ROR2*, tyrosine kinase transmembrane
317 receptor Ror2) and phosphatases (*Ptp69D*, tyrosine phosphatase 69D; *Ptp99A*, tyrosine
318 phosphatase 99A), which participate in the biological process of protein
319 dephosphorylation, were identified (Supplementary Data 8). Based on the sequence
320 comparison of Ubx and the results of the dual fluorescence reporter experiment
321 described above, FcUbx contains a putative regulatory phosphorylated site (S) in the
322 C-terminus, and we posit that collembolan Ubx may be directly or indirectly subject to
323 the regulation of protein phosphorylation and dephosphorylation during embryogenesis.
324

325 **Discussion**

326 **Genes involved in the development of collembolan appendages**

327 Collembola, as a basal hexapod group, exhibits morphological traits that intermediate
328 between crustaceans and insects (Gao et al., 2008; Luan et al., 2005; Misof et al., 2014).
329 Exploring the evolutionary and developmental mechanisms underlying these
330 phenotypes could provide valuable insights into the terrestrialization of hexapods (van

331 Straalen, 2021). In this study, we employed two distinct data mining approaches to
332 comprehensively identify genes involved in collembolan appendage development.
333 The first approach, the "transcriptome-wide screening strategy," involves selecting
334 specific appendage formation stages to identify genes that follow the corresponding
335 developmental pattern. In addition to extensively studied genes implicated in
336 morphogenesis, such as *Scr*, *Antp*, *Ubx*, *Notch*, *En*, *Exd*, and *Hth* (Table 1), our analysis
337 recognized the genes *ROR1* and *ROR2* (Table 1), which are associated with limb
338 formation in mice (Schwabe et al., 2004; Takeuchi et al., 2000). In particular, our
339 screening identified a LIM gene (Table 1) known for its role in shaping distal structures
340 during the development of thoracic legs in *D. melanogaster*. Mutations or deletions in
341 this gene have been associated with structural abnormalities (Costello et al., 2015; Tsuji
342 et al., 2000).
343 The second approach, the "candidate-gene focusing strategy", is initiated by focusing
344 on the pivotal *Ubx* gene and extracting coexpressed genes during embryogenesis. This
345 analysis not only uncovered the canonical *AbdA* genes (Averof & Patel, 1997; Casares
346 et al., 1996; Konopova & Akam, 2014) but also identified a series of genes related to
347 appendage formation (Supplementary Data 8). *DAAM2* (dishevelled-associated
348 activator of morphogenesis 2) functions as a factor binding and regulating actin, playing
349 a role in organogenesis, axonal growth, and muscle tissue formation in *D. melanogaster*.
350 Deletion of DAAM2 results in abnormalities in actin filament structures (S. Barkó et
351 al., 2010; Prokop et al., 2011). *Lhx1* is essential for establishing the proximal-distal axis
352 (PD axis) of limbs and specifically regulates the formation of the coxa and femur in
353 *Drosophila* (Szilvia Barkó et al., 2010).
354 These two synergistic approaches have elucidated a set of genes that have not
355 undergone extensive investigation in arthropods except for *D. melanogaster*. Our
356 analysis introduces new perspectives and broadens the scope of research directions.
357 However, it is crucial to note that the transcriptomes we examined were derived from
358 whole animals, and the identified genes might also be involved in various biological

359 processes, such as organogenesis and neurogenesis. Further research studies are
360 necessary to verify the accurate functionalities of these genes during collembolan
361 appendage development.
362
363 **The evolution of functional domains in arthropod Ubx**
364 Throughout its evolution, arthropod Ubx progressively acquired the capacity to inhibit
365 *Dll* expression in the abdomen of insect adults, consequently resulting in the loss of
366 abdominal appendages (Jockusch & Smith, 2015). In this study, by integrating
367 functional assays in collembolan Ubx, we reconstructed the evolutionary trajectory of
368 functional domains in Ubx (Table 2): (1) The ancestral Ubx in Panarthropoda
369 (onychophorans and arthropods) exhibited a consistent DNA binding capability (Galant
370 & Carroll, 2002; Gebelein et al., 2004; Ronshaugen et al., 2002). However, their
371 effectiveness in inhibiting the downstream target *Dll* varied among species (Table 2)
372 (Galant & Carroll, 2002; Gebelein et al., 2002; Grenier & Carroll, 2000; Ronshaugen
373 et al., 2002). (2) In Crustacea, the crustacean Ubx (AfUbx) is unable to inhibit *Dll*, and
374 previous research has proposed that potential regulatory phosphorylation sites may be
375 located in the C-terminus of AfUbx (Galant & Carroll, 2002; Ronshaugen et al., 2002).
376 (3) In Hexapoda, *Drosophila* Ubx (DmUbx_Ib) robustly represses *Dll* expression, with
377 particular emphasis on the linker region and the C-terminus (including both the QAQA
378 domain and the poly-Ala stretch), which play pivotal roles in this regulatory process
379 (Galant & Carroll, 2002; Gebelein et al., 2002; Ronshaugen et al., 2002). (4)
380 Remarkably, collembolan Ubx demonstrated the ability to inhibit *Drosophila Dll*
381 expression, mirroring the capability observed in DmUbx_Ib (Figure 6). Nonetheless,
382 (5) unlike *Drosophila* Ubx, the linker region of collembolan Ubx was not deemed
383 essential for this repression (Figure 6). Rather, (6) it appeared that the functional
384 repression domain and potential regulatory phosphorylation sites may be in the C-
385 terminus of collembolan Ubx (Figure 4, 6).
386 On the basis of these results, we summarize and propose a functional mechanism of

387 how the collembolan Ubx suppresses *Dll* transcription that appears to be intermediate
388 between that of crustaceans and insects (Table 2). Moreover, given the evidence of
389 characters (QAQA domain, poly Ala stretch and S/T site) in the C-terminus of dipluran
390 and proturan Ubx (Figure 4), we speculate that the ancestral hexapod Ubx could bind
391 and repress the expression of *Dll*. However, the scarcity of Ubx sequences in basal
392 hexapods limits our understanding. Currently, only two complete collembolan Ubx
393 sequences are publicly available from NCBI, exhibiting inconsistencies in sequence
394 features (Figure 4). With the advancement of genomic sequencing techniques, we
395 anticipate that more Ubx sequences could be extracted from the genomes of basal
396 hexapods. This, in turn, would facilitate the consolidation of sequence features in
397 hexapods and enlighten the exploration of the evolutionary trajectory of functional
398 domains in arthropod Ubx.
399
400 **The regulatory mechanism of collembolan Ubx on collembolan *Dll***
401 It is essential to highlight that the preceding discourse on the functional evolution of
402 Ubx primarily relies on the mechanism through which arthropod Ubx suppresses
403 *Drosophila Dll*, and the *bona fide* mechanism of how arthropod Ubx (*trans-factor*)
404 interacts with its respective regulatory element of *Dll* (*cis-element*) remains
405 underexplored. Previous research revealed that collembolan Ubx and *Dll* are
406 concurrently expressed in the abdominal segments (A1, 3, 4) (Palopoli & Patel, 1998),
407 while *Dll* and appendages are absent in the A2 segment, although the knockout of *Ubx*
408 and *AbdA* did not result in appendages in the A2 segment (Konopova & Akam, 2014).
409 In combination with our functional assays and transcriptomic analyses, we propose a
410 model elucidating the regulation of collembolan Ubx in the abdominal segments: the
411 intrinsic functionality of collembolan Ubx exhibits binding and repression of *Dll*
412 expression uniformly across all segments (Figures 5, 6). Nevertheless, within each of
413 the abdominal segments, collembolan Ubx implements its repression function in a
414 context-specific manner (see the depicted model). In abdominal segments 1 and 3 (A1

415 and A3), Ubx is likely subject to protein phosphorylation, potentially in conjunction
416 with ROR1 or ROR2 (Table 1, Supplementary Data 2, 8). This phosphorylation event
417 leads to the attenuation of its repression capacity on *Dll*, facilitating the development
418 of the ventral tube and retinaculum. We postulate that the distinctive morphology of
419 these abdominal appendages could be influenced by different cofactors, contributing to
420 morphological malformation, such as Lhx1 (Supplementary Data 8) and FBN2 (Table
421 1). Conversely, in abdominal segment 2 (A2), it presumably undergoes
422 dephosphorylation by phosphatases Ptp69D or Ptp99A (Supplementary Data 8). This
423 dephosphorylation allows Ubx to exert its "native" repression function on *Dll*, thus
424 resulting in the absence of appendages in A2.
425 Despite the current paucity of available genetic tools in collembolans, the validation of
426 our model demands prospective experiments and functional genomics. Specifically, the
427 application of state-of-the-art technologies such as CRISPR/CAS (Akinci et al., 2021)
428 to manipulate the Putative FcDll Element and ascertain its role as a transcriptional
429 regulatory element in collembolans. Furthermore, the incorporation of spatial
430 transcriptomics (Marx, 2021) and spatial ATAC-seq (Deng et al., 2022) to
431 comprehensively map the intricate context of transcription factors and the regulatory
432 landscape in each segment. These methodologies will empower us to investigate how
433 Ubx works precisely, provide compelling evidence for holistic reconstruction of its
434 functional evolution, and ultimately shed light on how arthropods lost their abdominal
435 appendages during evolution.
436

437 ## Materials and Methods

438 ### Sample collection and transcriptome sequencing

439 First, *F. candida* (Denmark strain) was synchronized for three generations: some adult
440 individuals (F_0 generation) were transferred to a new Petri dish (10 cm), a container
441 with a layer of a mixture of plaster of Paris and activated charcoal at a ratio of 9:1 by
442 weight (Krogh, 2009). After 24 hours of oviposition, the eggs (F_1 generation) were

443 subsequently transferred to a new Petri dish for one month of culture to reach sexual
444 maturity. The procedure was repeated twice to obtain synchronized adults of the F_3
445 generation. Approximately 300 adults of the F_3 generation were then transferred to three
446 new Petri dishes for 24 hours of oviposition, and all eggs were transferred to a new
447 culture dish. Next, all eggs laid by the F_3 generation were collected and transferred to
448 new culture dishes (5 cm) every 24 hours, except on the last day, when eggs were
449 collected at 12 hours after the adults had been transferred.
450 After 51 days of collection and culture, we obtained 51 samples ranging from day 0.5
451 to day 50.5. Fifteen developmental samples, including eggs (from days 0.5 to 9.5),
452 juveniles (days 12.5, 19.5 and 28.5), and adults (days 31.5 and 45.5), were chosen for
453 transcriptome sequencing. RNA was prepared and extracted separately from 15 samples
454 by using a miRNeasy mini kit (QIAGEN, Germany) according to the manufacturer's
455 instructions. The transcriptomes were sequenced on the BGISEQ-500 platform by BGI
456 (Yuzuki, 2015) (Shenzhen, China). All of the raw reads had been published (Liang et
457 al., 2019).
458

459 **Transcriptome analysis and gene annotation**

460 The reads mapping and differential gene expression analysis were described in a
461 previous study (Liang et al., 2019). Gene annotations were accomplished by running
462 BLAST2GO (Conesa et al., 2005) against the UniProt database (UniProt release
463 2018_02) (Bairoch et al., 2005). The expression profile was obtained from the data
464 matrix representing the expression abundance (RPKM, reads per kilobase of transcript
465 per million fragments mapped) from 15 transcriptomes. The distribution of gene
466 expression was visualized as Ridgeline Plots in 'ggplot2' (Wickham, 2016).
467 Additionally, to examine the relationships among the developmental stages, we
468 exploited principal component analysis, Pearson correlation analysis, and hierarchical
469 clustering analysis on all the genes and visualized them by 'pheatmap' (Kolde, 2012)
470 in RStudio (2023.3.0.386) (RStudio, 2020).

471

472 **Mining of genes involved in appendage formation**

473 The embryonic samples from day 1.5 (E_1.5d, blastula, no appendages), day 3.5
474 (E_3.5d, early stage of appendage formation), day 5.5 (E_5.5d, mid-stage of appendage
475 formation) and day 7.5 (E_7.5d, mature stage of appendage formation) were chosen for
476 mining genes were involved in appendage formation. Short Time-series Expression
477 Miner (STEM) (Ernst & Bar-Joseph, 2006; Ernst et al., 2005) software was used to
478 classify all the clusters according to the abundance of gene expression. Genes in the
479 clusters that were supposed to correlate with appendage formation were sorted out and
480 then annotated by Blast2GO software (Conesa et al., 2005).

481

482 **Expression profiling of Hox genes in *D. melanogaster* and *F. candida***

483 The gene expression profile of *D. melanogaster* was downloaded from flybase
484 (http://ftp.flybase.org/releases/current/precomputed_files/genes/gene_rpkm_matrix_fb
485 _2023_05.tsv.gz) (Brown et al., 2014). The expression values of Hox genes were subset
486 from expression profiles of *D. melanogaster* and *F. candida*. Notably, the *Ubx* in *F.*
487 *candida* was fragmented into five transcripts (XLOC_011518, XLOC_011661,
488 XLOC_011777, XLOC_011778 and XLOC_011779) (Supplementary Data 1). To
489 calculate the expression of *Ubx*, we computed and accumulated the RPKM of those
490 transcripts. The normalization of gene expression was applied in two approaches.
491 Normalized RPKM, the original RPKM was normalized by comparing the minimum
492 and maximum of each gene throughout the embryonic stages to elucidate the expression
493 pattern and the lowest and highest expression stage of a gene. Z-score, the original
494 RPKM was normalized by the z-scale in R (RStudio, 2020), which can reflect the
495 expression pattern of all the selected genes.

496

497 **Mining of genes correlated with *Ubx* from transcriptomes**

498 To predict the putative function or the underlying regulation of Ubx, based on the gene
499 expression profile, we used the Spearman correlation algorithm (Spearman correlation
500 coefficient $r > 0.9$, $p < 0.01$) to mine genes coexpressed with *Ubx* during embryogenesis.
501 The coexpressed genes were extracted and then annotated by Blast2GO of gene
502 ontology terms (GO terms) and KEGG pathway annotation (Conesa et al., 2005;
503 Kanehisa & Goto, 2000).

504

505 **Acquisition of *Ubx*, *Exd* and *Hth* sequences**

506 The complete sequences of two isoforms of *Ubx* in *F. candida* (Collembola) and the
507 partial sequences from *Sinentomon erythranum* (Protura) and *Campodea augens*
508 (Diplora) were cloned by degenerate PCR from the homeodomain, further obtained by
509 5' RACE and 3' RACE. Specifically, two isoforms of FcUbx were validated by using
510 exon-intron and exon-exon junction PCR with overlapping primers spanning the linker
511 region. *Exd* and *Hth* were identified via BLAST (Altschul et al., 1990; McGinnis &
512 Madden, 2004) from transcriptomes and further validated by PCR cloning. All
513 sequences were submitted to NCBI (OR593736, OR593737, OR593738, OR593739,
514 OR604006, OR604007). The sequences of Panarthropod Ubx were aligned by MAFFT
515 (Rozewicki et al., 2019), and the gaps were removed manually. The multiple sequence
516 alignment was visualized in Jalview (Waterhouse et al., 2009).

517

518 **Protein expression and purification**

519 The N-terminus truncated collembolan proteins (Supplementary Data 4), FcU1, FcU2,
520 Exd and Hth used for protein expression were constructed in the pGEX-4t-1 plasmid
521 and transformed into *E. coli* competent cells (BL21 strain). Protein expression was first
522 started from 500 mL of Luria–Bertani (LB) medium (pH = 7) cultured at 37 °C at 220
523 rpm. Until it reached the exponential growth phase (OD600 = 0.8~1.0), isopropyl β-D-

524 1-thiogalactopyranoside (IPTG) was added (0.1 mM) to 2 L of LB medium. The whole
525 culture was then subjected to low-temperature induction of protein expression at 16 °C
526 and 220 rpm for 12 hours. All the GST-tagged proteins were purified in native
527 conditions according to the manufacturer's instructions (Sangon Biotech, Shanghai,
528 China). All proteins were quantitated by comparing them to a set of bovine serum
529 albumin (BSA) concentration gradients: 750, 500, 250, 125, 100, 75, 50, 25, and 0 ng/μl
530 by Coomassie blue staining and further confirmed by anti-GST western blotting.
531

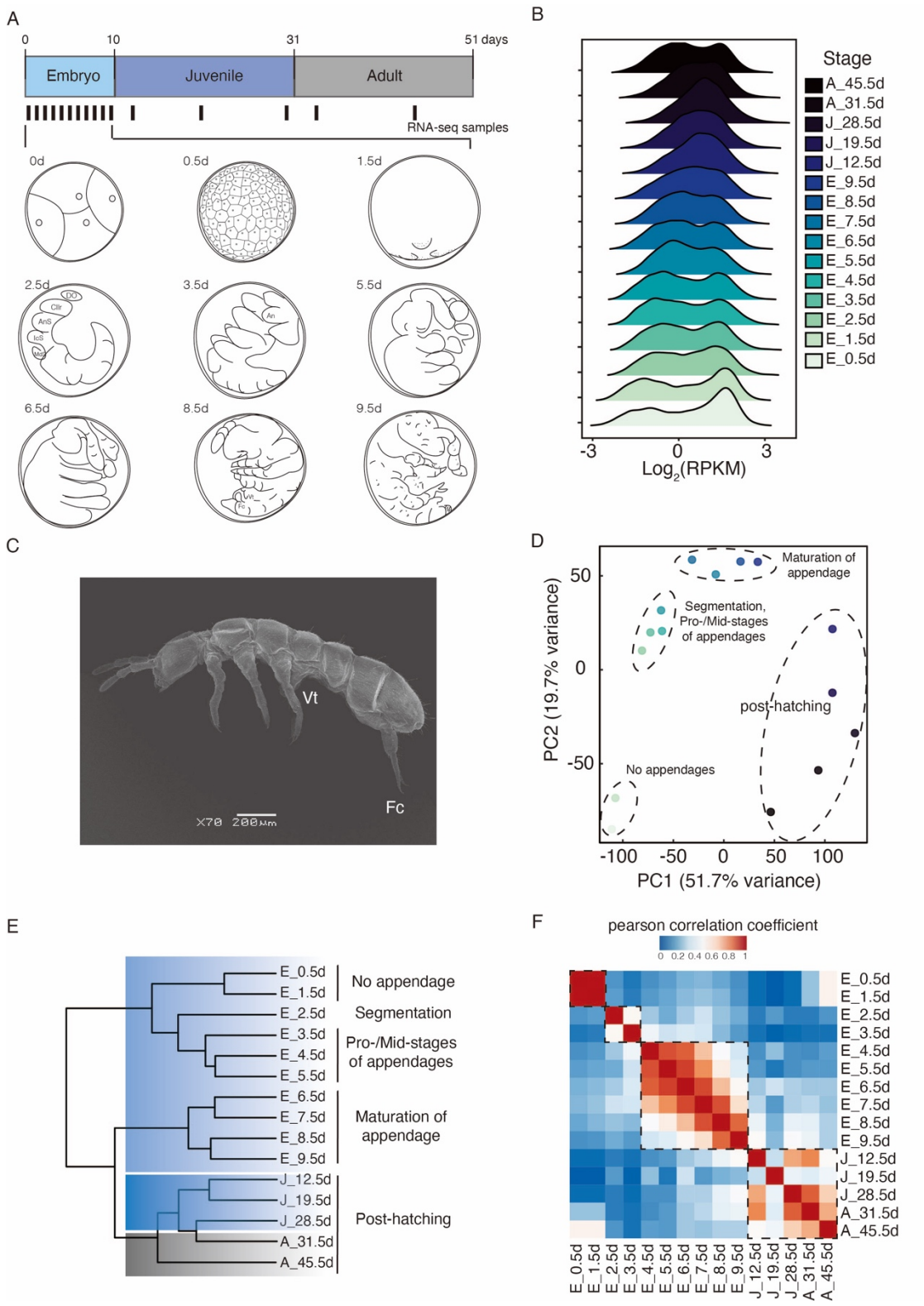
532 **Electrophoretic mobility shift assays (EMSAs)**

533 The repression regulatory element of *Dll* (Dll304) in *D. melanogaster*, the DMXR
534 element (Gebelein et al., 2002; Gebelein et al., 2004) and a putative regulatory element
535 of *Dll* in *F. candida* (Putative FcDll Element, PFE) were utilized as a DNA probe for
536 EMSAs separately. Each probe was first synthesized as two separate primers (forward
537 and reverse), which were tagged with Cy5 at the 5' end. The primers were diluted to a
538 concentration of 1 μM. Subsequently, 50 μl of each primer was mixed and incubated at
539 95 °C and then cooled to room temperature. The primers were annealed, resulting in
540 the formation of double-stranded DNA probes. 20 ng of DNA probe was used in each
541 EMSA, and the amount of protein used in each EMSA was 10 pmol of GST, 0.2, 1, and
542 1.5 pmol of FcU1, 0.4, 2, and 3 pmol of FcU2 and 1 pmol of Exd and Hth. The
543 procedures for electrophoretic mobility shift assays (EMSAs) were adjusted and
544 visualized according to the protocol (He et al., 2016). The 20 μl EMSA solution: 5×
545 EMSA buffer: Tris (0.1 M), glycerol (25%), and BSA (0.2 mg/ml). 2.5 M $MgCl_2$, 1 M
546 DTT, double-distilled water, poly(dI-dC), protein, and Cy5-labelled DNA probe. The
547 reaction was incubated for 25 minutes at 25 °C. The 12% non-denaturing
548 polyacrylamide gel contained double-distilled water, 5× TBE buffer, 50% glycerol,
549 polyacrylamide (the monomer: dimer ratio was 80:1), 10% ammonium persulfate, and
550 tetramethyl ethylenediamine (TEMED). Electrophoresis was carried out on ice, with
551 the blank gel pre-electrophoresed at 120 V for 1 hour. Subsequently, 20 μl of the

552 reaction mixture was loaded into each vial of the gel, and the samples were run at 150
553 V for 1-1.5 hours. For reaction visualization, the entire gel with the glass container was
554 scanned directly with the A Starion FLA-900 phosphorimager (Fujifilm, Japan).
555

556 ***Drosophila* S2 cell transfection and dual luciferase reporter assays**

557 The complete sequences of *D. melanogaster* Ubx (DmUbx_Ib, DmUbx_IVa) and
558 collembolan Ubx (FcU1, FcU2), the truncated collembolan Ubx (FcU1/△C, FcU2/△C)
559 and chimeric Ubx of *Drosophila* and collembolan (Dm/Fc_L, Dm/Fc_C) were cloned
560 into the pAC5.1 expression vector (Supplementary Data 4). To test for repression on
561 DMXR, this regulatory sequence was constructed into a pGL3-promoter vector
562 (Promega), which contains the SV40 enhancer and firefly luciferase. Transfection
563 experiments were carried out using the Qiagen Effectene reagent (Qiagen, Germany).
564 Initially, *Drosophila S2* cells were cultured in Schneider's *Drosophila* medium (Sigma–
565 Aldrich, USA) supplemented with 10% fetal bovine serum (HyClone, USA). After 24
566 hours, the cells were aliquoted into 48-well plates at 150 μl per well. After 12 hours, 50
567 μl of ENA-enhancer mixture, 1.2 μl of enhancer, 0.075 μg of pGL3-promoter vector,
568 0.075 μg of pAC5.1 vector, and 0.2 μl of the Renilla fluorescent vector (pRL) were
569 added to the mixture, followed by an 8-minute incubation. Then, 4 μl of Effectene and
570 200 μl of cell culture medium were added, and the transfected cells were cultured for
571 48 h in a 27 °C incubator. Each protein vector was set with technical triplicates.
572 The Dual luciferase reporter assay kit (Promega, USA) was used to measure the reporter
573 gene expression, and the luciferase activities were detected by a Modulus^{TM} Microplate
574 Luminometer (Turner BioSystems, USA). The repression activity of each protein is
575 represented by the average ratio of Firefly:Renilla luciferase activity. To estimate the
576 relative repression of the proteins on the DMXR, we compared the repression activity
577 of each sample with that of DmUbx_Ib. For pairwise comparisons, we performed
578 Student's t-test in R (RStudio, 2020).
579


A
0
10
31
51 days
Embryo
Juvenile
Adult
RNA-seq samples
0d
0.5d
1.5d
2.5d
3.5d
5.5d
6.5d
8.5d
9.5d
B
Stage
A_45.5d
A_31.5d
J_28.5d
J_19.5d
J_12.5d
E_9.5d
E_8.5d
E_7.5d
E_6.5d
E_5.5d
E_4.5d
E_3.5d
E_2.5d
E_1.5d
E_0.5d
-3
0
3
Log2(RPKM)
C
Vt
Fc
X70 200µm
D
Maturation of appendage
Segmentation, Pro-/Mid-stages of appendages
post-hatching
No appendages
PC2 (19.7% variance)
PC1 (51.7% variance)
50
0
-50
-100
-50
0
50
100
E
No appendage
Segmentation
Pro-/Mid-stages of appendages
Maturation of appendage
Post-hatching
F
pearson correlation coefficient
0 0.2 0.4 0.6 0.8 1

580

581

582 **Figure 1 Analysis of developmental transcriptomes of *F. candida*.**

583 **A**. Schematic of samples for developmental transcriptome sequencing and the
584 observation of embryonic development. Whole-animal collections were made from
585 embryos (blue), juveniles (grey blue) and adults (grey). The bars indicate that different
586 developmental samples were collected. 0d (d, days after oviposition): 4-cell stage; 0.5d:
587 blastula stage; 1.5d: gastrula stage; 2.5d: prophase of tissue differentiation stage; 3.5-
588 6.5d: metaphase of tissue differentiation stage; 7.5-8.5d: anaphase of tissue
589 differentiation stage; 9.5d: prehatching stage. An: antenna; AnS: antenna segment; Cllr:
590 clypeolabrum; DO: dorsal organ; Fc: furca; IcS: intercalary segment; MdS: mandibular
591 segment; Vt: ventral tube. **B**. Distribution of gene expression of 25,038 genes from 15
592 developmental transcriptomes. RPKM, reads per kilobase per million, normalizes the
593 raw count by transcript length and sequencing depth. **C**. Scanning electron microscopy
594 (SEM) of the whole animal of adult *F. candida*. Vt: ventral tube, Fc: furca. **D**. Principal
595 component analysis of the developmental transcriptomes. The dashed ellipses indicate
596 the stages of appendage development. **E**. Simplified hierarchical clustering of all
597 developmental transcriptomes. The stages of appendage development are shown on the
598 right. **F**. Heatmap of the Pearson correlation coefficient of all the samples; the dashed
599 rectangles indicate the relationship of samples.
600

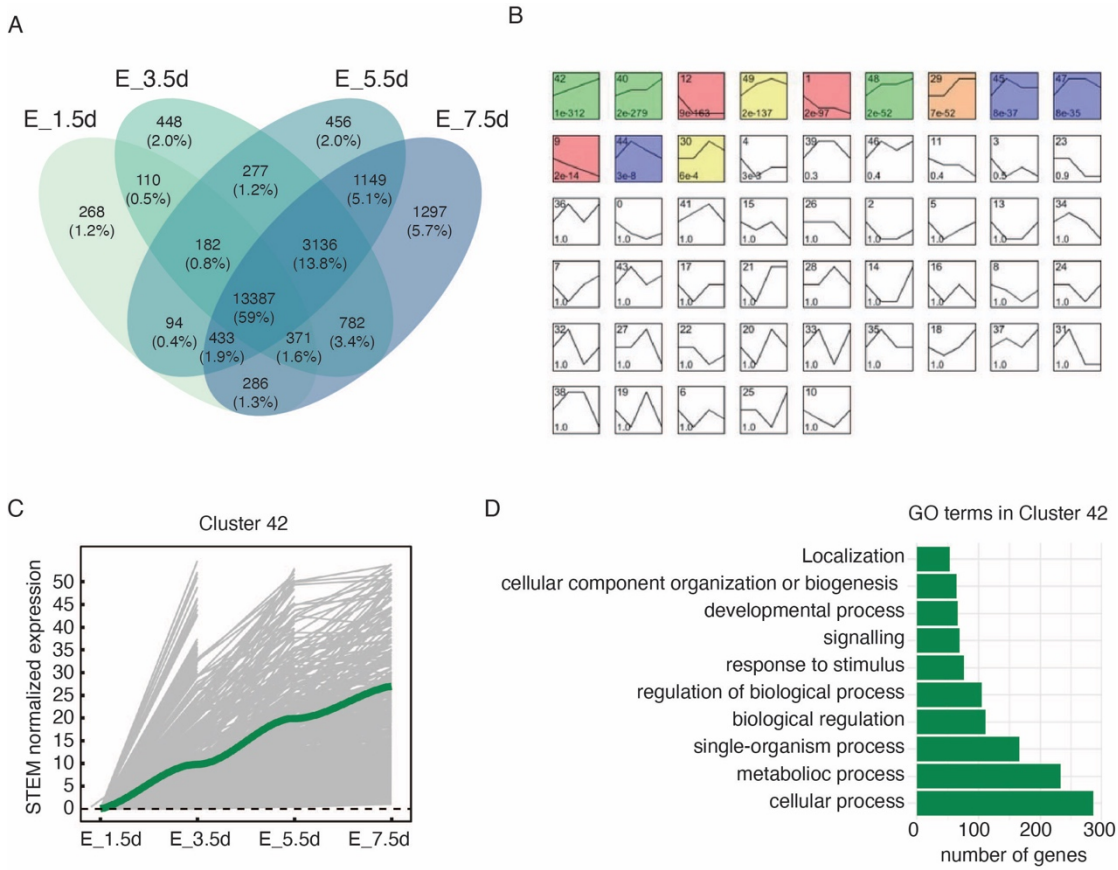


Figure 2 Identification of genes involved in appendage formation.

A. Venn diagram illustrating the number and percentage of expressed genes of each intersection among the four appendage formation stages. **B.** A screenshot of different gene expression profiles identified by short time-series expression miner (STEM) analysis. The number of each profile is shown at the upper left in the square. The coloured profiles indicated that genes were clustered significantly (permutation test, $p < 0.001$). **C.** The STEM normalized expression of Cluster 42. Values higher than 50 were excluded. **D.** Annotated GO terms of Cluster 42.

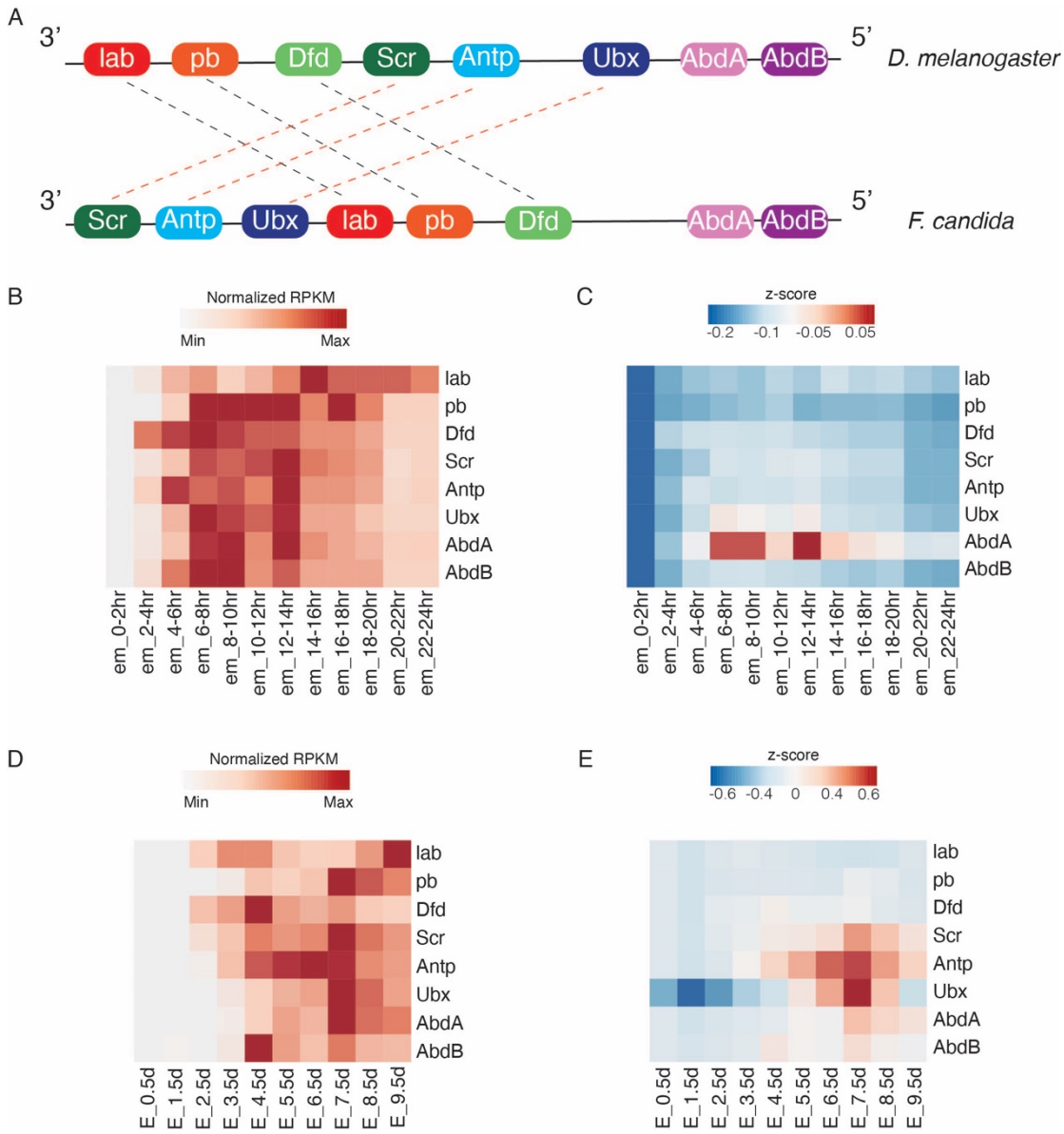


Figure 3 Comparison of the Hox cluster between *D. melanogaster* and *F. candida*.

A. Genomic architecture of the Hox cluster in *D. melanogaster* (Refined from Pearson, et al. 2005) and *F. candida* (Refined from Faddeeva-Vakhrusheva, et al. 2017). The distances between genes were not scaled in proportion to the original genomic distance. Dashed lines indicate that the Hox genes are rearranged. **B, C, D, E.** Expression profiles of Hox genes in *D. melanogaster* **(B, C)** and *F. candida* **(D, E)**. Normalized RPKM, the RPKM were normalized by the minimum and maximum expression value of each gene throughout embryonic stages. Z score, the RPKM values were normalized by z-scale.

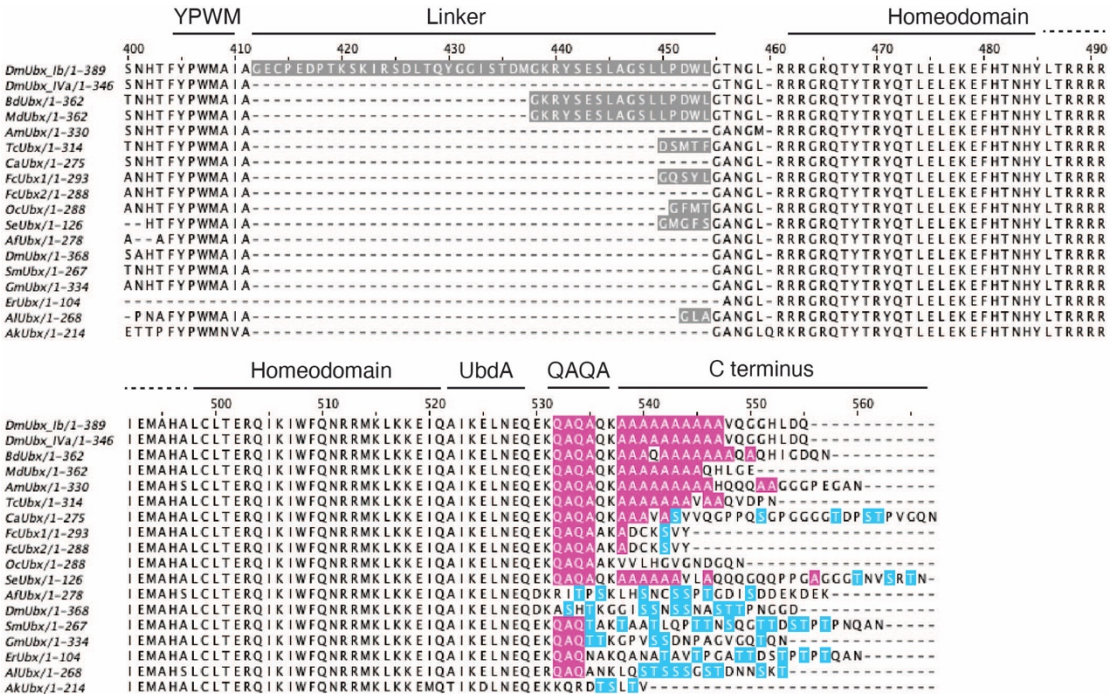
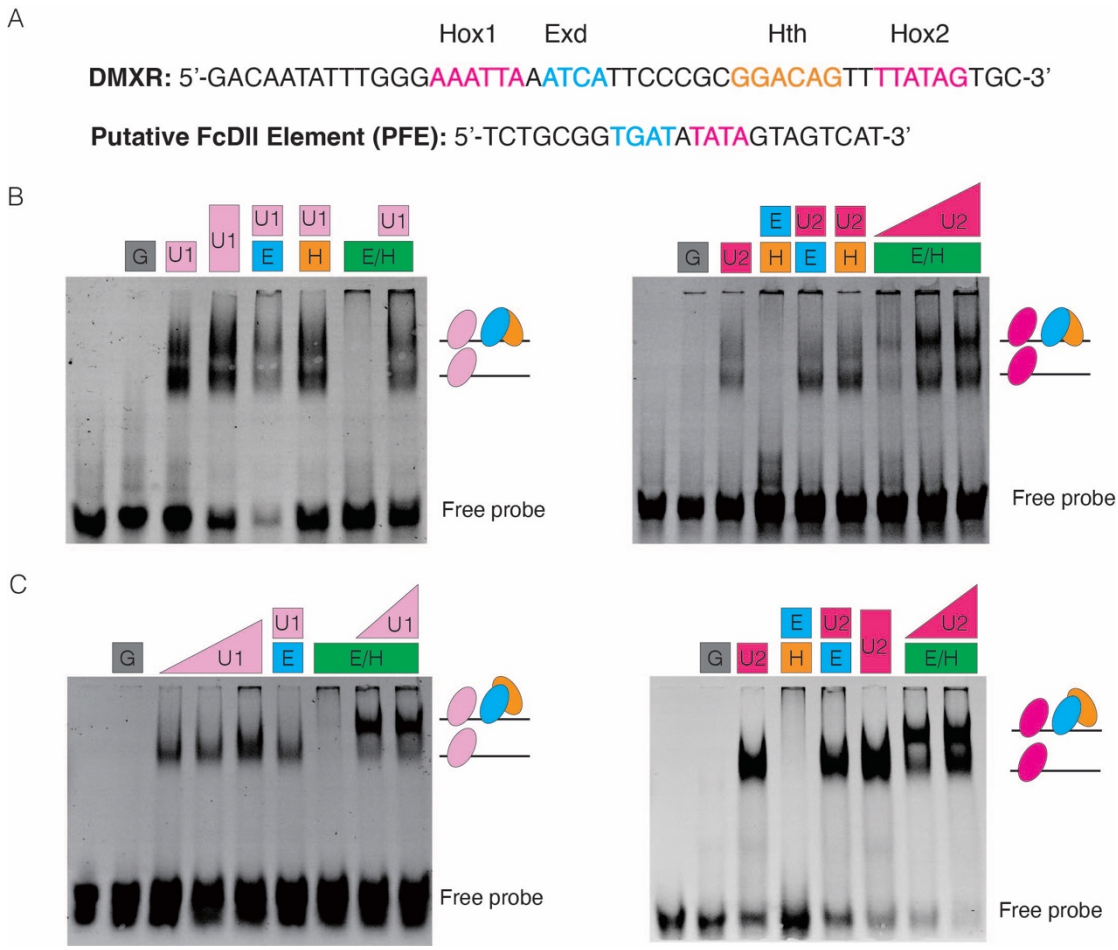


Figure 4 Sequence comparison of the linker and C-terminus of the panarthropod Ubx

Partial multiple sequence alignment of Ubx across Panarthropoda, encompassing Insecta, Diplora, Collembola, Protura, Crustacea, Myriapoda, Chelicerata and Onychophora. Insecta: *Drosophila melanogaster* (DmUbx_Ib, DmUbx_IVa), *Bactrocera dorsalis* (BdUbx), *Musca domestica* (MdUbx), *Apis mellifera* (AmUbx), *Tribolium castaneum* (TcUbx); Diplora: *Campodea augens* (CaUbx); Collembola: *Folsomia candida* (FcUbx1, FcUbx2), *Orchesella cincta* (OcUbx); Protura: *Sinentomon erythranum* (SeUbx); Crustacea: *Artemia franciscana* (AfUbx), *Daphnia magna* (DmUbx); Myriapoda: *Strigamia maritima* (SmUbx), *Glomeris marginata* (GmUbx), *Ethmostigmus rubripes* (ErUbx); Chelicerata: *Archegozetes longisetosus* (AlUbx); Onychophora: *Acanthokara kaputensis* (AkUbx). The functional domains are indicated. YPWM, YPWM motif; UbdA, UbdA motif. Features of sequences are highlighted in colours. The QAQA domain and poly-Ala stretch are marked in pink, and putative phosphorylated sites S/T (Ser/Thr) are marked in blue. Linkers are marked in grey.



638

639 **Figure 5 Collembolan Ubx and its cofactors could bind to the DNA elements that**

640 **contain Hox/Exd/Hth binding motifs.**

641 **A.** DNA probes used for EMSAs. DMXR, the transcription regulatory element of *D.*

642 *melanogaster Dll* (Gebelein et al., 2002; Gebelein et al., 2004); Putative FcDll Element

643 (PFE), a screened DNA region of *F. candida*, containing the Hox and Exd binding

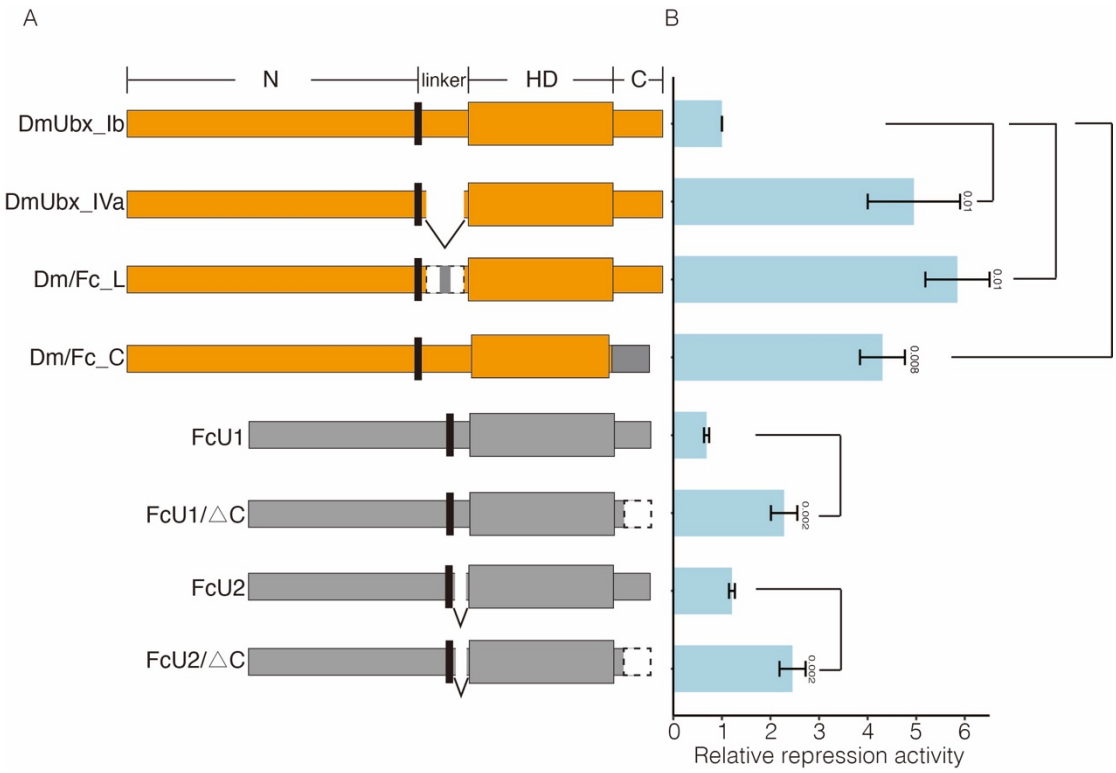
644 motifs. The binding sites of Hox/Exd/Hth are shown in colours. **B, C.** Assemblies of

645 collembolan Ubx/Exd/Hth on DMXR and Putative FcDll Element (PFE), respectively.

646 G, GST; U1, FcU1; U2, FcU2; E, Exd; H, Hth. Simplified complexes are indicated on

647 the right.

648



649

650 **Figure 6 Collembolan Ubx could repress *Dll* transcription while its C-terminus**
651 **contains regulatory domain(s).**

652 **A.** Constructs of proteins used for transcription repression assays, including complete
653 sequences, chimaeras, and truncations of *D. melanogaster* (yellow shade) and *F.*
654 *candida* (grey shade) Ubx. Dm/Fc_L, the linker of DmUbx_Ib was replaced by the
655 linker (GQSYL) of FcU; Dm/Fc_C, the C-terminus of DmUbx_Ib was replaced by the
656 C-terminus (AKADCKSVY) of FcU; FcU1△C and FcU2△C, the C-terminus (QAQA
657 AKADCKSVY) of FcUbx were deleted. **B.** The relative transcriptional repressive
658 activity of Ubx. The number on the Y-axis indicates transcriptional repression activity
659 relative to DmUbx_Ib. Significant P values of selected pairwise comparisons are shown
660 (Student's t-test).
661

662 **Table 1** Genes classified in morphogenesis from Cluster 42.

Gene_ID	Gene_symbol	Molecular function
XLOC_013891	FBN2	embryonic limb morphogenesis
XLOC_001350	Exd	somatic muscle development
XLOC_002645	Actin	muscle contraction
XLOC_017923	BMI1	segment specification
XLOC_019142	GATA-3	anatomical structure morphogenesis
XLOC_016846	Wnt16	Wnt signaling pathway
XLOC_022735	En	trunk segmentation
XLOC_011518	Ubx	specification of segmental identity
XLOC_011519	Antp	specification of segmental identity
XLOC_015540	ROR1	inner ear development
XLOC_005027	FOX	anatomical structure morphogenesis
XLOC_013406	MYO1E	microtubule-based movement
XLOC_016945	Notch	anterior/posterior pattern specification
XLOC_011524	Scr	specification of segmental identity
XLOC_009469	Ski	embryonic limb morphogenesis
XLOC_019633	Hth	segmentation
XLOC_007689	LARGE2	muscle cell cellular homeostasis
XLOC_016637	Lhx1	anatomical structure morphogenesis

663

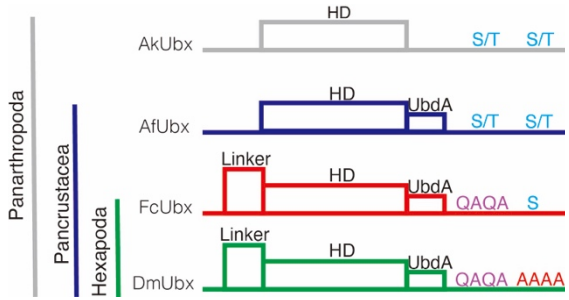
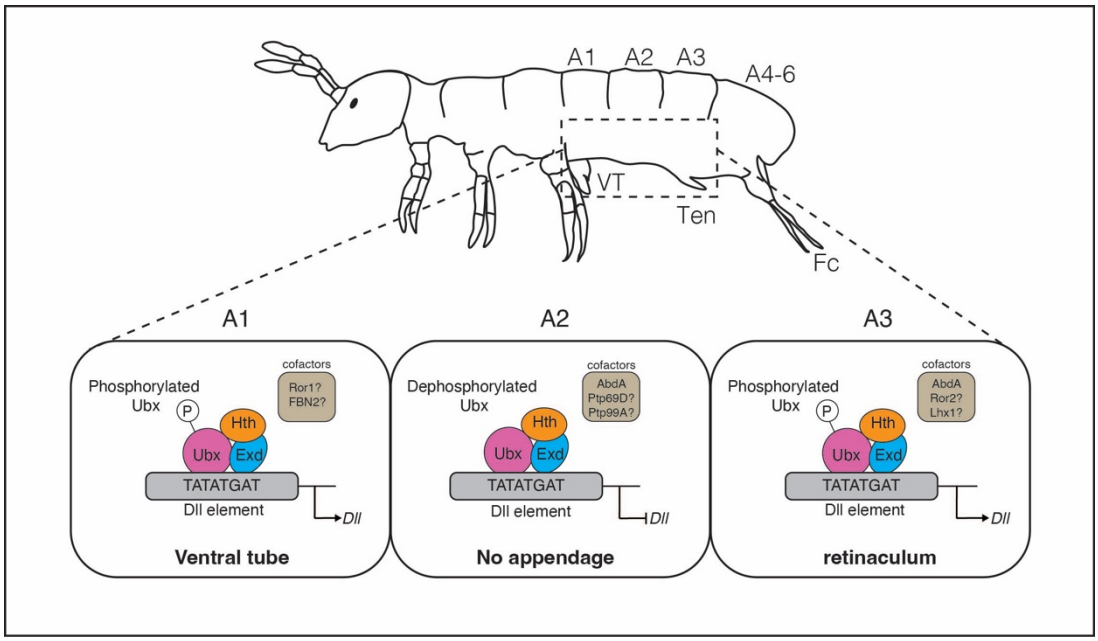


Table 2 Functional evolution of Ubx.

DNA binding	**Functional dissection of Ubx on *Dll* repression**			
	Repression capacity of complete protein	Indispensability of linker	C-terminus	
			Repression domain	phosphorylation site
+	-	-	ND	+
ND	-	-	-	+
+	+	-	+	+
+	+	+	+	-

Note: Schematic summary of the functional domains in Ubx of panarthropods. +: strong phenotype; -: no phenotype; ND: no data.



The proposed model for the regulation of collembolan Ubx on *Dll* expression

A schematic morphology of adult *F. candida*. The abdominal segments are shown as A1 to A6. VT, ventral tube; Ten, Retinulum; Fc, Furca. The functional regulatory model of Ubx: In A2, Ubx exerts its repression function on *Dll* expression, therefore suppressing appendage formation. Conversely, in A1 and 3, Ubx is phosphorylated, resulting in the derepression of *Dll* and facilitating the formation of appendages. See the text in the discussion for a detailed explanation.

676 **Supplementary Data**

677 1. The gene expression dynamics of *F. candida*

678 2. Short Time-series Expression Miner analysis

679 3. The GO annotation and KEGG pathways of Cluster 42

680 4. Multiple sequence alignment of Ubx

681 5. Truncated proteins of Ubx/Exd/Hth

682 6. Genomic sequence of *F. candida Dll*

683 7. Dual luciferase assays of collembolan and *Drosophila* proteins

684 8. Genes coexpressed with *FcUbx*

685

686 **Competing Interests**

687 The authors have declared that no competing interests exist.

688

689 **Authors' contributions**

690 YL and Y-XL conceived the project. YL performed all the experiments and data
691 analyses and drafted the manuscript. The authors read and approved the manuscript.

692

693 **Acknowledgements**

694 We thank all the lab members from Yin and Luan's previous laboratory for their help
695 and valuable discussion about this work. We would like to express appreciation to many
696 colleagues. Dr. WU Wei and the team from the Core Facility of Drosophila Resource
697 and Technology at the Centre for Excellence in Molecular Cell Science, CAS for their
698 assistance in establishing the transgenic fly strains of collembolan Ubx and providing
699 troubleshooting support. Miss LI Jiqing and Mr GAO Xiaoyan from the Electron
700 Microscope Platform at the Centre for Excellence in Molecular Plant Sciences, CAS
701 for the aid in the SEM of the collembolans. Dr. ZHAO Qin and Dr. WANG Jing for
702 their suggestions on protein expression. Dr. HE Juanmei and Dr. LANG Nannan for
703 generously sharing the EMSA protocol. Dr. JIA Qiangqiang for assistance with
704 luciferase assays in *Drosophila* S2 cells. Dr. ZHANG Yu for help in bioinformatic
705 analysis. Dr. Chema Martin at Queen Mary University of London and Dr. Kamil Jaron
706 at Wellcome Sanger Institute for the comments on the manuscript.

707

References

Abzhanov, A., & Kaufman, T. C. (2000). Homologs of Drosophila Appendage Genes in the Patterning of Arthropod Limbs. *Developmental Biology*, *227*(2), 673-689. https://doi.org/https://doi.org/10.1006/dbio.2000.9904

Akam, M., Averof, M., Castelli-Gair, J., Dawes, R., Falciani, F., & Ferrier, D. (1994). The evolving role of Hox genes in arthropods. *Dev Suppl*, 209-215.

Akinci, E., Hamilton, M. C., Khowpinitchai, B., & Sherwood, R. I. (2021). Using CRISPR to understand and manipulate gene regulation. *Development*, *148*(9). https://doi.org/10.1242/dev.182667

Ala, U., Piro, R. M., Grassi, E., Damasco, C., Silengo, L., Oti, M., Provero, P., & Di Cunto, F. (2008). Prediction of human disease genes by human-mouse conserved coexpression analysis. *PLoS Comput Biol*, *4*(3), e1000043. https://doi.org/10.1371/journal.pcbi.1000043

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol*, *215*(3), 403-410. https://doi.org/10.1016/s0022-2836(05)80360-2

Angelini, & Kaufman. (2005a). Comparative Developmental Genetics and the Evolution of Arthropod Body Plans. *Annual Review of Genetics*, *39*(1), 95-119. https://doi.org/10.1146/annurev.genet.39.073003.112310

Angelini, & Kaufman. (2005b). Insect appendages and comparative ontogenetics. *Dev Biol*, *286*(1), 57-77. https://doi.org/10.1016/j.ydbio.2005.07.006

Averof, M., & Patel, N. H. (1997). Crustacean appendage evolution associated with changes in Hox gene expression. *Nature*, *388*(6643), 682-686. https://doi.org/10.1038/41786

Bairoch, A., Apweiler, R., Wu, C. H., Barker, W. C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R., Magrane, M., Martin, M. J., Natale, D. A., O'Donovan, C., Redaschi, N., & Yeh, L. S. (2005). The Universal Protein Resource (UniProt). *Nucleic Acids Res*, *33*(Database issue), D154-159. https://doi.org/10.1093/nar/gki070

Bar-Joseph, Z. (2004). Analyzing time series gene expression data. *Bioinformatics*, *20*(16), 2493-2503. https://doi.org/10.1093/bioinformatics/bth283

Barkó, S., Bugyi, B., Carlier, M.-F., Gombos, R., Matusek, T., Mihály, J., & Nyitrai, M. (2010). Characterization of the Biochemical Properties and Biological Function of the Formin Homology Domains of Drosophila DAAM. *Journal of Biological Chemistry*, *285*(17), 13154-13169. https://doi.org/10.1074/jbc.M109.093914

Barkó, S., Bugyi, B., Carlier, M. F., Gombos, R., Matusek, T., Mihály, J., & Nyitrai, M. (2010). Characterization of the biochemical properties and biological function of the formin homology domains of Drosophila DAAM. *J Biol Chem*, *285*(17), 13154-13169. https://doi.org/10.1074/jbc.M109.093914

Berger, M. F., Badis, G., Gehrke, A. R., Talukder, S., Philippakis, A. A., Peña-Castillo, L., Alleyne, T. M., Mnaimneh, S., Botvinnik, O. B., Chan, E. T., Khalid, F., Zhang, W., Newburger, D., Jaeger, S. A., Morris, Q. D., Bulyk, M. L., & Hughes,

T. R. (2008). Variation in homeodomain DNA binding revealed by high-resolution analysis of sequence preferences. *Cell*, *133*(7), 1266-1276. https://doi.org/10.1016/j.cell.2008.05.024

Bhar, A., Haubrock, M., Mukhopadhyay, A., Maulik, U., Bandyopadhyay, S., & Wingender, E. (2013). Coexpression and coregulation analysis of time-series gene expression data in estrogen-induced breast cancer cell. *Algorithms Mol Biol*, *8*(1), 9. https://doi.org/10.1186/1748-7188-8-9

Brown, J. B., Boley, N., Eisman, R., May, G. E., Stoiber, M. H., Duff, M. O., Booth, B. W., Wen, J., Park, S., Suzuki, A. M., Wan, K. H., Yu, C., Zhang, D., Carlson, J. W., Cherbas, L., Eads, B. D., Miller, D., Mockaitis, K., Roberts, J., Davis, C. A., Frise, E., Hammonds, A. S., Olson, S., Shenker, S., Sturgill, D., Samsonova, A. A., Weiszmann, R., Robinson, G., Hernandez, J., Andrews, J., Bickel, P. J., Carninci, P., Cherbas, P., Gingeras, T. R., Hoskins, R. A., Kaufman, T. C., Lai, E. C., Oliver, B., Perrimon, N., Graveley, B. R., & Celniker, S. E. (2014). Diversity and dynamics of the Drosophila transcriptome. *Nature*, *512*(7515), 393-399. https://doi.org/10.1038/nature12962

Budd, G. E., & Telford, M. J. (2009). The origin and evolution of arthropods. *Nature*, *457*(7231), 812-817. https://doi.org/10.1038/nature07890

Buffry, A. D., Kittelmann, S., & McGregor, A. P. (2023). Characterisation of the role and regulation of Ultrabithorax in sculpting fine-scale leg morphology [Original Research]. *Frontiers in Cell and Developmental Biology*, *11*. https://doi.org/10.3389/fcell.2023.1119221

Casares, F., Calleja, M., & Sánchez-Herrero, E. (1996). Functional similarity in appendage specification by the Ultrabithorax and abdominal-A Drosophila HOX genes. *Embo j*, *15*(15), 3934-3942. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC452108/pdf/emboj00015-0170.pdf

Cohen, B., Simcox, A. A., & Cohen, S. M. (1993). Allocation of the thoracic imaginal primordia in the Drosophila embryo. *Development*, *117*(2), 597-608. https://doi.org/10.1242/dev.117.2.597

Cohen, B., Wimmer, E. A., & Cohen, S. M. (1991). Early development of leg and wing primordia in the Drosophila embryo. *Mech Dev*, *33*(3), 229-240. https://doi.org/10.1016/0925-4773(91)90030-a

Cohen, S. M. (1990). Specification of limb development in the Drosophila embryo by positional cues from segmentation genes. *Nature*, *343*(6254), 173-177. https://doi.org/10.1038/343173a0

Conesa, A., Götz, S., García-Gómez, J. M., Terol, J., Talón, M., & Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, *21*(18), 3674-3676. https://doi.org/10.1093/bioinformatics/bti610

Cook, C. E., Smith, M. L., Telford, M. J., Bastianello, A., & Akam, M. (2001). Hox genes and the phylogeny of the arthropods. *Current Biology*, *11*(10), 759-763.

https://doi.org/https://doi.org/10.1016/S0960-9822(01)00222-6

Costello, I., Nowotschin, S., Sun, X., Mould, A. W., Hadjantonakis, A. K., Bikoff, E. K., & Robertson, E. J. (2015). Lhx1 functions together with Otx2, Foxa2, and Ldb1 to govern anterior mesendoderm, node, and midline development. *Genes Dev*, *29*(20), 2108-2122. https://doi.org/10.1101/gad.268979.115

Deng, Y., Bartosovic, M., Ma, S., Zhang, D., Kukanja, P., Xiao, Y., Su, G., Liu, Y., Qin, X., Rosoklija, G. B., Dwork, A. J., Mann, J. J., Xu, M. L., Halene, S., Craft, J. E., Leong, K. W., Boldrini, M., Castelo-Branco, G., & Fan, R. (2022). Spatial profiling of chromatin accessibility in mouse and human tissues. *Nature*, *609*(7926), 375-383. https://doi.org/10.1038/s41586-022-05094-1

Durston, A. J., Jansen, H. J., In der Rieden, P., & Hooiveld, M. H. (2011). Hox collinearity - a new perspective. *Int J Dev Biol*, *55*(10-12), 899-908. https://doi.org/10.1387/ijdb.113358ad

Ekker, S. C., Young, K. E., von Kessler, D. P., & Beachy, P. A. (1991). Optimal DNA sequence recognition by the Ultrabithorax homeodomain of Drosophila. *Embo j*, *10*(5), 1179-1186. https://doi.org/10.1002/j.1460-2075.1991.tb08058.x

Ernst, J., & Bar-Joseph, Z. (2006). STEM: a tool for the analysis of short time series gene expression data. *BMC Bioinformatics*, *7*, 191. https://doi.org/10.1186/1471-2105-7-191

Ernst, J., Nau, G. J., & Bar-Joseph, Z. (2005). Clustering short time series gene expression data. *Bioinformatics*, *21 Suppl 1*, i159-168. https://doi.org/10.1093/bioinformatics/bti1022

Faddeeva-Vakhrusheva, A., Kraaijeveld, K., Derks, M. F. L., Anvar, S. Y., Agamennone, V., Suring, W., Kampfraath, A. A., Ellers, J., Le Ngoc, G., van Gestel, C. A. M., Mariën, J., Smit, S., van Straalen, N. M., & Roelofs, D. (2017). Coping with living in the soil: the genome of the parthenogenetic springtail Folsomia candida. *BMC Genomics*, *18*(1), 493. https://doi.org/10.1186/s12864-017-3852-x

Fountain, M. T., & Hopkin, S. P. (2005). FOLSOMIA CANDIDA (COLLEMBOLA): A "Standard" Soil Arthropod. *Annual Review of Entomology*, *50*(1), 201-222. https://doi.org/10.1146/annurev.ento.50.071803.130331

Fujioka, M., Gebelein, B., Cofer, Z. C., Mann, R. S., & Jaynes, J. B. (2012). Engrailed cooperates directly with Extradenticle and Homothorax on a distinct class of homeodomain binding sites to repress sloppy paired. *Dev Biol*, *366*(2), 382-392. https://doi.org/10.1016/j.ydbio.2012.04.004

Galant, R., & Carroll, S. B. (2002). Evolution of a transcriptional repression domain in an insect Hox protein. *Nature*, *415*(6874), 910-913. https://doi.org/10.1038/nature717

Gao, Y., Bu, Y., Luan, Y.-X., & Yin, W.-Y. (2006). Preliminary Observation on the Embryonic Development of Folsomia candida (Collembola: Isotomidae). *Zoological Research*, *5*(27), 519-524.

Gao, Y., Bu, Y., & Luan, Y. X. (2008). Phylogenetic relationships of basal hexapods reconstructed from nearly complete 18S and 28S rRNA gene sequences. *Zoolog*

Sci, *25*(11), 1139-1145. https://doi.org/10.2108/zsj.25.1139

Gaunt, S. J. (2015). The significance of Hox gene collinearity. *Int J Dev Biol*, *59*(4-6), 159-170. https://doi.org/10.1387/ijdb.150223sg

Gebelein, B., Culi, J., Ryoo, H. D., Zhang, W., & Mann, R. S. (2002). Specificity of Distalless repression and limb primordia development by abdominal Hox proteins. *Dev Cell*, *3*(4), 487-498. https://doi.org/10.1016/s1534-5807(02)00257-5

Gebelein, B., McKay, D. J., & Mann, R. S. (2004). Direct integration of Hox and segmentation gene inputs during Drosophila development. *Nature*, *431*(7009), 653-659. https://doi.org/10.1038/nature02946

Geyer, A., Koltsaki, I., Hessinger, C., Renner, S., & Rogulja-Ortmann, A. (2015). Impact of Ultrabithorax alternative splicing on Drosophila embryonic nervous system development. *Mechanisms of Development*, *138*, 177-189. https://doi.org/https://doi.org/10.1016/j.mod.2015.08.007

Giribet, G., Edgecombe, G. D., & Wheeler, W. C. (2001). Arthropod phylogeny based on eight molecular loci and morphology. *Nature*, *413*(6852), 157-161. https://doi.org/10.1038/35093097

Grenier, J. K., & Carroll, S. B. (2000). Functional evolution of the Ultrabithorax protein. *Proc Natl Acad Sci U S A*, *97*(2), 704-709. https://doi.org/10.1073/pnas.97.2.704

He, J. M., Zhu, H., Zheng, G. S., Liu, P. P., Wang, J., Zhao, G. P., Zhu, G. Q., Jiang, W. H., & Lu, Y. H. (2016). Direct Involvement of the Master Nitrogen Metabolism Regulator GlnR in Antibiotic Biosynthesis in Streptomyces. *J Biol Chem*, *291*(51), 26443-26454. https://doi.org/10.1074/jbc.M116.762476

Headrick, D. H., & Gordh, G. (2009). Chapter 5 - Anatomy: Head, Thorax, Abdomen, and Genitalia. In V. H. Resh & R. T. Cardé (Eds.), *Encyclopedia of Insects (Second Edition)* (pp. 11-21). Academic Press. https://doi.org/https://doi.org/10.1016/B978-0-12-374144-8.00005-9

Hughes, & Kaufman. (2002a). Exploring the myriapod body plan: expression patterns of the ten Hox genes in a centipede. *Development*, *129*(5), 1225-1238. https://doi.org/10.1242/dev.129.5.1225

Hughes, & Kaufman. (2002b). Hox genes and the evolution of the arthropod body plan. *Evol Dev*, *4*(6), 459-499. https://doi.org/10.1046/j.1525-142x.2002.02034.x

Jockusch, E. L., & Smith, F. W. (2015). Hexapoda: Comparative Aspects of Later Embryogenesis and Metamorphosis. In A. Wanninger (Ed.), *Evolutionary Developmental Biology of Invertebrates 5: Ecdysozoa III: Hexapoda* (pp. 111-208). Springer Vienna. https://doi.org/10.1007/978-3-7091-1868-9_3

Jockusch, E. L., Williams, T. A., & Nagy, L. M. (2004). The evolution of patterning of serially homologous appendages in insects. *Development Genes and Evolution*, *214*(7), 324-338. https://doi.org/10.1007/s00427-004-0412-6

Kanehisa, M., & Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*, *28*(1), 27-30. https://doi.org/10.1093/nar/28.1.27

Kobayashi, M., Fujioka, M., Tolkunova, E. N., Deka, D., Abu-Shaar, M., Mann, R. S., & Jaynes, J. B. (2003). Engrailed cooperates with extradenticle and homothorax to repress target genes in Drosophila. *Development*, *130*(4), 741-751. https://doi.org/10.1242/dev.00289

Kolde, R. (2012). Pheatmap: pretty heatmaps. *R package version*, *1*(2), 726.

Konopova, B., & Akam, M. (2014). The Hox genes Ultrabithorax and abdominal-A specify three different types of abdominal appendage in the springtail Orchesella cincta (Collembola). *Evodevo*, *5*(1), 2. https://doi.org/10.1186/2041-9139-5-2

Krogh, P. H. (2009). Toxicity testing with the collembolans Folsomia fimetaria and Folsomia candida and the results of a ringtest.

Kurata, S., Go, M. J., Artavanis-Tsakonas, S., & Gehring, W. J. (2000). Notch signaling and the determination of appendage identity. *Proc Natl Acad Sci U S A*, *97*(5), 2117-2122. https://doi.org/10.1073/pnas.040556497

Liang, Y., Xie, W., & Luan, Y. X. (2019). Developmental expression and evolution of hexamerin and haemocyanin from Folsomia candida (Collembola). *Insect Molecular Biology*, *28*(5), 716-727. https://doi.org/https://doi.org/10.1111/imb.12585

Luan, Y. X., Cui, Y., Chen, W. J., Jin, J. F., Liu, A. M., Huang, C. W., Potapov, M., Bu, Y., Zhan, S., Zhang, F., & Li, S. (2023). High-quality genomes reveal significant genetic divergence and cryptic speciation in the model organism Folsomia candida (collembola). *Mol Ecol Resour*, *23*(1), 273-293. https://doi.org/10.1111/1755-0998.13699

Luan, Y. X., Mallatt, J. M., Xie, R. D., Yang, Y. M., & Yin, W. Y. (2005). The phylogenetic positions of three Basal-hexapod groups (protura, diplura, and collembola) based on ribosomal RNA gene sequences. *Mol Biol Evol*, *22*(7), 1579-1592. https://doi.org/10.1093/molbev/msi148

Mann, R. S., Lelli, K. M., & Joshi, R. (2009). Hox specificity unique roles for cofactors and collaborators. *Curr Top Dev Biol*, *88*, 63-101. https://doi.org/10.1016/s0070-2153(09)88003-4

Marx, V. (2021). Method of the Year: spatially resolved transcriptomics. *Nature Methods*, *18*(1), 9-14. https://doi.org/10.1038/s41592-020-01033-y

Matsuda, R. (2017). *Morphology and evolution of the insect abdomen: with special reference to developmental patterns and their bearings upon systematics*. Elsevier.

McGinnis, S., & Madden, T. L. (2004). BLAST: at the core of a powerful and diverse set of sequence analysis tools. *Nucleic Acids Res*, *32*(Web Server issue), W20-25. https://doi.org/10.1093/nar/gkh435

McIntosh, B. B., & Ostap, E. M. (2016). Myosin-I molecular motors at a glance. *J Cell Sci*, *129*(14), 2689-2695. https://doi.org/10.1242/jcs.186403

Mead, T. J., Martin, D. R., Wang, L. W., Cain, S. A., Gulec, C., Cahill, E., Mauch, J., Reinhardt, D., Lo, C., Baldock, C., & Apte, S. S. (2022). Proteolysis of fibrillin-

2 microfibrils is essential for normal skeletal development. *eLife*, *11*, e71142. https://doi.org/10.7554/eLife.71142

Misof, B., Liu, S., Meusemann, K., Peters, R. S., Donath, A., Mayer, C., Frandsen, P. B., Ware, J., Flouri, T., Beutel, R. G., Niehuis, O., Petersen, M., Izquierdo-Carrasco, F., Wappler, T., Rust, J., Aberer, A. J., Aspöck, U., Aspöck, H., Bartel, D., Blanke, A., Berger, S., Böhm, A., Buckley, T. R., Calcott, B., Chen, J., Friedrich, F., Fukui, M., Fujita, M., Greve, C., Grobe, P., Gu, S., Huang, Y., Jermiin, L. S., Kawahara, A. Y., Krogmann, L., Kubiak, M., Lanfear, R., Letsch, H., Li, Y., Li, Z., Li, J., Lu, H., Machida, R., Mashimo, Y., Kapli, P., McKenna, D. D., Meng, G., Nakagaki, Y., Navarrete-Heredia, J. L., Ott, M., Ou, Y., Pass, G., Podsiadlowski, L., Pohl, H., von Reumont, B. M., Schütte, K., Sekiya, K., Shimizu, S., Slipinski, A., Stamatakis, A., Song, W., Su, X., Szucsich, N. U., Tan, M., Tan, X., Tang, M., Tang, J., Timelthaler, G., Tomizuka, S., Trautwein, M., Tong, X., Uchifune, T., Walzl, M. G., Wiegmann, B. M., Wilbrandt, J., Wipfler, B., Wong, T. K., Wu, Q., Wu, G., Xie, Y., Yang, S., Yang, Q., Yeates, D. K., Yoshizawa, K., Zhang, Q., Zhang, R., Zhang, W., Zhang, Y., Zhao, J., Zhou, C., Zhou, L., Ziesmann, T., Zou, S., Li, Y., Xu, X., Zhang, Y., Yang, H., Wang, J., Wang, J., Kjer, K. M., & Zhou, X. (2014). Phylogenomics resolves the timing and pattern of insect evolution. *Science*, *346*(6210), 763-767. https://doi.org/10.1126/science.1257570

Monteiro, A. S., & Ferrier, D. E. K. (2006). Hox genes are not always Colinear [Review]. *International Journal of Biological Sciences*, *2*(3), 95-103. https://doi.org/10.7150/ijbs.2.95

Noyes, M. B., Christensen, R. G., Wakabayashi, A., Stormo, G. D., Brodsky, M. H., & Wolfe, S. A. (2008). Analysis of homeodomain specificities allows the family-wide prediction of preferred recognition sites. *Cell*, *133*(7), 1277-1289. https://doi.org/10.1016/j.cell.2008.05.023

O'Day, K. E. (2006). *Notch Signaling and Segmentation in Parhyale Hawaiensis*. University of California, Berkeley. https://books.google.co.uk/books?id=FbFPAQAAMAAJ

Pace, R. M., Grbić, M., & Nagy, L. M. (2016). Composition and genomic organization of arthropod Hox clusters. *Evodevo*, *7*, 11. https://doi.org/10.1186/s13227-016-0048-4

Palopoli, M. F., & Patel, N. H. (1998). Evolution of the interaction between Hox genes and a downstream target. *Curr Biol*, *8*(10), 587-590. https://doi.org/10.1016/s0960-9822(98)70228-3

Panganiban, G., Irvine, S. M., Lowe, C., Roehl, H., Corley, L. S., Sherbon, B., Grenier, J. K., Fallon, J. F., Kimble, J., Walker, M., Wray, G. A., Swalla, B. J., Martindale, M. Q., & Carroll, S. B. (1997). The origin and evolution of animal appendages. *Proceedings of the National Academy of Sciences*, *94*(10), 5162-5166. https://doi.org/doi:10.1073/pnas.94.10.5162

Passner, J. M., Ryoo, H. D., Shen, L., Mann, R. S., & Aggarwal, A. K. (1999). Structure

of a DNA-bound Ultrabithorax-Extradenticle homeodomain complex. *Nature*, *397*(6721), 714-719. https://doi.org/10.1038/17833

Pearson, J. C., Lemons, D., & McGinnis, W. (2005). Modulating Hox gene functions during animal body patterning. *Nature Reviews Genetics*, *6*(12), 893-904. https://doi.org/10.1038/nrg1726

Peel, A. D., Chipman, A. D., & Akam, M. (2005). Arthropod Segmentation: beyond the Drosophila paradigm. *Nature Reviews Genetics*, *6*(12), 905-916. https://doi.org/10.1038/nrg1724

Prokop, A., Sánchez-Soriano, N., Gonçalves-Pimentel, C., Molnár, I., Kalmár, T., & Mihály, J. (2011). DAAM family members leading a novel path into formin research. *Commun Integr Biol*, *4*(5), 538-542. https://doi.org/10.4161/cib.4.5.16511

Reed, H. C., Hoare, T., Thomsen, S., Weaver, T. A., White, R. A., Akam, M., & Alonso, C. R. (2010). Alternative splicing modulates Ubx protein function in Drosophila melanogaster. *Genetics*, *184*(3), 745-758. https://doi.org/10.1534/genetics.109.112086

Ronshaugen, M., McGinnis, N., & McGinnis, W. (2002). Hox protein mutation and macroevolution of the insect body plan. *Nature*, *415*(6874), 914-917. https://doi.org/10.1038/nature716

Rozewicki, J., Li, S., Amada, K. M., Standley, D. M., & Katoh, K. (2019). MAFFT-DASH: integrated protein sequence and structural alignment. *Nucleic Acids Research*, *47*(W1), W5-W10. https://doi.org/10.1093/nar/gkz342

RStudio. (2020). *RStudio: Integrated Development for R. RStudio*. http://www.rstudio.com/.

Ruddle, F. H., Bartels, J. L., Bentley, K. L., Kappen, C., Murtha, M. T., & Pendleton, J. W. (1994). EVOLUTION OF HOX GENES. *Annual Review of Genetics*, *28*(1), 423-442. https://doi.org/10.1146/annurev.ge.28.120194.002231

Ryoo, H. D., Marty, T., Casares, F., Affolter, M., & Mann, R. S. (1999). Regulation of Hox target genes by a DNA bound Homothorax/Hox/Extradenticle complex. *Development*, *126*(22), 5137-5148. https://doi.org/10.1242/dev.126.22.5137

Schwabe, G. C., Trepczik, B., Süring, K., Brieske, N., Tucker, A. S., Sharpe, P. T., Minami, Y., & Mundlos, S. (2004). Ror2 knockout mouse as a model for the developmental pathology of autosomal recessive Robinow syndrome. *Dev Dyn*, *229*(2), 400-410. https://doi.org/10.1002/dvdy.10466

Slattery, M., Riley, T., Liu, P., Abe, N., Gomez-Alcala, P., Dror, I., Zhou, T., Rohs, R., Honig, B., Bussemaker, Harmen J., & Mann, Richard S. (2011). Cofactor Binding Evokes Latent Differences in DNA Binding Specificity between Hox Proteins. *Cell*, *147*(6), 1270-1282. https://doi.org/10.1016/j.cell.2011.10.053

Takeuchi, S., Takeda, K., Oishi, I., Nomi, M., Ikeya, M., Itoh, K., Tamura, S., Ueda, T., Hatta, T., Otani, H., Terashima, T., Takada, S., Yamamura, H., Akira, S., & Minami, Y. (2000). Mouse Ror2 receptor tyrosine kinase is required for the heart development and limb formation. *Genes Cells*, *5*(1), 71-78.

https://doi.org/10.1046/j.1365-2443.2000.00300.x

Timmermans, M., Roelofs, D., Mariën, J., & van Straalen, N. M. (2008). Revealing pancrustacean relationships: Phylogenetic analysis of ribosomal protein genes places Collembola (springtails) in a monophyletic Hexapoda and reinforces the discrepancy between mitochondrial and nuclear DNA markers. *BMC Evolutionary Biology*, *8*(1), 83. https://doi.org/10.1186/1471-2148-8-83

Tsuji, T., Sato, A., Hiratani, I., Taira, M., Saigo, K., & Kojima, T. (2000). Requirements of Lim1, a Drosophila LIM-homeobox gene, for normal leg and antennal development. *Development*, *127*(20), 4315-4323. https://doi.org/10.1242/dev.127.20.4315

Tully, T., & Potapov, M. (2015). Intraspecific Phenotypic Variation and Morphological Divergence of Strains of Folsomia candida (Willem) (Collembola: Isotomidae), the "Standard" Test Springtaill. *PLoS One*, *10*(9), e0136047. https://doi.org/10.1371/journal.pone.0136047

van Dam, S., Võsa, U., van der Graaf, A., Franke, L., & de Magalhães, J. P. (2018). Gene co-expression analysis for functional classification and gene-disease predictions. *Brief Bioinform*, *19*(4), 575-592. https://doi.org/10.1093/bib/bbw139

van Straalen, N. M. (2021). Evolutionary terrestrialization scenarios for soil invertebrates. *Pedobiologia*, *87-88*, 150753. https://doi.org/https://doi.org/10.1016/j.pedobi.2021.150753

Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., & Barton, G. J. (2009). Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, *25*(9), 1189-1191. https://doi.org/10.1093/bioinformatics/btp033

Weatherbee, S. D., & Carroll, S. B. (1999). Selector Genes and Limb Identity in Arthropods and Vertebrates. *Cell*, *97*(3), 283-286. https://doi.org/10.1016/S0092-8674(00)80737-0

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. https://ggplot2.tidyverse.org

Yuzuki, D. (2015). *BGISEQ-500 debuts at the International Congress of Genomics 10*. In *Next Generation Technologist [online]* http://www.yuzuki.org/bgiseq-500-debut-at-the-international-congress-of-genomics-10