Hox gene duplications correlate with posterior heteronomy in scorpions

Prashant P. Sharma¹, Evelyn E. Schwager², Cassandra G. Extavour² and Ward C. Wheeler¹

¹Division of Invertebrate Zoology, American Museum of Natural History, Central Park West at 79th Street, New York, NY 10024, USA
²Department 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 Euchelicerata, 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
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. 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, stemum. Scale bar in (e): 200 μm. (Online version in colour.)

The 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 1a) [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 1b) [28]. In spiders, the differentiation of the opisthosoma into segments forming the pedicel (O1), the respiratory organs (book lungs and/or tracheal tubes; 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 1c) [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 1d). 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 1f,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 \textit{Ubx} and \textit{abd-A} (termed ‘\textit{UbdA}’) has been surveyed in the scorpion \textit{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 \textit{Centruroides sculpturatus}. We show that the paralogues of the posterior class Hox genes except the \textit{labial} (lab), \textit{pb}, \textit{Dfd}, \textit{Scr} and \textit{Ubx} and the horseshoe crab \textit{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 \textit{lab}, \textit{Dfd}, \textit{Scr} and \textit{Ubx} using spider, scorpion and harvestman sequences. 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 \textit{Mesobuthus martensii} [25] (electronic supplementary material, table S2). In no gene tree do scorpion and spider orthologues form parologue-specific clusters (figure 3). Instead, gene trees generated from these alignments showed that the spider and scorpion 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 \textit{Cscu-Antp} paralogues

The anterior expression boundary of \textit{Cscu-Antp-1} occurs in the L4 (fourth walking leg) segment, including in the limb buds (figure 4c). By contrast, the anterior expression boundary of \textit{Cscu-Antp-2} occurs in the posterior part of the L4 segment, and it is absent from the L4 limb buds (figure 4d; electronic supplementary material, figure S1). In other chelicerates, the anterior expression boundary of the opisthosomal Hox group genes (\textit{Antp}, \textit{Ubx}, \textit{abd-A} and \textit{abd-B}) extend to the posterior terminus of the embryo [10,28,29,40,41]. Consistent with this pattern, expression of \textit{Cscu-Antp-1} and \textit{Cscu-Antp-2} extends to the posterior (figure 4e,h). \textit{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 \textit{Cscu-Antp-2} in posterior segments is more uniform than that of \textit{Cscu-Antp-1}. Both genes are expressed in the neuroectoderm along the ventral midline.

(c) Expression of \textit{Cscu-Ubx} paralogues

Distinct anterior expression boundaries of \textit{Cscu-Ubx-1} and \textit{Cscu-Ubx-2} occur in ventral portions of O1 and O2, respectively (figure 4e,f). In \textit{Cscu-Ubx-1}, the anterior boundary comprises a single pair of small domains in the posterior of O1, whereas the anterior expression boundary of \textit{Cscu-Ubx-2} consists of a pair of hemispherical domains in O2 that abut the O1-O2 boundary. \textit{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.

(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 4i,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 4i,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 5a). 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
hypothesis 1: lineage-specific duplications

hypothesis 2: duplication before divergence of spiders and scorpions

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 4f to fig. 2a of [30]). In late stages of S. mesaensis,
expression of UbdA is similar to that of Cscu-Ubx-1 (i.e. with strong expression in O2 and posterior segments; compare figure 4e 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
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,

![Diagram](image-url)

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.)
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.

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).)

<table>
<thead>
<tr>
<th>Trait</th>
<th>lnL (unconstrained)</th>
<th>lnL (Hypothesis 1)</th>
<th>lnL (Hypothesis 2)</th>
<th>Δ(lnL_H1, H2)</th>
<th>σr (lnL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>labial</td>
<td>−991.94</td>
<td>−1090.34</td>
<td>−1094.68</td>
<td>4.332493</td>
<td>6.291582</td>
</tr>
<tr>
<td>proboscipedia</td>
<td>−1606.00</td>
<td>−1715.38</td>
<td>−1717.49</td>
<td>2.112443</td>
<td>2.312595</td>
</tr>
<tr>
<td>Deformed</td>
<td>−1327.66</td>
<td>−1447.39</td>
<td>−1463.84</td>
<td>16.450287</td>
<td>11.739177</td>
</tr>
<tr>
<td>Sex combs reduced</td>
<td>−1166.96</td>
<td>−1284.29</td>
<td>−1292.70</td>
<td>8.406873</td>
<td>4.859868</td>
</tr>
<tr>
<td>Antennapedia</td>
<td>−2937.54</td>
<td>−3157.80</td>
<td>−3158.49</td>
<td>0.694105</td>
<td>1.626799</td>
</tr>
<tr>
<td>Ultrabithorax</td>
<td>−1191.54</td>
<td>−1616.86</td>
<td>−1613.57</td>
<td>3.292049</td>
<td>4.592267</td>
</tr>
</tbody>
</table>

References


