PROCEEDINGS OF THE ROYAL SOCIETY B

rspb.royalsocietypublishing.org

Research

Cite this article: Sharma PP, Schwager EE, Extavour CG, Wheeler WC. 2014 Hox gene duplications correlate with posterior heteronomy in scorpions. *Proc. R. Soc. B* **281**: 20140661.
http://dx.doi.org/10.1098/rspb.2014.0661

Received: 19 March 2014
Accepted: 22 July 2014

Subject Areas:
evolution

Keywords:
Scorpiones, opisthosoma, Arthropoda, bauplan, tagmosis

Author for correspondence:
Prashant P. Sharma
e-mail: psharma@amnh.org

Electronic supplementary material is available at http://dx.doi.org/10.1098/rspb.2014.0661 or via http://rspb.royalsocietypublishing.org.

Royal Society Publishing

Hox gene duplications correlate with posterior heteronomy in scorpions

Prashant P. Sharma[1], Evelyn E. Schwager[2], Cassandra G. Extavour[2] and Ward C. Wheeler[1]

[1]Division of Invertebrate Zoology, American Museum of Natural History, Central Park West at 79th Street, New York, NY 10024, USA
[2]Department of Organismic and Evolutionary Biology, Harvard University, 16 Divinity Avenue, Cambridge, MA 02138, USA

The evolutionary success of the largest animal phylum, Arthropoda, has been attributed to tagmatization, the coordinated evolution of adjacent metameres to form morphologically and functionally distinct segmental regions called tagmata. Specification of regional identity is regulated by the Hox genes, of which 10 are inferred to be present in the ancestor of arthropods. With six different posterior segmental identities divided into two tagmata, the bauplan of scorpions is the most heteronomous within Chelicerata. Expression domains of the anterior eight Hox genes are conserved in previously surveyed chelicerates, but it is unknown how Hox genes regionalize the three tagmata of scorpions. Here, we show that the scorpion *Centruroides sculpturatus* has two paralogues of all Hox genes except *Hox3*, suggesting cluster and/or whole genome duplication in this arachnid order. Embryonic anterior expression domain boundaries of each of the last four pairs of Hox genes (two paralogues each of *Antp*, *Ubx*, *abd-A* and *Abd-B*) are unique and distinguish segmental groups, such as pectines, book lungs and the characteristic tail, while maintaining spatial collinearity. These distinct expression domains suggest neofunctionalization of Hox gene paralogues subsequent to duplication. Our data reconcile previous understanding of Hox gene function across arthropods with the extreme heteronomy of scorpions.

1. Introduction

The evolutionary success of Arthropoda is attributed to their segmented bauplan and its modularization through tagmosis, whereby groups of adjacent segments evolve in concert to achieve morphological and functional distinction from other such groups along the anteroposterior (AP) axis [1]. A conserved cluster of transcription factors called the Hox genes play key roles in conferring segmental identity and have been implicated as the driving force in the evolution of tagmata [2–6]. Loss or gain of Hox gene function has been shown to cause homeotic transformations in multiple arthropod species [6–10]. Ten Hox genes are inferred to have been present in the common ancestor of arthropods and their sister group, Onychophora, and are typically closely genetically linked on a chromosome [4,11–14]. They are expressed collinearly along the AP axis, i.e. in the same order as they occur on the chromosome in the cluster [15,16], and spatial shifts in Hox gene expression are associated with morphological change in body plans [9,17–21].

Scorpions are an unusual and iconic lineage of arthropods, due to an ancient origin and marked heteronomy. This lineage is unique among arthropods in possessing a tagma dedicated exclusively to prey capture and defence: the flexible and photosensitive metasoma (tail) of scorpions harbours the venom glands and expresses one of the opsins shared with the eyes, conferring light signal transduction to this eyeless body region [22–25]. Like most Eucheliceratа, the prosoma (anterior tagma) of scorpions consists of seven segments: the protocerebral, cheliceral, pedipalpal and four leg-bearing segments (but see [26] for a review of the segmental nature of the protocerebrum). Hox gene expression surveys of multiple chelicerate orders (spiders, mites and harvestmen) have demonstrated the

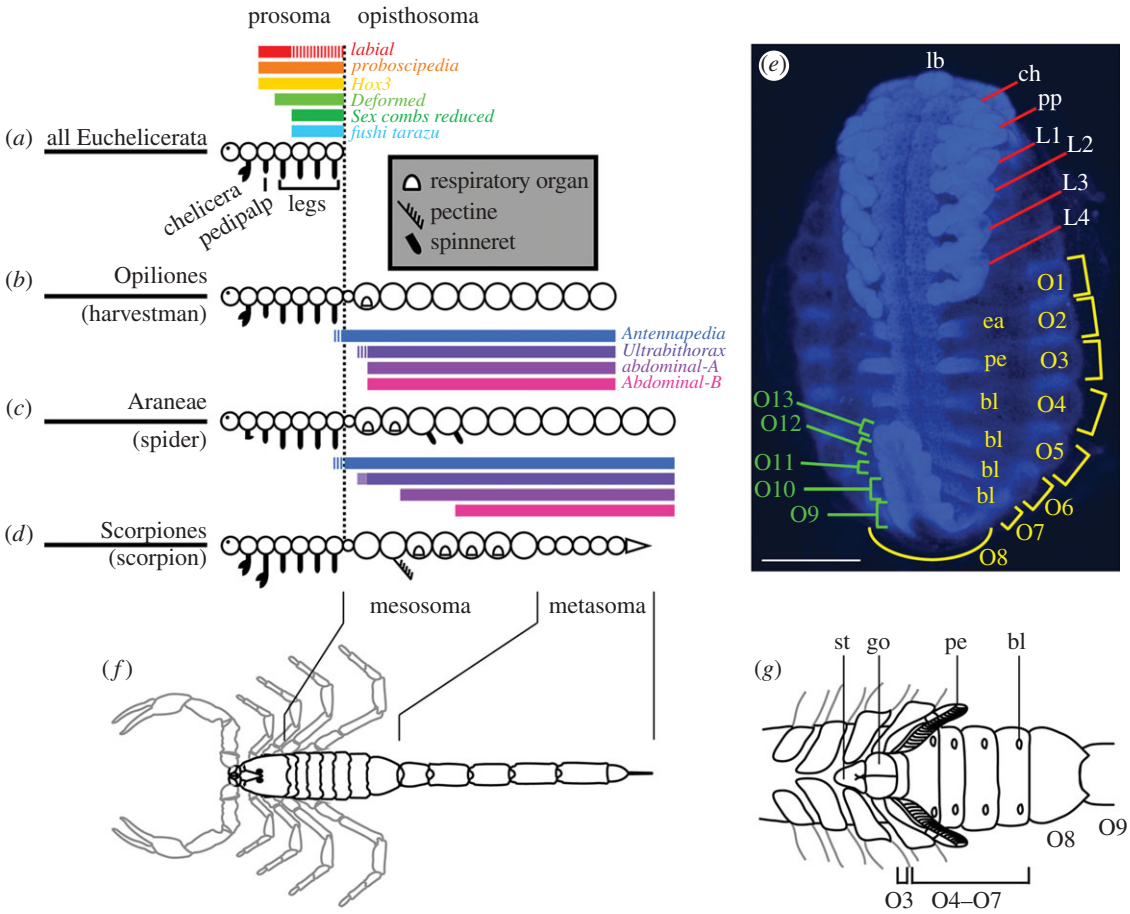

Figure 1. Four Hox genes are expressed in the chelicerate opisthosoma. (*a*) Summary of Hox gene expression domains in the chelicerate prosoma, based on data from spiders, mites and harvestmen. (*b*) Opisthosomal Hox expression in Opiliones is correlated with the homonomy of the harvestman posterior tagma. Note coincident anterior expression boundaries of *Ubx*, *abd-A* and *Abd-B*. (*c*) Differentiation of four distinct anterior opisthosomal segment types in spiders is correlated with staggered expression domains of the opisthosomal Hox group. (*d*) The bauplan of scorpions has six segmental identities divided into two tagmata, which cannot be reconciled with only four Hox genes in this region. (*e*) Limb bud stage embryo of *C. sculpturatus*, with segments of the prosoma, mesosoma and metasoma completely formed. (*f*) Generalized scorpion anatomy in dorsal view. (*g*) Generalized scorpion mesosomal anatomy in ventral view. bl, book lung; ch, chelicera; ea, embryonic appendage; go, genital operculum; L1, first walking leg; lb, labrum; O1, first opisthosomal segment; pe, pectine; pp, pedipalp; st, sternum. Scale bar in (*e*): 200 µm. (Online version in colour.)

evolutionary conservation of Hox patterning in the prosoma. As in the anterior tagma of all arthropods surveyed, the patterning of the prosoma is associated with staggered anterior expression boundaries of the first six Hox genes, engendering a unique combination of Hox transcripts for each segmental identity (figure 1*a*) [4,27,28]. In chelicerates, few of these anterior six Hox genes are expressed posterior to the last prosomal segment, with transient expression of *proboscipedia* (*pb*), *Hox3*, *Deformed* (*Dfd*) and *Sex combs reduced* (*Scr*) observed in posterior segments of some, but not all, species. This leaves four Hox genes, *Antennapedia* (*Antp*), *Ultrabithorax* (*Ubx*), *abdominal-A* (*abd-A*) and *Abdominal-B* (*Abd-B*), to pattern the posterior tagma, the opisthosoma [4,28].

Functional studies of *Antp* in the spider *Parasteatoda tepidariorum* have demonstrated the role of this gene in enforcing the prosomal–opisthosomal boundary [10]. The earliest expression of *Antp* consistently occurs in the first opisthosomal (O1) segment of multiple chelicerates [10,12,28]. Additionally, a previous study illustrated the lability of *abd-A* and *Abd-B* anterior expression boundaries in some chelicerate orders. In harvestmen (Opiliones), the homonomous opisthosoma is correlated with coincident anterior boundaries of *Ubx*, *abd-A* and *Abd-B* in the second opisthosomal segment (figure 1*b*) [28]. In spiders, the differentiation of the opisthosoma into segments forming the pedicel (O1), the respiratory organs (book lungs and/or tracheal tubules; O2–O3), spinnerets (silk-spinning organs; O4–O5) or no external organs at all (O6–O12), is in turn correlated with the staggered expression domains of *Ubx*, *abd-A* and *Abd-B*, which result in unique combinations of Hox transcripts in each segment type (figure 1*c*) [27,29].

Unlike all other Chelicerata, scorpions have a highly differentiated opisthosoma that contains six segmental identities conventionally grouped into two tagmata: the mesosoma ('abdomen,' segments O1 to O8) and the metasoma (segments O9 to O13) (figure 1*d*). The mesosoma and the metasoma are functionally distinct tagmata, with the former harbouring major organ systems, and the latter dedicated to prey immobilization and defence through envenomation (figure 1*f*,*g*). In the mesosoma, O1 develops a transient pair of embryonic appendage rudiments that are resorbed before hatching, and the remainder of the segment becomes part of the genital chamber in adults. O2 bears the genital pores, O3 the pectines (sensory appendages unique to scorpions) and O4–O7 the book lungs. No organs form on O8, the tapering last segment of the mesosoma. O9–O13 narrow to form the characteristic metasoma, which bears the venom glands and the aculeus (stinger).

Among the arthropods, only insects, some crustaceans and scorpions bear three tagmata (reviewed by [4]). Yet it is

unknown how Hox genes regionalize the three tagmata in scorpions. Only the collective protein expression of *Ubx* and *abd-A* (termed 'UbdA') has been surveyed in the scorpion *Smeringurus mesaensis*, which shows an anterior expression boundary in part of the O2 segment, comparable to harvestmen [28,30]. Scorpion life-history characteristics, including live birth in all extant species and gestation periods lasting several months have heretofore precluded extensive developmental genetic study in this lineage [30–33]. Here, we investigated the genetic basis of posterior tagmosis in the bark scorpion *Centruroides sculpturatus*. We show that *C. sculpturatus* has two paralogues of every Hox gene class excepting *Hox3*, and thus possesses a total of 19 Hox genes. We show that the paralogues of the posterior class Hox genes each have distinct expression patterns that result in unique combinations of Hox transcripts for each segmental identity. The paralogues retain archetypal Hox gene spatial collinearity, reconciling the heteronomy of scorpions with established mechanisms of Hox gene regulation in Arthropoda.

2. Material and methods

Adult females of the scorpion *C. sculpturatus* were purchased from Hatari Invertebrates (AZ, USA). Females were anaesthetized with CO_2 and embryos dissected from the ovary following a published protocol [31]. RNA was extracted from one clutch of limb bud stage embryos using Trizol (Invitrogen) and first strand cDNA synthesis was performed using SuperScriptIII (Invitrogen). A developmental transcriptome of *C. sculpturatus* was generated by sequencing this cDNA in a single flowcell on an Illumina HiSeq 2500 platform subsequent to strand-specific library preparation with the Apollo 324TM system (IntegenX), using paired-end 150 bp-long reads at the Harvard University Bauer Core Facility. Terminal bases with a Phred quality score under 33 were trimmed and over-represented rRNA sequences were discarded. Assembly was conducted using the software Trinity with a path reinforcement distance of 75 and enforcing strand-specific assembly, with a resulting estimated coverage of 23.9× [34].

Fragments of Hox genes were identified by BLAST. Phylogenetic analysis of Hox gene amino acid sequences was conducted as follows: amino acid sequences of chelicerate and mandibulate Hox orthologues with known expression patterns were aligned using Muscle v. 3.6 [35] and culled to 68 conserved, adjacent positions using GBlocks v. 0.91b [36]. Maximum-likelihood analysis was performed using RAxML v. 7.7.5 [37] and Bayesian inference analysis using MrBayes v. 3.1.2 [38]. Details of heuristics and all multiple sequence alignments are provided in the electronic supplementary material.

Templates for riboprobe synthesis were generated by PCR-amplified gene-specific primers, and cloning amplicons using the TOPO TA Cloning Kit with One Shot Top10 chemically competent *Escherichia coli* (Invitrogen, Carlsbad, CA, USA), following the manufacturer's protocol. Amplicon identities were verified by Sanger sequencing. Primer sequences for templates are provided in the electronic supplementary material, table S1. The developmental transcriptome is deposited in the NCBI Sequence Read Archive, accession no. SRR1515193.

Centruroides sculpturatus embryos were fixed and whole mount *in situ* hybridization performed following a published protocol [39]. Approximately 15–20 embryos were assayed for every paralogue. For each pair of paralogues, clutches were divided and expression examined in equivalent stages. Embryos were mounted in glycerol and images were captured using an HrC AxioCam and an Axio Zoom V.16 fluorescence stereomicroscope driven by Zen (Zeiss).

3. Results

(a) Identification of scorpion Hox genes

Phylogenetic analysis of a 68 site conserved region of Hox gene amino acid sequences obtained from a *C. sculpturatus* transcriptome was conducted using maximum-likelihood and Bayesian inference analyses. Tree topologies obtained indicated exactly two copies of all Hox genes in *C. sculpturatus*, except for *Hox3*, for which a single copy was recovered (figure 2). Among chelicerates, duplicates of Hox genes have been previously reported for the spiders *Cupiennius salei* and *P. tepidariorum*, in the genes *labial* (*lab*), *pb*, *Dfd*, *Scr* and *Ubx*; and the horseshoe crab *Limulus polyphemus*, in which up to four putative copies occur of some Hox genes [11,40] (electronic supplementary material, table S2). To assess whether the scorpion and spider paralogues represent lineage-specific duplications, we constructed larger nucleotide sequence alignments for *lab*, *Dfd*, *Scr* and *Ubx* using spider, scorpion and harvestman sequences only. Consideration of only closely related chelicerate Hox sequences enabled retention of more sites for analysis upon masking of ambiguously aligned sites, as non-chelicerate sequences diverge greatly outside of the conserved region. We included paralogues of some Hox genes found in the genome of the scorpion *Mesobuthus martensii* [25] (electronic supplementary material, table S2). In no gene tree do scorpion and spider orthologues form paralogue-specific clusters (figure 3). Instead, gene trees generated from these alignments showed that the spider and scorpions sequences frequently form clusters corresponding to their respective orders when rooted with orthologous harvestman sequences, supporting lineage-specific duplications in both spiders and scorpions (figure 3).

(b) Expression of *Cscu-Antp* paralogues

The anterior expression boundary of *Cscu-Antp-1* occurs in the L4 (fourth walking leg) segment, including in the limb buds (figure 4*c*). By contrast, the anterior expression boundary of *Cscu-Antp-2* occurs in the posterior part of the L4 segment, and it is absent from the L4 limb buds (figure 4*d*; electronic supplementary material, figure S1). In other chelicerates, the expression domain boundaries of the opisthosomal Hox group genes (*Antp*, *Ubx*, *abd-A* and *Abd-B*) extend to the posterior terminus of the embryo [10,28,29,40,41]. Consistent with this pattern, expression of *Cscu-Antp-1* and *Cscu-Antp-2* extends to the posterior (figure 4*a*,*b*). *Cscu-Antp-1* is expressed at higher levels in O1 and O2 than in O3 through to O13, with the strongest expression occurring in the O2 neural lobes. Expression of *Cscu-Antp-2* in posterior segments is more uniform than that of *Cscu-Antp-1*. Both genes are expressed in the neuroectoderm along the ventral midline.

(c) Expression of *Cscu-Ubx* paralogues

Distinct anterior expression boundaries of *Cscu-Ubx-1* and *Cscu-Ubx-2* occur in ventral portions of O1 and O2, respectively (figure 4*e*,*f*). In *Cscu-Ubx-1*, the anterior boundary comprises a single pair of small domains in the posterior of O1, whereas the anterior expression boundary of *Cscu-Ubx-2* consists of a pair of hemispherical domains in O2 that abut the O1-O2 boundary. *Cscu-Ubx-2* expression is stronger in the opisthosomal organs and ventral ectoderm than in the pleura, and both orthologues show similar levels of expression in posterior segments.

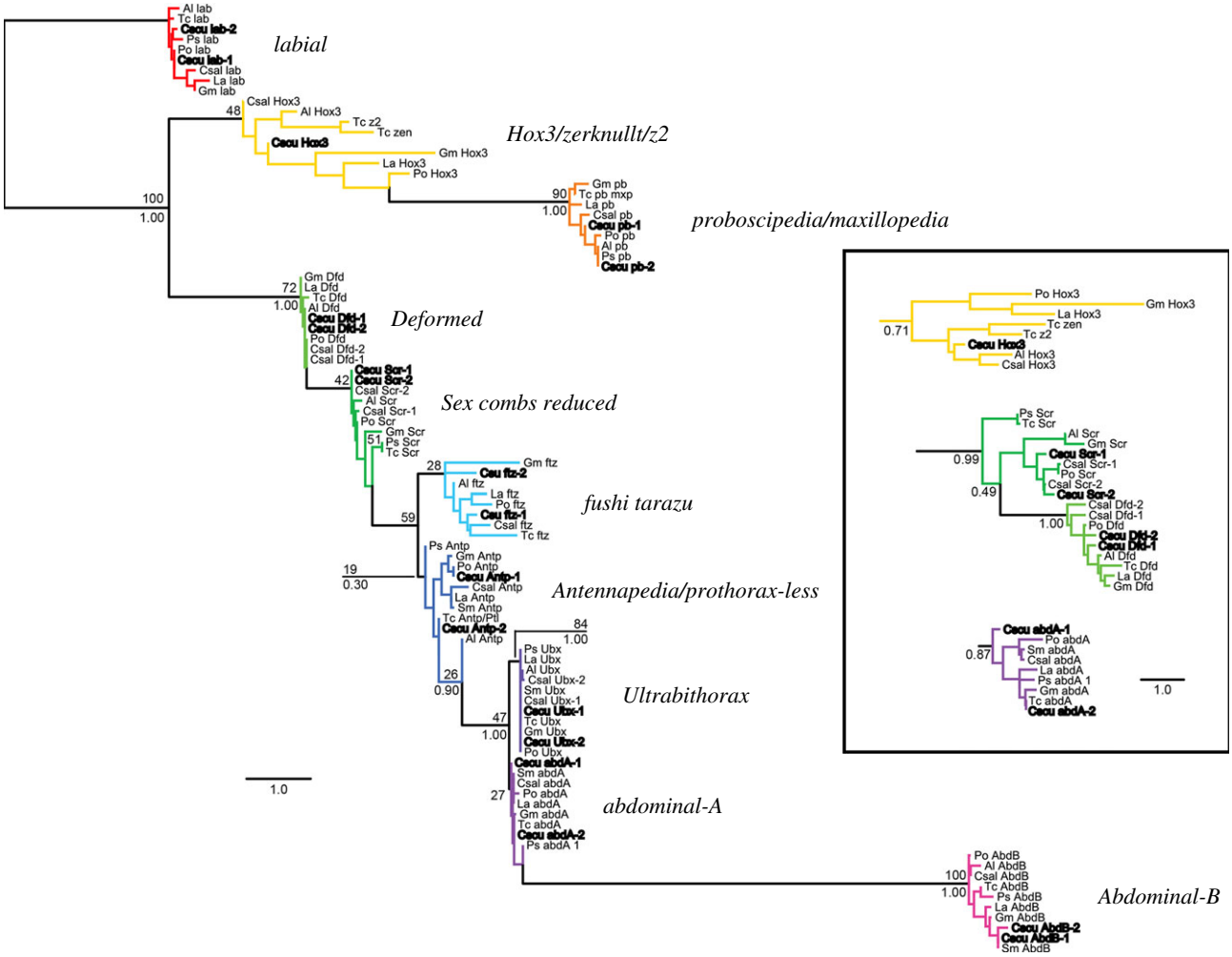

Figure 2. Tree topology inferred from maximum-likelihood analysis of conserved regions (68 amino acid characters), using selected arthropod taxa for which gene expression has been reported (ln $L = -2074.37$). *Centruroides sculpturatus* sequences are indicated in bold. Colours correspond to Hox gene identities, as in figure 1. Inset: alternative topologies of *Hox3*, *Scr*, *Dfd* and *abd-A* subtrees based on Bayesian inference analyses. Values on nodes indicate bootstrap resampling frequencies (integers) and posterior probabilities (values ≤ 1.00). (Online version in colour.)

(d) Expression of *Cscu-abd-A* paralogues

Anterior expression boundaries of *Cscu-abd-A-1* and *Cscu-abd-A-2* occur in O3 and O4, respectively (figure 4*g*,*h*). Clear expression of *Cscu-abd-A-1* is observed in the developing pectines, the largest of the embryonic opisthosomal organs in this species. *Cscu-abd-A-2* is expressed in the first book lung primordia in O4. In posterior segments, expression level of *Cscu-abd-A-1* is higher in the metasoma than in the mesosoma. Both *Cscu-abd-A-1* and *Cscu-abd-A-2* are strongly expressed in the metasoma.

(e) Expression of *Cscu-Abd-B* paralogues

The strong anterior expression boundary of *Cscu-Abd-B-1* in the metasomal segments occurs in O8, the recurved mesosomal segment that occurs in the posterior of the scorpion embryo (but not at the terminus of the AP axis). By contrast, the strong anterior expression boundary of *Cscu-Abd-B-2* occurs in the posterior part of O9 (figure 4*i*,*j*). Restricted and unique expression domains of each *Abd-B* paralogue are also observed in the ventral ectoderm of mesosomal segments, as follows. *Cscu-Abd-B-1* is additionally expressed in the ventral neuroectoderm of O3–O7, with expression restricted to the medial part of the neuroectoderm. By contrast, additional expression of *Cscu-Abd-B-2* occurs in the ventral ectoderm of O4–O7, but in domains that comprise most of the neuroectoderm. Expression in the book lungs is observed in later stages (electronic supplementary material, figure S2). Both paralogues are uniformly expressed in the metasoma posterior to O9 (figure 4*i*,*j*). Expression of neither paralogue was observed in the genital segment (only stages prior to formation of genital pores were examined in this study).

4. Discussion

The degree of tagmatization occurring in the scorpion opisthosoma is exceptional among chelicerates. Our gene expression data demonstrate that both paralogues of each of the last four Hox genes are expressed in a canonical Hox-like pattern in scorpions. The staggering of the anterior expression boundaries, resulting from the offset of the anterior expression boundaries by one or more segments, establishes a unique combination of Hox transcripts for each segmental identity (figure 5*a*).

Expression data in the spider and the harvestman implicate roles for *abd-A* and/or *Abd-B* in opisthosomal differentiation [28]. Apropos, offsets in anterior expression domains among *Cscu-abd-A* and *Cscu-Abd-B* paralogues appear to delimit the segmental identities of the scorpion pectine- and book lung-bearing segments (O3 and O4–O7, respectively). The weaker expression domains of the *Cscu-Abd-B* paralogues are similar to their counterparts in the spider *C. salei*, wherein *Abd-B* is

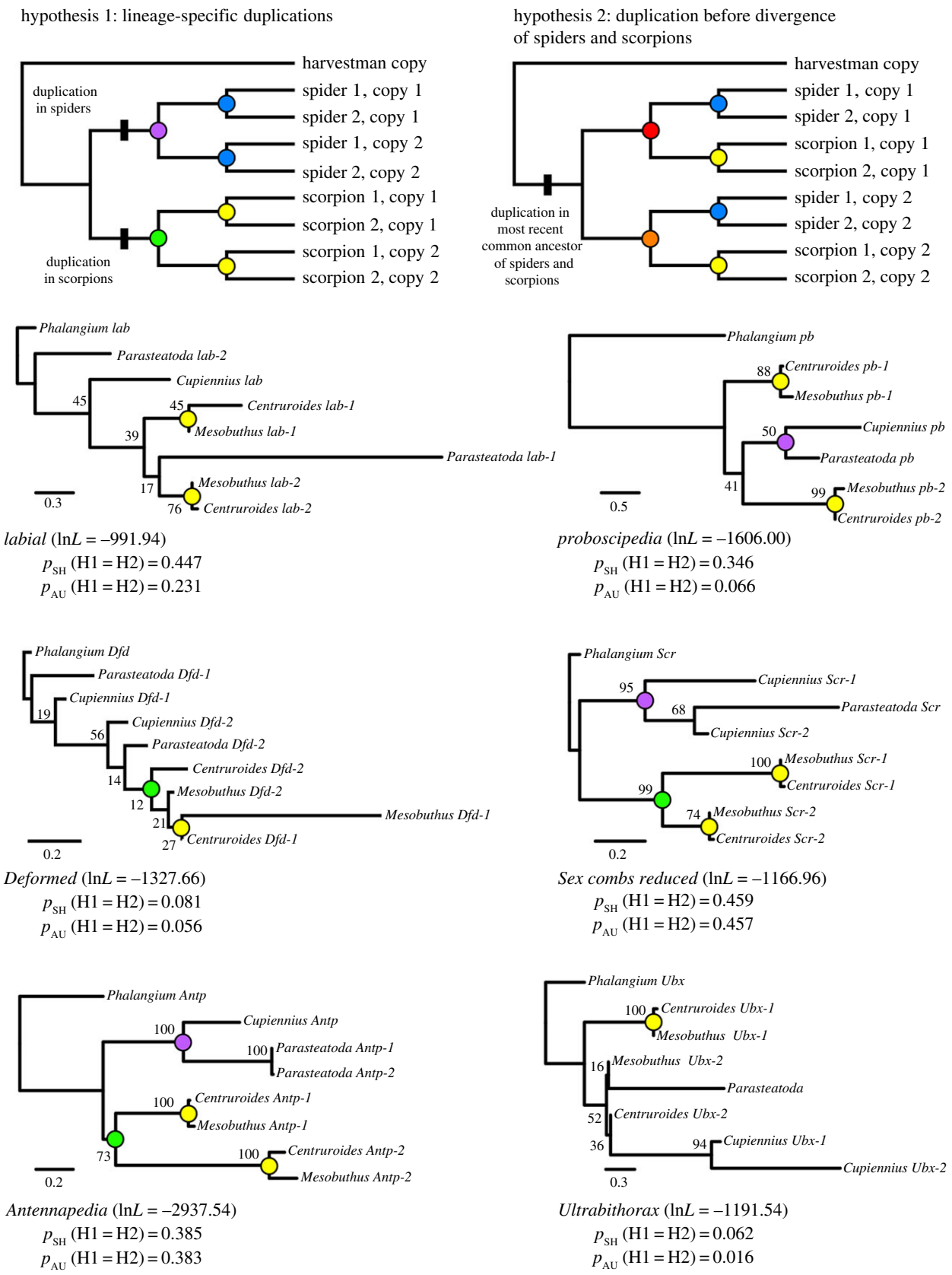

Figure 3. Chelicerate Hox gene trees inferred from maximum-likelihood analysis of nucleotide sequences, wherein multiple paralogues are present in scorpions, spiders or both. Hypothetical trees in the top row indicate alternative scenarios of lineage-specific duplications (left) and a single duplication in the common ancestor of spiders and scorpions (right). Coloured circles indicate nodes expected under a particular hypothesis. Pairs of *p*-values under each gene tree indicate results of Shimodaira-Hasegawa (SH) and approximately unbiased (AU) tests. Harvestman: *P. opilio*; spiders: *C. salei* and *P. tepidariorum*; scorpions: *C. sculpturatus* and *M. martensii*. (Online version in colour.)

strongly expressed throughout posterior segments (O5–O10), and weakly in O3–O4 [28]. The functional significance of the weaker *Abd-B* domains is not known in chelicerates [10,29].

Paralogues of all four scorpion opisthosomal Hox genes follow spatial collinearity. However, protein expression of UbdA in another scorpion species (*S. mesaensis*) shows an anterior expression boundary similar to the one observed for *Cscu-Ubx-2* in early stages (i.e. in the ventral part of O2, with expression throughout O3 and posterior segments; compare figure 4*f* to fig. 2a of [30]). In late stages of *S. mesaensis*,

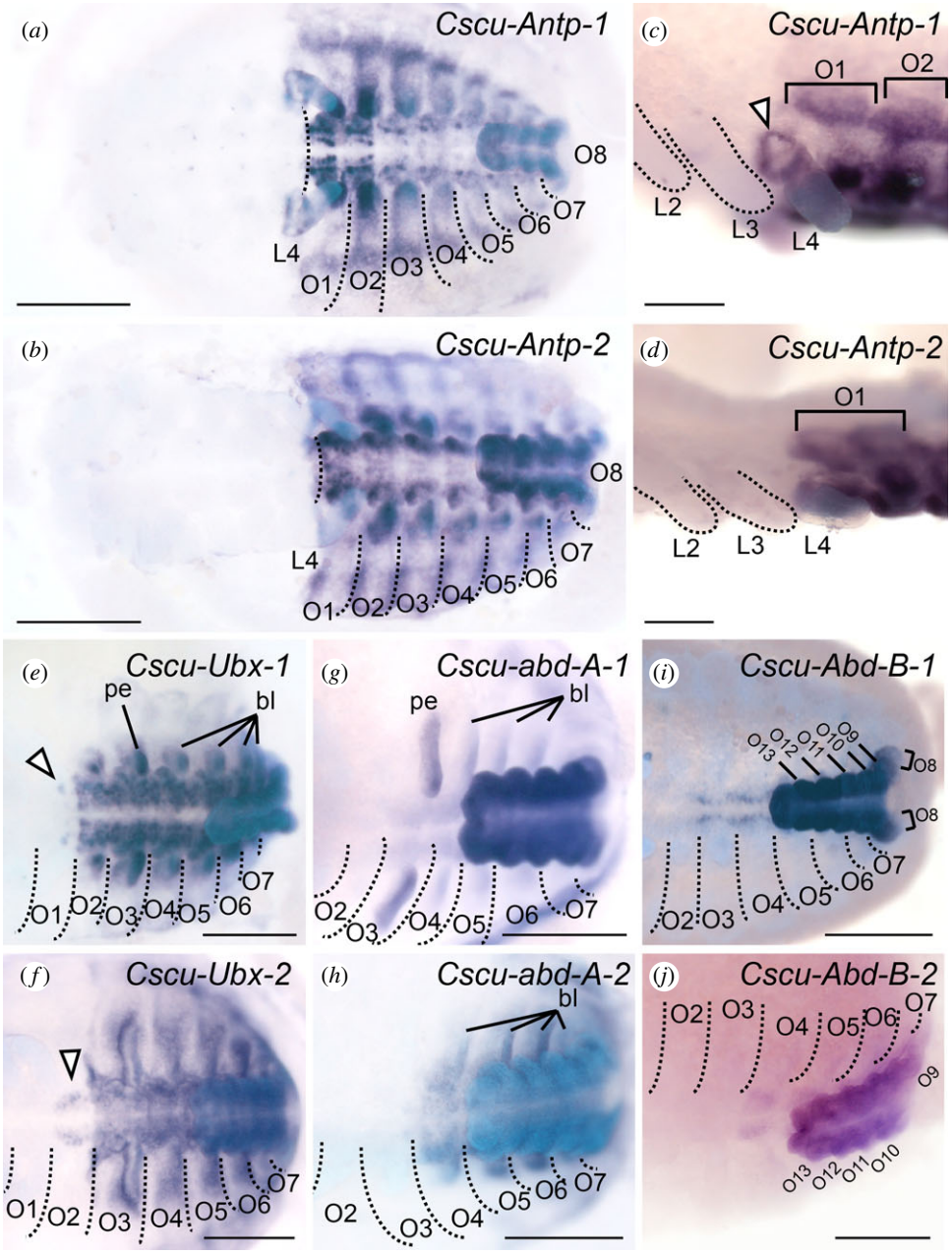

Figure 4. Staggered expression of opisthosomal Hox group paralogues yields unique Hox transcript combinations for every segmental identity. Embryos are shown in ventral view unless otherwise specified. (*a*,*b*) Expression of *Cscu-Antp-1* and *Cscu-Antp-2* spans L4 to posterior terminus. (*c*,*d*) Higher magnification lateral views of embryos in (*a*,*b*). Arrowheads indicate unique anterior expression domains of *Cscu-Antp-1* in the L4 limb bud, and *Cscu-Antp-2* in the posterior part of the L4 neural lobe only. (*e*,*f*) Expression of *Cscu-Ubx-1* and *Cscu-Ubx-2* shows anterior expression boundaries offset by one segment, i.e. posterior and ventral part of O1, and ventral part of O2, respectively. Arrowheads show expression in a more restricted area as compared with posterior segments. (*g*,*h*) Like the *Ubx* paralogues, the anterior expression boundaries of *Cscu-abd-A-1* and *Cscu-abd-A-2* are offset by a single segment, in O3 and O4, respectively. (*i*,*j*) *Cscu-Abd-B-1* and *Cscu-Abd-B-2* are strongly expressed starting from O8 (brackets in (*i*) show upturning O8 segment) and from O9, respectively. Weaker expression domains occur along the ventral midline in both paralogues, but in distinct tissues. pe, pectine; bl, book lung. Scale bars: 200 µm (*c*,*d*); 400 µm (*e*–*j*); and 500 µm (*a*,*b*). (Online version in colour.)

expression of UbdA is similar to that of *Cscu-Ubx-1* (i.e. with strong expression in O2 and posterior segments; compare figure 4*e* to fig. 2c of [30], but includes more tissue in the O1 segment. It is unclear whether these differences in expression correspond to differences in embryonic stage or differences between the two lineages.

In the harvestman *Phalangium opilio*, a chelicerate with putatively single orthologues of *Antp* and *Ubx*, the anterior expression boundary of both Hox genes shifts during development [28]. Early expression of *Antp* commences in O1, but expands into the L4 segment (including the limb bud) in later stages [28]; the same phenomenon occurs with spiders' *Antp* [10,27]. Curiously, in *C. sculpturatus*, the *Antp* paralogues' expression domains resemble the two temporally distinct expression profiles of the single harvestman and spider *Antp* orthologues. Similarly, *Ubx* paralogues of both spiders and scorpions have offset anterior expression boundaries (although neither spider *Ubx* paralogue is expressed in O1 [40]), whereas *Ubx* expression in the harvestman commences in the posterior part of O2 but shifts forward in later stages [28].

These data are suggestive of subfunctionalization of *Antp* in scorpions and of *Ubx* in both scorpions and spiders. A similar pattern is observed in spider and harvestman *Scr*; the harvestman *Scr* orthologue's expression domain resembles the union of expression domains of the two *Scr* paralogues of *C. salei* [28,40]. However, we add the caveat that gene silencing of

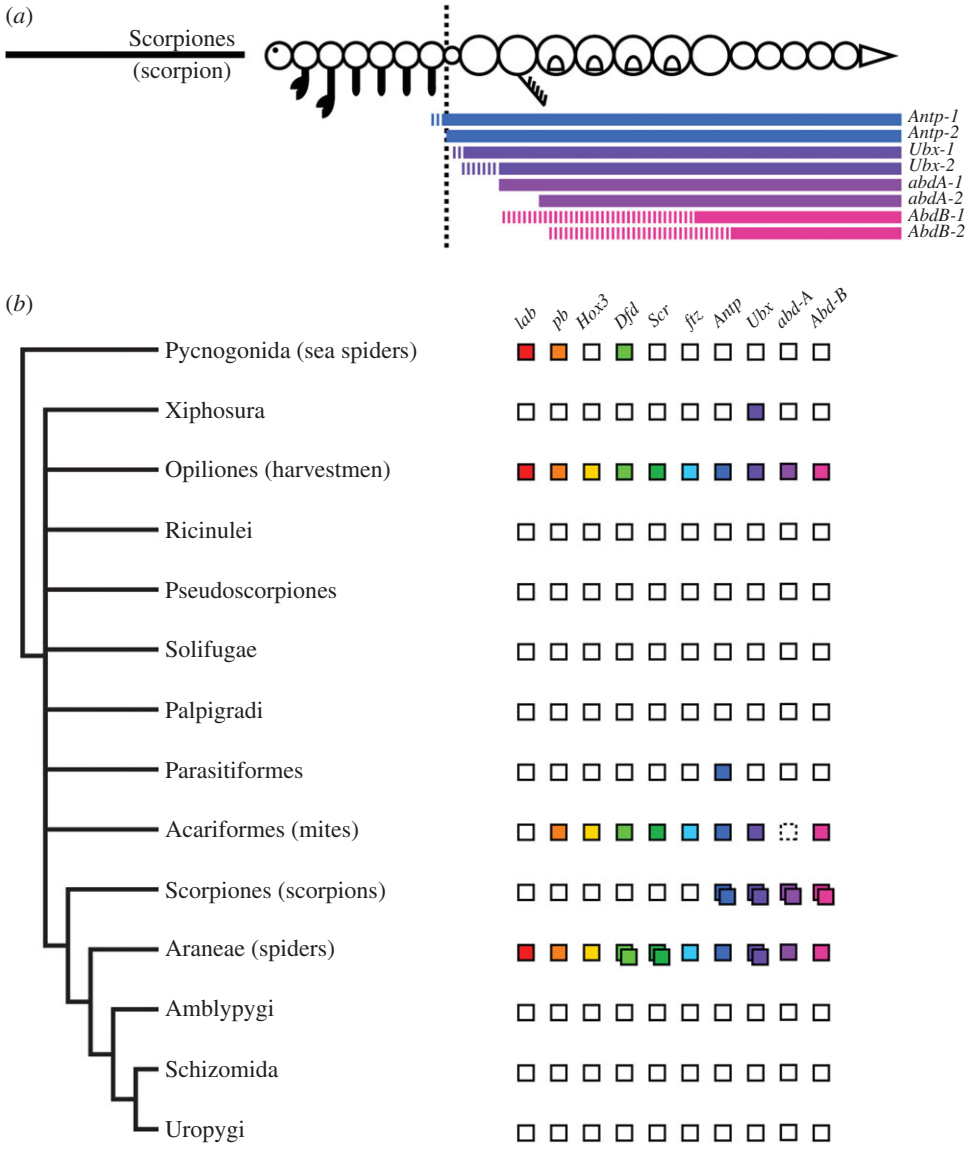

Figure 5. (*a*) Summary of opisthosomal Hox expression patterns in *C. sculpturatus*. (*b*) Summary of known Hox expression patterns in Chelicerata. Coloured squares indicate known expression domains. Dashed square for acariform *abd-A* indicates gene loss. Duplicated genes without verified expression data not indicated. Tree topology based on [42]. (Online version in colour.)

spider *Antp*, the only functional data available to date for chelicerate Hox genes, affects only the O1 segment, de-repressing appendage formation [10]. No *Antp* phenotype has been reported in L4, and it is therefore not clear what, if anything, is achieved by expansion of *Antp* expression into the L4 segment, which has now been reported in two spiders (*C. salei* and *P. tepidariorum*), a mite and a harvestman [12,27,28,40,43]. Furthermore, phylogenetic placements of scorpion and spider Hox paralogues, when these are retained and topologically informative (specifically, *Dfd*, *Scr* and *Antp*), do not support with statistical significance a scenario of lineage-specific duplications (figure 3). Likelihood-based tests of tree topology demonstrate that neither lineage-specific duplications nor a common duplication event in the most recent common ancestor of scorpions and spiders is the topologically superior scenario (at $\alpha = 0.01$) in any of the six test cases (figure 3 and table 1; electronic supplementary material). A lack of statistical significance in this case may be attributable to: (i) the paucity of sampled lineages, (ii) rate heterogeneity in one or more arachnid species, (iii) the degree of evolutionary conservation of the homeobox domain in multiple sequence alignments, or (iv) some combination of these systematic biases. In spite of accruing genomic resources for arachnids, we were unable to discover duplicated scorpion Hox genes in published databases, as well as our own unpublished scorpion libraries (P.P. Sharma and W.C. Wheeler 2014, personal communication), particularly because no other embryonic transcriptomes have been generated for this group of chelicerates.

It therefore remains to be investigated how many other Hox paralogues have been retained in spider genomes, and what function the spider paralogues serve during embryogenesis. Similarly, the degree of retention of Hox gene paralogues within Scorpiones cannot presently be assessed because both *C. sculpturatus* and *M. martensii* are members of the same family of scorpions (Buthidae), and basal relationships of the extant scorpions are unresolved [44,45]. Given that both spiders and scorpions have a conserved prosoma, but a more differentiated opisthosoma, than other euchelicerates, it is possible that some prosomal Hox paralogues in spiders and scorpions have been lost or undergone subfunctionalization, whereas some opisthosomal paralogues have neofunctionalized. The recently sequenced genome of the scorpion *M. martensii* is reported to bear over 32 000 genes, more than any other arthropod genome, and demonstrably accelerated gene family turnover,

Table 1. Likelihood scores for Hox gene tree alignments with topologies unconstrained or constrained to a specific scenario (as in figure 3). (Difference in log likelihoods and standard deviation of this value are indicated for the comparison of hypothesis 1 (lineage-specific duplication) and hypothesis 2 (duplication in most recent common ancestor of spiders and scorpions).)

	ln*L* (unconstrained)	ln*L* (Hypothesis 1)	ln*L* (Hypothesis 2)	$\lvert\Delta(\ln L_{H1,\ H2})\rvert$	σ (ln*L*)
labial	−991.94	−1090.34	−1094.68	4.332493	6.291582
proboscipedia	−1606.00	−1715.38	−1717.49	2.112443	2.312595
Deformed	−1327.66	−1447.39	−1463.84	16.450287	11.739177
Sex combs reduced	−1166.96	−1284.29	−1292.70	8.406873	4.858968
Antennapedia	−2937.54	−3157.80	−3158.49	0.694105	1.626799
Ultrabithorax	−1191.54	−1616.86	−1613.57	3.292049	4.592267

but a putative whole genome duplication in scorpions has not been assessed [25].

An alternative route to testing a common mechanism of Hox gene duplications in spiders and scorpions may be concerted expansion of available genomic resources for other arachnid orders. Specifically, the discovery of Hox gene paralogues in non-spider tetrapulmonates (Amblypygi, Schizomida and Thelyphonida) would empower statistical assessments of the likelihood of a common duplication mechanism. Moreover, such a discovery could lend credence to the recently proposed sister relationship between scorpions and Tetrapulmonata [42].

5. Conclusion

We propose that the duplication of scorpion Hox genes and putative neofunctionalization of paralogues of the opisthosomal group facilitated the extreme heteronomy of scorpions, the only group of arthropods with a tagma dedicated exclusively to prey capture and defence. Our results suggest that gene duplications may have been a plausible mechanism for achieving morphological diversification in this ancient group of arthropods.

Acknowledgements. We are indebted to Philip Donoghue and two anonymous reviewers whose comments greatly improved an earlier draft of this manuscript. Douglas Richardson facilitated imaging at the Harvard Center for Biological Imaging. P.P.S. designed and performed the research, analysed data and wrote the paper. E.E.S. performed research and analysed data. C.G.E. contributed reagents. W.W. contributed analytical tools. All authors edited and approved the manuscript.

Funding statement. This work was partially supported by NSF grant no. IOS-1257217 to C.G.E. and AMNH funds to W.W. P.P.S. was supported by the National Science Foundation Postdoctoral Research Fellowship in Biology under grant no. DBI-1202751.

References

1. Cisne JL. 1974 Evolution of the world fauna of aquatic free-living arthropods. *Evolution* **28**, 337–366. (doi:10.2307/2407157)
2. Carroll SB. 1995 Homeotic genes and the evolution of arthropods and chordates. *Nature* **376**, 479–485. (doi:10.1038/376479a0)
3. Popadić A, Abzhanov A, Rusch D, Kaufman TC. 1998 Understanding the genetic basis of morphological evolution: the role of homeotic genes in the diversification of the arthropod bauplan. *Int. J. Dev. Biol.* **42**, 453–461.
4. Hughes CL, Kaufman TC. 2002 Hox genes and the evolution of the arthropod body plan. *Evol. Dev.* **4**, 459–499. (doi:10.1046/j.1525-142X.2002.02034.x)
5. Lewis EB. 1978 A gene complex controlling segmentation in *Drosophila*. *Nature* **276**, 565–570. (doi:10.1038/276565a0)
6. Denell RE, Hummels KR, Wakimoto BT, Kaufman TC. 1981 Developmental studies of lethality associated with the *Antennapedia* gene complex in *Drosophila melanogaster*. *Dev. Biol.* **81**, 43–50. (doi:10.1016/0012-1606(81)90346-8)
7. Pultz MA, Diederich RJ, Cribbs DL, Kaufman TC. 1988 The *proboscipedia* locus of the Antennapedia complex: a molecular and genetic analysis. *Genes Dev.* **2**, 901–920. (doi:10.1101/gad.2.7.901)
8. Angelini DR, Liu PZ, Hughes CL, Kaufman TC. 2005 Hox gene function and interaction in the milkweed bug *Oncopeltus fasciatus* (Hemiptera). *Dev. Biol.* **287**, 440–455. (doi:10.1016/j.ydbio.2005.08.010)
9. Liubicich DM *et al*. 2009 Knockdown of *Parhyale Ultrabithorax* recapitulates evolutionary changes in crustacean appendage morphology. *Proc. Natl Acad. Sci. USA* **106**, 13 892–13 896. (doi:10.1073/pnas.0903105106)
10. Khadjeh S *et al*. 2012 Divergent role of the Hox gene *Antennapedia* in spiders is responsible for the convergent evolution of abdominal limb repression. *Proc. Natl Acad. Sci. USA* **109**, 4921–4926. (doi:10.1073/pnas.1116421109)
11. Cook CE *et al*. 2001 Hox genes and the phylogeny of the arthropods. *Curr. Biol.* **11**, 759–763. (doi:10.1016/S0960-9822(01)00222-6)
12. Telford MJ, Thomas RH. 1998 Expression of homeobox genes shows chelicerate arthropods retain their deutocerebral segment. *Proc. Natl Acad. Sci. USA* **95**, 10 671–10 675. (doi:10.1073/pnas.95.18.10671)
13. Telford MJ, Thomas RH. 1998 Of mites and *zen*: expression studies in a chelicerate arthropod confirm *zen* is a divergent Hox gene. *Dev. Genes Evol.* **208**, 591–594. (doi:10.1007/s004270050219)
14. Janssen R, Eriksson BJ, Tait NN, Budd GE. 2014 Onychophoran Hox genes and the evolution of arthropod Hox gene expression. *Front. Zool.* **11**, 22. (doi:10.1186/1742-9994-11-22)
15. Sánchez-Herrero E, Vernos I, Marco R, Morata G. 1985 Genetic organization of *Drosophila* bithorax complex. *Nature* **313**, 108–113. (doi:10.1038/313108a0)
16. Kaufman TC, Seeger MA, Olsen G. 1990 Molecular and genetic organization of the *Antennapedia* gene complex of *Drosophila melanogaster*. *Adv. Gen.* **27**, 309–362. (doi:10.1016/S0065-2660(08)60029-2)
17. González-Reyes A, Morata G. 1990 The developmental effect of overexpressing a Ubx product in *Drosophila* embryos is dependent on its interactions with other homeotic products. *Cell* **61**, 515–522. (doi:10.1016/0092-8674(90)90533-K)
18. Averof M, Patel NH. 1997 Crustacean appendage evolution associated with changes in Hox gene expression. *Nature* **388**, 682–686. (doi:10.1038/41786)
19. Abzhanov A, Kaufman TC. 2000 Crustacean (malacostracan) Hox genes and the evolution of the arthropod trunk. *Development* **127**, 2239–2249.
20. Tomoyasu Y, Wheeler SR, Denell RE. 2005 *Ultrabithorax* is required for membranous wing

identity in the beetle *Tribolium castaneum*. *Nature* **433**, 643–647. (doi:10.1038/nature03272)

21. Pavlopoulos A *et al*. 2009 Probing the evolution of appendage specialization by Hox gene misexpression in an emerging model crustacean. *Proc. Natl Acad. Sci. USA* **106**, 13 897–13 902. (doi:10.1073/pnas.0902804106)
22. Zwicky KT. 1968 A light response in the tail of *Urodacus*, a scorpion. *Life Sci.* **7**, 257–262. (doi:10.1016/0024-3205(68)90020-9)
23. Zwicky KT. 1970 The spectral sensitivity of the tail of *Urodacus*, a scorpion. *Experientia* **26**, 317. (doi:10.1007/BF01900120)
24. Rao G, Rao KP. 1973 A metasomatic neuronal photoreceptor in the scorpion. *J. Exp. Biol.* **58**, 189–196.
25. Cao Z *et al*. 2013 The genome of *Mesobuthus martensii* reveals a unique adaptation model of arthropods. *Nat. Commun.* **4**, 2602. (doi:10.1038/ncomms3602)
26. Scholtz G, Edgecombe GD. 2006 The evolution of arthropod heads: reconciling morphological, developmental and palaeontological evidence. *Dev. Genes Evol.* **216**, 395–415. (doi:10.1007/s00427-006-0085-4)
27. Damen WGM, Hausdorf M, Seyfarth E-A, Tautz D. 1998 A conserved mode of head segmentation in arthropods revealed by the expression pattern of Hox genes in a spider. *Proc. Natl Acad. Sci. USA* **95**, 10 665–10 670. (doi:10.1073/pnas.95.18.10665)
28. Sharma PP, Schwager EE, Extavour CG, Giribet G. 2012 Hox gene expression in the harvestman *Phalangium opilio* reveals divergent patterning of the chelicerate opisthosoma. *Evol. Dev.* **14**, 450–463. (doi:10.1111/j.1525-142X.2012.00565.x)
29. Damen WGM, Tautz D. 1999 *Abdominal-B* expression in a spider suggests a general role for *Abdominal-B* in specifying the genital structure. *J. Exp. Zool.* **285**, 85–91. (doi:10.1002/(SICI)1097-010X(19990415)285:1<85::AID-JEZ10>3.0.CO;2-N)
30. Popadić A, Nagy L. 2001 Conservation and variation in *Ubx* expression among chelicerates. *Evol. Dev.* **3**, 391–396. (doi:10.1046/j.1525-142X.2001.01049.x)
31. Farley RD. 2001 Development of segments and appendages in embryos of the desert scorpion *Paruroctonus mesaensis* (Scorpiones: Vaejovidae). *J. Morphol.* **250**, 70–88. (doi:10.1002/jmor.1060)
32. Simonnet F, Deutsch J, Quéinnec E. 2004 *hedgehog* is a segment polarity gene in a crustacean and a chelicerate. *Dev. Genes Evol.* **214**, 537–545. (doi:10.1007/s00427-004-0435-z)
33. Simonnet F, Célérier M-L, Quéinnec E. 2006 *Orthodenticle* and *empty spiracles* genes are expressed in a segmental pattern in chelicerates. *Dev. Genes Evol.* **216**, 467–480. (doi:10.1007/s00427-006-0093-4)
34. Grabherr MG *et al*. 2011 Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–652. (doi:10.1038/nbt.1883)
35. Edgar RC. 2004 Muscle: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797. (doi:10.1093/nar/gkh340)
36. Castresana J. 2000 Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* **17**, 540–552. (doi:10.1093/oxfordjournals.molbev.a026334)
37. Stamatakis A. 2006 RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688–2690. (doi:10.1093/bioinformatics/btl446)
38. Huelsenbeck JP, Ronquist F. 2005 Bayesian analyses of molecular evolution using MrBayes. In *Statistical methods in molecular evolution* (ed. R Nielsen), pp. 1–52. New York, NY: Springer.
39. Sharma PP, Gupta T, Schwager EE, Wheeler WC, Extavour CG. 2014 Subdivision of *cap-n-collar* domains is restricted to Mandibulata. *EvoDevo* **5**, 3. (doi:10.1186/2041-9139-5-3)
40. Schwager EE, Schoppmeier M, Pechmann M, Damen WGM. 2007 Duplicated Hox genes in the spider *Cupiennius salei*. *Front. Zool.* **4**, 10. (doi:10.1186/1742-9994-4-10)
41. Barnett AA, Thomas RH. 2013 Posterior Hox gene reduction in an arthropod: *Ultrabithorax* and *Abdominal-B* are expressed in a single segment in the mite *Archegozetes longisetosus*. *EvoDevo* **4**, 23. (doi:10.1186/2041-9139-4-23)
42. Regier JC, Shultz JW, Zwick A, Hussey A, Ball B, Wetzer R, Martin JW, Cunningham CW. 2010 Arthropod relationships revealed by phylogenomic analysis of nuclear protein-coding sequences. *Nature* **463**, 1079–1083. (doi:10.1038/nature08742)
43. Schwager EE, Pechmann M, Feitosa NM, McGregor AP, Damen WGM. 2009 *hunchback* functions as a segmentation gene in the spider *Achaearanea tepidariorum*. *Curr. Biol.* **19**, 1333–1340. (doi:10.1016/j.cub.2009.06.061)
44. Prendini L, Volschenk ES, Maaliki S, Gromov AV. 2006 A 'living fossil' from Central Asia: the morphology of *Pseudochactas ovchinnikovi* Gromov, 1998 (Scorpiones: Pseudochactidae), with comments on its phylogenetic position. *Zool. Anz.* **245**, 211–248. (doi:10.1016/j.jcz.2006.07.001)
45. Soleglad ME, Fet V. 2003 High-level systematics and phylogeny of the extant scorpions (Scorpiones: Orthosterni). *Euscorpius* **11**, 1–57.