EVOLUTION & DEVELOPMENT 17:6, 347–355 (2015)
DOI: 10.1111/ede.12166

Evidence of duplicated Hox genes in the most recent common ancestor of extant scorpions

Prashant P. Sharma,[a,b,*] Marc A. Santiago,[a] Edmundo González-Santillán,[c] Lionel Monod,[d] and Ward C. Wheeler[a]

[a] Division of Invertebrate Zoology, American Museum of Natural History, Central Park West at 79th Street, New York, NY 10024, USA
[b] Department of Zoology, University of Wisconsin-Madison, Madison, WI 53706, USA
[c] Laboratorio Nacional de Genómica para la Biodiversidad, Centro de Investigaciones y de Estudios Avanzados del Instituto Politecnico Nacional, and Laboratorio de Aracnología, Departamento de Biología Comparada, Facultad de Ciencias, Universidad Nacional Autónoma de México, Coyoacán, C.P. 04510, México D.F., México
[d] Département des Arthropodes et d'Entomologie I, Muséum d'histoire naturelle de la Ville de Genève, Route de Malagnou 1, CH-1208 Genève, Suisse
*Author for correspondence (e-mail: prashant.sharma@wisc.edu)

SUMMARY Scorpions (order Scorpiones) are unusual among arthropods, both for the extreme heteronomy of their bauplan and for the high gene family turnover exhibited in their genomes. These phenomena appear to be correlated, as two scorpion species have been shown to possess nearly twice the number of Hox genes present in most arthropods. Segmentally offset anterior expression boundaries of a subset of Hox paralogs have been shown to correspond to transitions in segmental identities in the scorpion posterior tagmata, suggesting that posterior heteronomy in scorpions may have been achieved by neofunctionalization of Hox paralogs. However, both the first scorpion genome sequenced and the developmental genetic data are based on exemplars of Buthidae, one of 19 families of scorpions. It is therefore not known whether Hox paralogy is limited to Buthidae or widespread among scorpions. We surveyed 24 high throughput transcriptomes and the single whole genome available for scorpions, in order to test the prediction that Hox gene duplications are common to the order. We used gene tree parsimony to infer whether the paralogy was consistent with a duplication event in the scorpion common ancestor. Here we show that duplicated Hox genes in non-buthid scorpions occur in six of the ten Hox classes. Gene tree topologies and parsimony-based reconciliation of the gene trees are consistent with a duplication event in the most recent common ancestor of scorpions. These results suggest that a Hox paralogy, and by extension the model of posterior patterning established in a buthid, can be extended to non-Buthidae scorpions.

INTRODUCTION

Depth of taxonomic sampling is a cardinal element of robustness in evolutionary inference. Especially prone to this consideration is evolutionary developmental biology. While advances in sequencing and functional tools have greatly ameliorated the paucity of available model systems within the last two decades, the number of established model systems available to evodevo research remains orders of magnitude smaller than the number of species used in such disciplines as molecular phylogenetics (Abzhanov et al. 2008).

For a handful of strongly conserved developmental processes, the scale of species sampling does not affect evolutionary inference. In such cases, a given development process is inferred to be phylotypic (i.e., synapomorphic) if it is shown that (a) the process occurs in multiple sampled species, and (b) the exemplars used to examine the process span the most recent common ancestor of that group. As an example within arthropods, the expression and functional dynamics of the segment polarity genes *engrailed*, *hedgehog*, and *wingless* appear to be conserved in the trunk segments of every sampled arthropod species, including both chelicerates and mandibulates (e.g., Nüsslein-Volhard and Wieschaus 1980; Fjose et al. 1985; Kornberg et al. 1985; Nusse and Varmus 1992; Klingensmith and Nusse 1994; Damen 2002, 2007; Hughes and Kaufman 2002a; Simonnet et al. 2004; O'Donnell and Jockusch 2010). A similar example is provided by the involvement of Notch signaling in arthropod appendage segmentation (de Celis et al. 1996; Bishop et al. 1999; Rauskolb and Irvine 1999; Prpic and Damen 2009).

In numerous other cases, previously held evolutionary scenarios are overturned when a gene's expression or function is shown to be homoplastic upon addition of data from new model systems. As examples, the gene *nubbin* was shown to

have different levels of involvement in patterning appendage joints across a variety of hemimetabolous and holometabolous insects (Li and Popadic 2004; Turchyn et al. 2011), as well as exhibiting species-specific expression patterns within spiders (Abzhanov and Kaufman 2000; Damen et al. 2002; Prpic and Damen 2005). Similarly, functional studies of the gene *Antennapedia* (*Antp*) in spiders have demonstrated functional convergence on the insect *Ultrabithorax* (*Ubx*) ortholog, suggesting that Hox gene functions are not always directly comparable between insects and arachnids (Khadjeh et al. 2012). Thus, generalizing a developmental process to a group of organisms requires supporting evidence from a range of model and satellite systems that corroborate evolutionary conservation.

These concerns weigh heavily upon a recent study addressing the patterning of the scorpion body plan (Sharma et al. 2014a). Arthropod tagmata and segmental identities are delimited by the action of the ten Hox genes common to all panarthropods (Hughes and Kaufman 2002b; Janssen et al. 2014). However, it was shown that 19 Hox genes (two paralogs of each Hox class except *Hox3*) are transcriptionally active during the embryogenesis of one scorpion exemplar *Centruroides sculpturatus* (Sharma et al. 2014a). Intriguingly, anterior expression domain boundaries of the last four Hox genes' duplicates correlate spatially with shifts in posterior segment identity (Fig. 1A). The paralogy and spatial correlation is striking, because scorpions have more segmental identities in their two posterior tagmata (mesosoma and metasoma) than any other extant chelicerate (Sharma et al. 2014a). These data were suggestive of a scenario wherein the segmental heteronomy of scorpions was achieved by Hox gene duplication and neofunctionalization of paralogs.

However, the extension of this model to the remaining ca. 2000 described scorpion species had been dubious for two reasons. First, paralogy of scorpion Hox gene duplicates was known only from two species, *Mesobuthus martensii* (sequenced genome; Cao et al. 2013; Di et al. 2015) and *Centruroides sculpturatus* (developmental transcriptome; Sharma et al. 2014a). Both are members of the scorpion family Buthidae, which is morphologically distinct from the remaining 18 scorpion families (Sissom 1990; Volschenk et al. 2008); it was therefore not known whether the duplications were unique to Buthidae. Second, at the time, a robust scorpion phylogeny was not available (Coddington et al. 2004). Without evidence of paralogy from non-buthid scorpions and the lack of a robust molecular phylogeny for the group, it was not possible to infer whether Hox gene paralogy is common to all scorpions.

More recently, the first molecular phylogeny of Scorpiones based on phylogenomic analyses of transcriptomic datasets was published (Fig. 1B) (Sharma et al. 2015). Given the present understanding of where Buthidae is placed in the scorpion tree, and clear identification of which lineages span the most recent common ancestor of scorpions, we are presently able to test the hypothesis that the duplication of Hox genes dates at least to the scorpion common ancestor. We combined the search for Hox

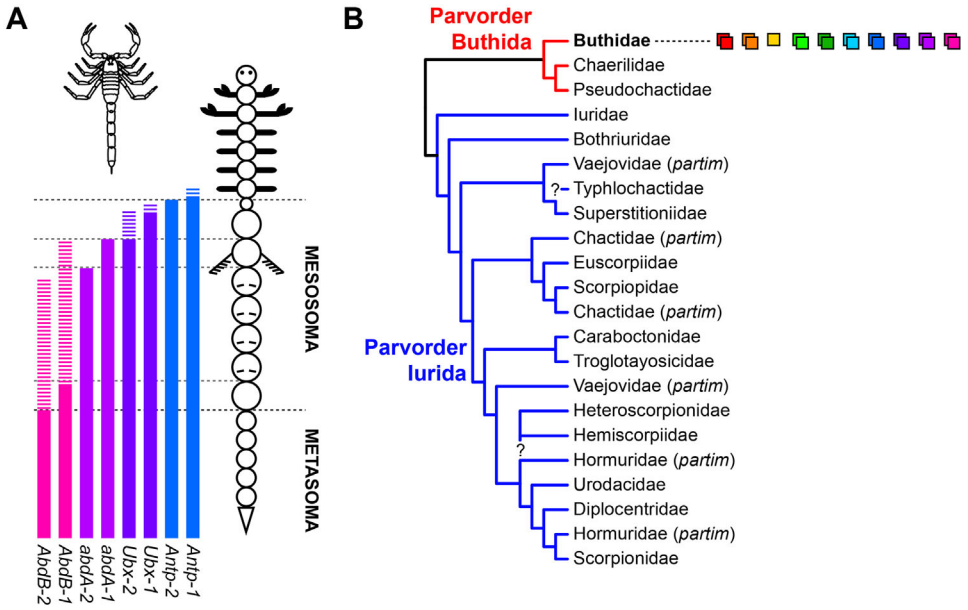

Fig. 1. A: Schematic representation of the scorpion bauplan and known expression domains of eight Hox paralogs in the posterior tagmata of the buthid *Centruroides sculpturatus* (Sharma et al. 2014a). B: Family level relationships of scorpions based on maximum likelihood analysis of 1557 genes (Sharma et al. 2015). Colored squares indicate the 19 Hox homologs that occur in *C. sculpturatus*, a member of Buthidae. Colors of branches indicate membership to parvorders Buthida (red) and Iurida (blue). Note that some scorpion families are not monophyletic and are pending systematic revision.

gene paralogs with topological tests, with the prediction that if the duplication occurred in the scorpion common ancestor, paralogs of Hox genes should form two mutually monophyletic clusters reflecting the putative duplication event.

MATERIALS AND METHODS

Identification of Hox gene orthologs

Fragments of homeodomain-containing genes were identified by tBLASTn from the assembled transcriptomes of 24 scorpions (Fig. 1B), previously used for phylogenomic investigation of scorpions (Sharma et al. 2015; accession numbers SRR1515193, SRR1721600-1767669). These were added to the available Hox sequences of the buthids *Centruroides sculpturatus* and *Mesobuthus martensii* previously analyzed (Sharma et al. 2014a). Nucleotide sequences of the ten Hox genes of *Phalangium opilio* were used as query sequences, as there is no evidence of Hox gene duplication in this basally-branching arachnid order (Sharma et al. 2012, 2014b). Inversely, sequences of spider Hox genes were not used, as spiders exhibit duplications of some Hox genes (Schwager et al. 2007).

Peptide translations of scorpion sequences were added to an existing 68-site alignment previously analyzed by us (Sharma et al. 2012, 2014a), which consists of Hox orthologs of various chelicerate and mandibulate species with verified embryonic expression patterns. Scorpion sequences that formed clusters exclusive of verified outgroup Hox genes in preliminary phylogenetic analyses were identified both topologically and using reciprocal best BLASTp hit; such non-Hox clusters were unambiguously identified as orthologs of other homeodomain-containing genes (e.g., *extradenticle*; *Distal-less*), and were excluded from the phylogenetic analyses described below.

Phylogenetic analyses

A 161-terminal, 68-site peptide alignment of all ten Hox genes was constructed after exclusion of non-Hox orthologs. Outgroup sequences consisted of Hox genes previously sequenced from a range of arthropod taxa for which gene identity was corroborated by expression studies (ref. Sharma et al. 2014a). Alignments were constructed de novo from nucleotide sequences using MUSCLE v. 3.6 (Edgar 2004). Ambiguously aligned sites were culled using GBlocks v.0.91b (Castresana 2000) using the commands $-b1 = 3$ (minimum of three sequences for a conserved position), $-b2 = 3$ (minimum of three sequences for a flanking position), $-b3 = 8$ (maximum of eight contiguous non-conserved positions), $-b4 = 5$ (minimum of five positions in a block), and $-b5 = a$ (gap positions allowed). Maximum likelihood analysis was conducted using RAxML v. 7.7.5 (Stamatakis 2006). A WAG model of sequence evolution with corrections for a discrete gamma distribution ($WAG + \Gamma$) was specified (Yang 1996; Whelan and Goldman 2001); 500 independent searches were conducted and nodal support estimated with 500 rapid bootstrap replicates (Stamatakis et al. 2008).

The restriction of the peptide alignment to the conserved homeodomain region was designed only to classify Hox genes using phylogenetic placement, as flanking sequences are often too diverged to align unambiguously across arthropods. The conserved Hox region used for phylogenetic analysis bears little sequence variation within individual Hox genes, which is reflected by near-zero terminal branch lengths across arthropods within each Hox gene cluster (e.g., Figure S1 of Sharma et al. 2012; Fig. 2 of Sharma et al. 2014a; Supplementary File 2). For this reason, we separately conducted phylogenetic analyses of nucleotide alignments for individual Hox gene trees, wherein novel duplicated Hox genes were identified (from the peptide alignment analysis described above). To facilitate sequence alignment of the more variable nucleotide sequences, analyses were restricted to scorpion Hox genes. A single copy Hox ortholog of the harvestman *Phalangium opilio* was used as an outgroup for each analysis, as Opiliones are basally branching with respect to Arachnopulmonata (scorpions + tetrapulmonates) and less prone to long branch attraction artifacts than such rapidly-evolving arachnid lineages as mites or ticks (Sharma et al. 2014b). Alignments were constructed de novo from nucleotide sequences using MUSCLE v. 3.6 (Edgar 2004), culled using GBlocks v. 0.91b (Castresana 2000) with the same commands as for the peptide alignment, and verified using peptide translations. A GTR model of sequence evolution with corrections for a discrete gamma distribution ($GTR + \Gamma$) was specified (Tavaré 1986; Yang 1996); 500 independent searches were conducted and nodal support estimated with 500 rapid bootstrap replicates. Topological placement and the previous designations of *C. sculpturatus* paralogs (Sharma et al. 2014a) were used to diagnose the orthology of newly discovered Hox genes (i.e., as "paralog 1" or "paralog 2"). All alignments are provided in Supplementary File 1.

Gene tree parsimony

To infer the timing of Hox gene duplication on the phylogenomic tree of scorpions previously inferred by us (Sharma et al. 2015), we conducted gene tree parsimony using DupTree v. 1.48 (Wehe et al. 2008), which seeks to minimize the number of gene duplications and does not score gene absence as loss events (appropriately suited to our case, as gene absence in our dataset is consistent with missing data, not necessarily gene loss). Maximum likelihood trees for all ten Hox genes were used as inputs. The leaf-adding heuristic with randomized hill-climbing was used to infer the most parsimonious species tree. Ten runs were independently conducted to ascertain heuristic stability.

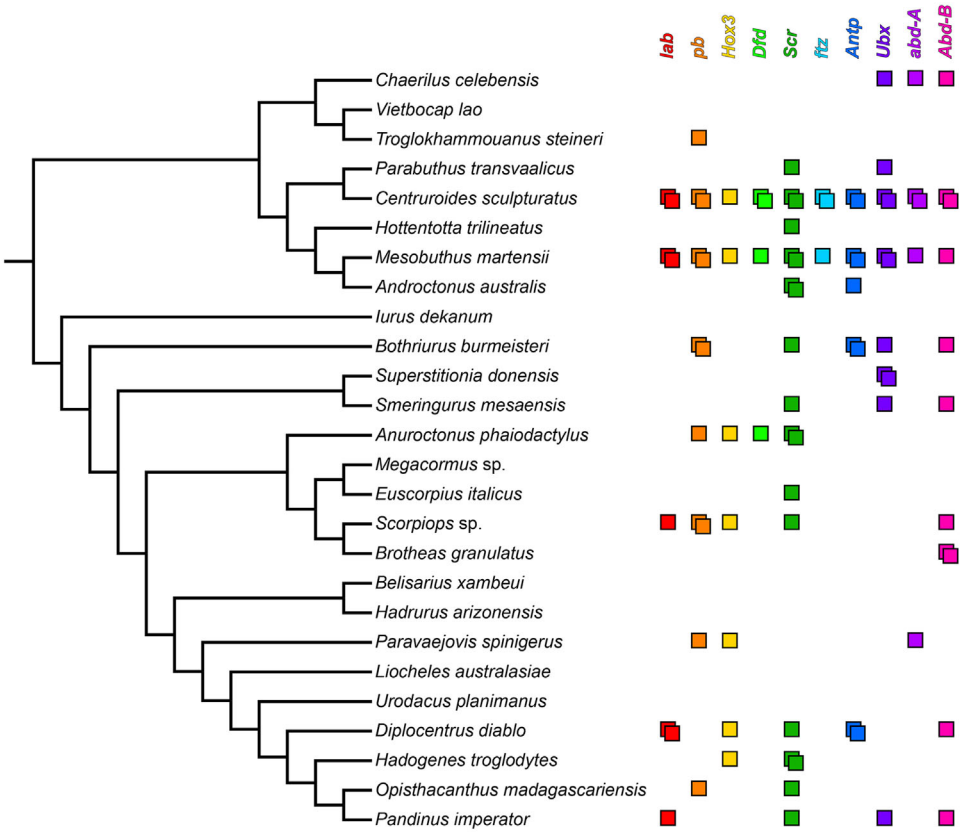

Fig. 2. Phylogenetic distribution of newly discovered Hox paralogs. Absences indicate missing data, not gene loss.

Because gene tree parsimony can be misled by the rare occurrence of a species in a set of partial gene trees, we conducted an additional run, where singly occurring terminals were culled, with the same heuristics as specified above.

RESULTS

Newly identified Hox genes

A total of 69 scorpion sequences were identified as bona fide Hox orthologs, as inferred from phylogenetic analysis of the 68-site alignment (Supplementary Files 1–2). In spite of incomplete Hox sequences in some taxa (minimum sequence length of 14 sites in the *abdominal-A* locus of the centipede *Lithobius atkinsoni*), the presence of key diagnostic sites in the homeodomain, together with the high level of sequence conservation within each Hox class, resulted in clear and diagnostic placements of all scorpion sequences among the Hox clusters (Supplementary File 2).

In every case of duplicated Hox genes, a maximum of two scorpion paralogs was discovered, consistent with the incidence of Hox genes in *C. sculpturatus* (Fig. 2). Newly identified orthologs were discovered for every Hox class except *fushi tarazu* (*ftz*). In nine of the 24 transcriptomes, one or more Hox genes were identified, but not in duplicate. Another eight transcriptomes bore duplicated Hox genes. In seven of the 25 scorpion transcriptomes, no Hox genes were identified at all.

Topological inference of Hox duplication

It was previously shown that two copies are present for all Hox genes except *Hox3* in the buthid *Centruroides sculpturatus* (19 Hox genes in total), whereas five duplicated Hox genes and another five in single copy were discovered in the low-coverage genome of *Mesobuthus martensii* (15 Hox genes in total) (Sharma et al. 2014a). As only the *C. sculpturatus* transcriptome was based wholly on embryonic tissues, we did not expect to recover all the Hox genes from every species in our dataset. Nevertheless, of the ten Hox classes, newly identified, duplicated Hox genes were discovered in six (*labial*, *proboscipedia*, *Sex combs reduced*, *Antennapedia*, *Ultrabithorax*, and *Abdominal-B*) (Figs. 3–5).

In all six cases, the duplicated Hox genes of a given species or derived species group never formed a sister pair in gene tree topologies, disfavoring the hypothesis of lineage-specific duplications. In three of the gene trees (*proboscipedia*, *Antennapedia*, and *Abdominal-B*), the recovered topology consisted of two mutually monophyletic clusters of paralogs

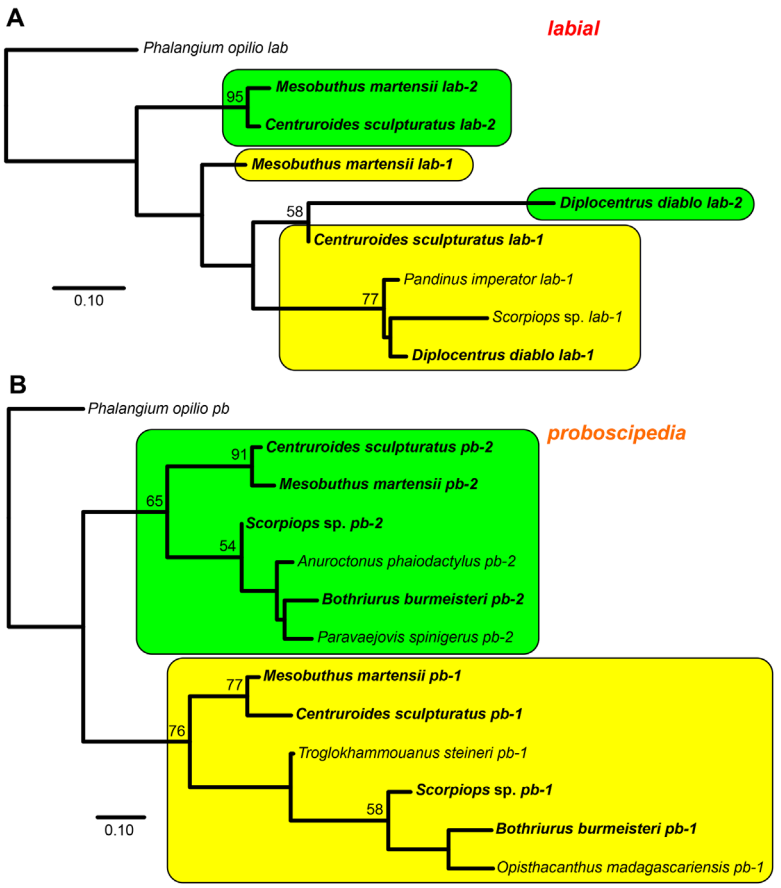

Fig. 3. Maximum likelihood tree topology of (A) *labial* orthologs ($\ln L = -884.89$), and (B) *proboscipedia* paralogs ($\ln L = -1470.36$), based on nucleotide sequence alignments. Shaded boxes indicated inferred orthology assignments, based on placement of *Centruroides sculpturatus* sequences. Boldfaced text indicates duplicates of a single species.

with nodal support (bootstrap frequencies ranging from 65% to 95%), with one member of each duplicated pair nesting within either cluster (Figs. 3B, 4B and 5B). In two cases (*Sex combs reduced* and *Ultrabithorax*), only one cluster of paralogs was recovered as monophyletic with support, with the remaining paralogs forming a grade without support (Figs. 4A and 5A). The gene tree of *labial* showed intermixed clusters of paralogs, and also without support for relationships between clusters (Fig. 3A).

Gene tree parsimony

The single most parsimonious solution reconciling the complete gene tree set was found twice (score: 22) and recovered a tree topology somewhat consistent with the phylogenomic analyses of Sharma et al. (2015), with four Hox gene duplications unambiguously and uniquely mapping to the common ancestor of scorpions (Fig. 6A). The sister relationship of Chaerilidae and Buthidae was recovered, as well as the monophyly of Iurida. However, Pseudochactidae was recovered as the sister group to all scorpions and relationships within Iurida did not accord with the understood relationships of scorpion families. As the pseudochactid exemplar and several other species occurred only in a single partial gene tree, we repeated the run after removing rare terminals, as specified above.

Reanalysis of a trimmed tree set lacking singly occurring terminals recovered five most parsimonious trees (score: 22), also with four Hox gene duplications unambiguously and uniquely mapping to the common ancestor of scorpions. A strict consensus of the most parsimonious solutions indicated the mutual monophyly of Buthida (represented by Chaerilidae and Buthidae) and Iurida, but no internal resolution within Iurida due to topological instability of multiple terminals (Fig. 6B).

DISCUSSION

Gene tree topologies reveal ancient paralogy in scorpion Hox genes

The purpose of this study was to test the hypothesis that the Hox gene paralogy exhibited by Buthidae, and by extension, the model of segmental heteronomy specified by novel expression domains of the Hox paralogs during embryogenesis, is common to all scorpions (Sharma et al. 2014a). This hypothesis is seemingly plausible, given the rigid evolutionary conservation of scorpion segmental architecture, but genetic evidence from

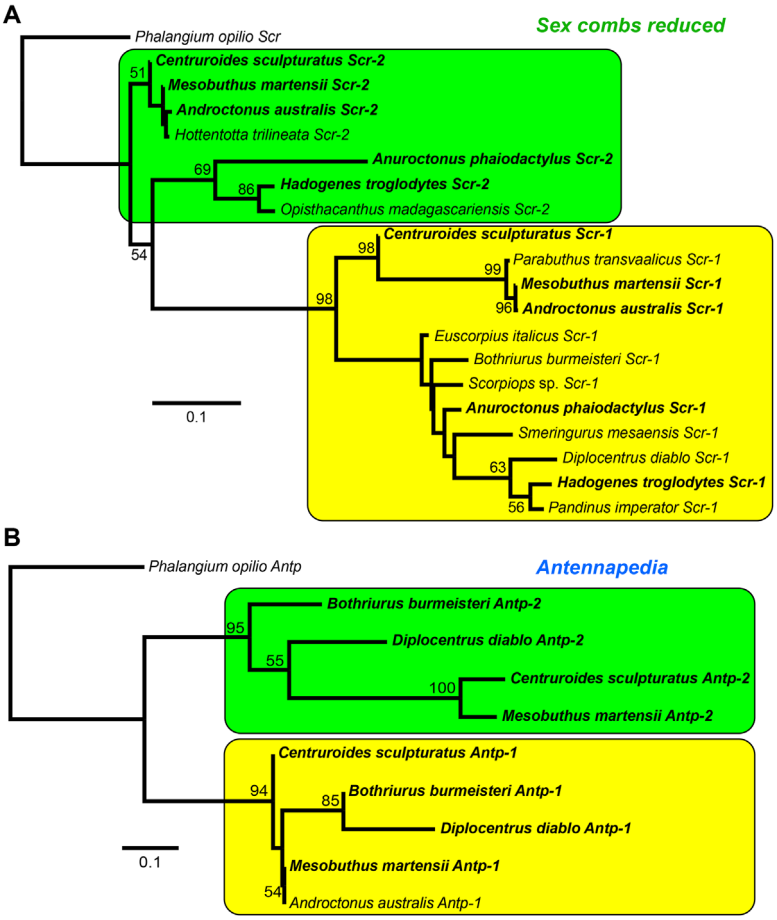

Fig. 4. Maximum likelihood tree topology of (A) *Sex combs reduced* paralogs ($\ln L = -1862.57$), and (B) *Antennapedia* paralogs ($\ln L = -2907.77$), based on nucleotide sequence alignments. Shaded boxes indicated inferred orthology assignments, based on placement of *Centruroides sculpturatus* sequences. Boldfaced text indicates duplicates of a single species.

the remaining scorpion families was lacking. Testing this hypothesis with molecular data was previously hindered by (a) the limitation of genomic resources to the family Buthidae and (b) the lack of a robust scorpion molecular phylogeny (Coddington et al. 2004). Having redressed the lack of a scorpion phylogeny with phylogenomic analyses (Sharma et al. 2015), we are presently able to test the two predictions stemming from a phylotypic developmental process. First, we expected duplicated Hox genes to occur in multiple scorpion species; and second, we expected evidence of a common duplication event to be present in both basal branches of the order Scorpiones (Buthida and Iurida).

In all six cases where duplicated Hox genes were discovered in one or more non-buthid scorpions, maximum likelihood analysis of the gene trees recovered evidence that the duplicates constituted out-paralogs. The tree topologies recovered are consistent with a scenario of ancient duplication, insofar as two mutually monophyletic clusters of paralogs (each consisting exclusively of one copy from a given species) were recovered with nodal support in three gene trees. In the other cases, the monophyly of one cluster was supported, whereas the non-monophyly of the second cluster was obtained without support. Only the gene tree of *labial* (Fig. 3A) is ambiguous, as the two *labial* paralogs of *Diplocentrus diablo* could be interpreted to belong to the same most-inclusive clade (albeit without nodal support). In no gene trees was the non-monophyly of a Hox cluster ever supported.

Because transcriptomes of subadult or adult tissues are not guaranteed to include all transcriptionally active Hox genes, we cannot interpret the absence of Hox genes as evolutionary losses. We therefore treat the absences as missing data, whose presence could be tested in future either by genomic sequencing of the targeted Hox genes, or by constructing transcriptomic libraries with embryos of that species (Sharma et al. 2014a).

We further add the caveat that, even when present, some scorpion sequences are incomplete in peptide and nucleotide alignments. Generally, given a uniform distribution of informative sites in an alignment, a shorter sequence results in a shorter terminal branch length, lowered power in phylogenetic resolution (i.e., lower nodal support), and higher likelihood of inaccurate phylogenetic placement. Therefore, in addition to improving the sampling of Hox genes, adding more complete sequences in alignments may improve the diagnostic power of tree topologies in detecting ancient paralogy. Nevertheless, in

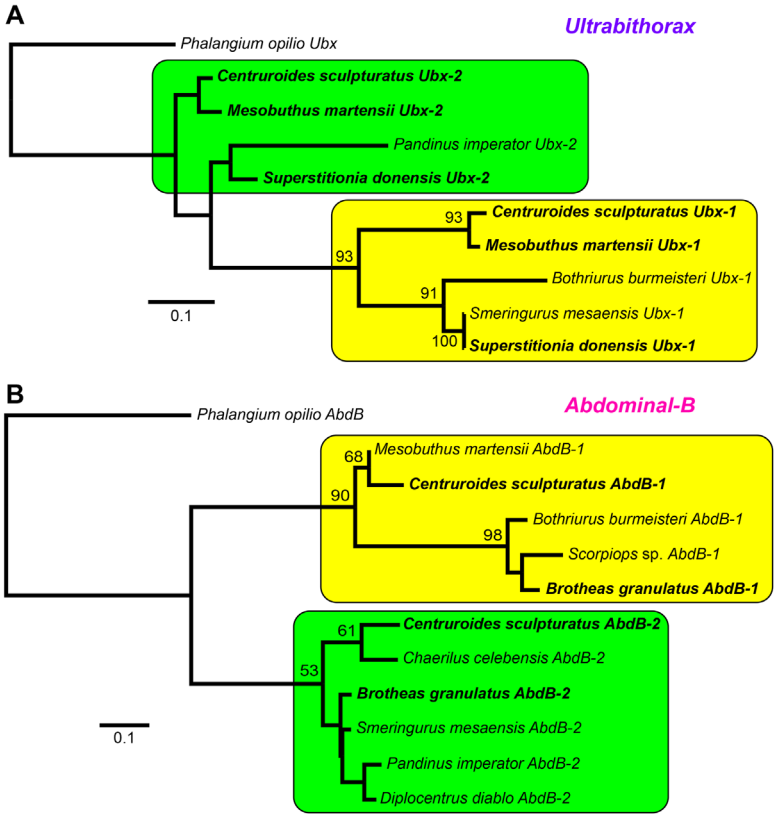

Fig. 5. Maximum likelihood tree topology of (A) *Ultrabithorax* paralogs ($\ln L = -1543.27$), and (B) *Abdominal-B* paralogs ($\ln L = -4315.21$), based on nucleotide sequence alignments. Shaded boxes indicated inferred orthology assignments, based on placement of *Centruroides sculpturatus* sequences. Boldfaced text indicates duplicates of a single species.

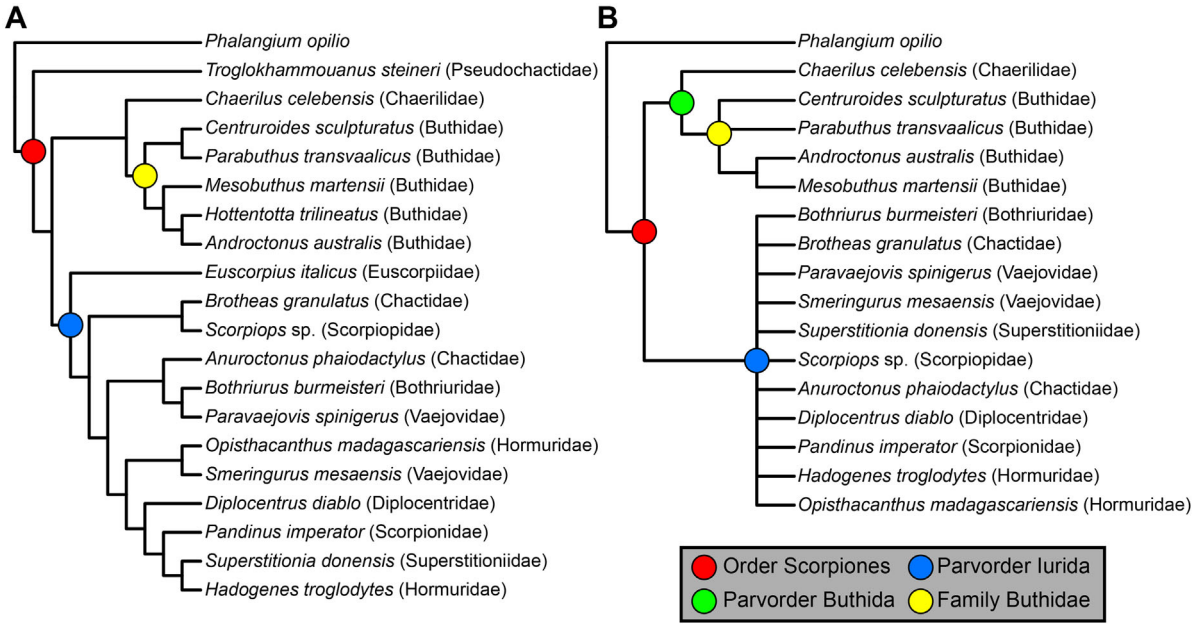

Fig. 6. A: Most parsimonious tree based on minimization of duplication events, using all maximum likelihood Hox gene tree topologies. B: Strict consensus of five most parsimonious trees based on minimization of duplication events, upon culling singly occurring terminals from input gene trees.

spite of the volume of missing data in our dataset (Fig. 2), the distribution of Hox paralogs, together with their gene tree topologies, corroborates our first expectation of systemic Hox gene paralogy in Scorpiones.

Duplication of Hox genes dates at least to the scorpion common ancestor

The presence of Hox paralogs in both basal clades of Scorpiones (Buthida and Iurida; Fig. 1B) is an expected outcome of duplication in the MRCA of scorpions, but is not sufficient to demonstrate a single basal duplication event. To infer the timing of the duplication, we conducted gene tree parsimony on all Hox gene trees we obtained. If the duplication event dates at least to the MRCA of scorpions, then the most parsimonious reconciliation minimizing duplication events was expected to reconstruct the basal tree topology of scorpions (*sensu* Sharma et al. 2015), with duplications mapping unambiguously to the scorpion MRCA.

Consistent with this prediction, the most parsimonious reconciliation of all gene trees recovered a tree topology largely comparable to the phylogenomic tree of scorpions (compare Figs. 1B to 6A). Due to the misleading effects of singly occurring taxa in partial gene tree inputs, anomalous placements were recovered for such taxa as *Troglokhammouanus steineri* and *Euscorpius italicus*. Upon culling poorly represented taxa from the gene tree set, the tree topology obtained recovered the basal split of Buthida and Iurida, and the monophyly of Buthidae (Fig. 6B). In either case, at least four Hox duplications mapped unambiguously and uniquely to the MRCA of scorpions. The lack of resolution within Iurida may reflect an abundance of missing data (resulting in topological instability of some terminals) or the decay of phylogenetic signal within a given Hox gene alignment, or some combination of the two. Nevertheless, these results from parsimony analysis of the constituent gene trees is strongly supportive of a duplication event predating the basal divergence of the extant scorpions, which has been estimated ca. 200 ± 50 Mya (Sharma and Wheeler 2014).

CONCLUSION

The discovery of Hox gene paralogy throughout the arachnid order Scorpiones, together with the tree topologies of constituent Hox gene trees, supports a duplication event concomitant with or prior to the basal divergence of scorpions. These observations suggest that a previously proposed model of posterior heteronomy achieved by duplication and putative functional divergence of Hox paralogs (Sharma et al. 2014a) is generalizable to all scorpions. Given that Hox duplicates have also been reported in a spider (Schwager et al. 2007; Sharma et al. 2014a), future studies should explore whether Hox cluster and/or whole genome duplication is synapomorphic of Arachnopulmonata (scorpions + tetrapulmonates) (Sharma et al. 2014b).

Acknowledgments

This material is based on work supported by the National Science Foundation Postdoctoral Research Fellowship in Biology under Grant No. DBI-1202751 to PPS. Participation of MAS in this research was made possible by the AMNH Volunteers Program. Comments from Lisa Nagy and two anonymous reviewers improved an earlier draft of this work.

Note Added in Proof

During the late stages of this manuscript's publication, an article was published in the journal Insect Biochemistry and Molecular Biology, titled, "Genome-wide analysis of homeobox genes from Mesobuthus martensii reveals Hox gene duplication in scorpions" (citation in text). This article reported additional paralogy of homeobox genes in the new version of the Mesobuthus martensii genome, and indicated the possibility of two complete Hox clusters in this buthid exemplar. These data are consistent with the scenario of a common duplication event dating at least to the scorpion MRCA, as supported by our gene tree analyses in the present study. We encourage interested readers to examine Di et al. (2015).

REFERENCES

Abzhanov, A., Extavour, C. G., Groover, A., Hodges, S. A., Hoekstra, H. E., Kramer, E. M., and Monteiro, A. 2008. Are we there yet? Tracking the development of new model systems. *Trends Genet.* 24: 353–360.

Abzhanov, A., and Kaufman, T. C. 2000. Homologs of *Drosophila* appendage genes in the patterning of arthropod limbs. *Dev. Biol.* 227: 673–689.

Bishop, S. A., Klein, T., Martinez Arias, A., and Couso, J. P. 1999. Composite signaling from *Serrate* and *Delta* establishes leg segments in *Drosophila* through *Notch*. *Development* 126: 2993–3003.

Cao, Z., et al. 2013. The genome of *Mesobuthus martensii* reveals a unique adaptation model of arthropods. *Nature Communications* 4: 2602.

Castresana, J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* 17: 540–552.

Coddington, J. A., Giribet, G., Harvey, M. S., Prendini, L., and Walter, D. E. 2004. Arachnida. In J. Cracraft, M. J. Donoghue (eds.). *Assembling the tree of life*. Oxford University Press: New York (NY), pp. 296–318.

Damen, W. G. M. 2002. Parasegmental organization of the spider embryo implies that the parasegment is an evolutionary conserved entity in arthropod embryogenesis. *Development* 129: 1239–1250.

Damen, W. G. M. 2007. Evolutionary conservation and divergence of the segmentation process in arthropods. *Dev. Dynam.* 236: 1379–1391.

Damen, W. G. M., Saridaki, T., and Averof, M. 2002. Diverse adaptations of an ancestral gill: a common evolutionary origin for wings, breathing organs, and spinnerets. *Curr. Biol.* 12: 1711–1716.

de Celis, J. F., de Celis, J., Ligoxygakis, P., Preiss, A., Delidakis, C., and Bray, S. J. 1996. Functional relationships between *N*, *Su(H)*, and the bHLH genes of the *E(spl)* complex: the *E(spl)* genes mediate only a subset of *Notch* activities during imaginal development. *Development* 122: 2719–2728.

Di, Z., et al. 2015. Genome-wide analysis of homeobox genes from Mesobuthus martensii reveals Hox gene duplication in scorpions. *Insect Biochem. Molec.* 61: 25–33.

Edgar, R. C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nuc. Acids. Res.* 32: 1792–1797.
Fjose, A., McGinnis, W. J., and Gehring, W. J. 1985. Isolation of a homeobox containing gene from the *engrailed* region of *Drosophila* and the spatial distribution of its transcripts. *Nature* 313: 284–289.
Hughes, C. L., and Kaufman, T. C. 2002a. Exploring myriapod segmentation: the expression patterns of *even-skipped*, *engrailed* and *wingless* in a centipede. *Dev. Biol.* 247: 47–61
Hughes, C. L., and Kaufman, T. C. 2002b. Hox genes and the evolution of the arthropod body plan. *Evol. Dev.* 4: 459–499.
Janssen, R., Eriksson, B. J., Tait, N. N., and Budd, G. E. 2014. Onychophoran Hox genes and the evolution of arthropod Hox gene expression. *Front. Zool.* 11: 22.
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.
Klingensmith, J., and Nusse, R. 1994. Signaling by *wingless* in *Drosophila*. *Dev. Biol.* 166: 396–414.
Kornberg, T., Siden, I., O'Farrell, P. H., and Simon, M. 1985. The *engrailed* locus of Drosophila: in situ localization of transcripts reveals compartment-specific expression. *Cell* 40: 45–63.
Li, H., and Popadic, A. 2004. Analysis of *nubbin* expression patterns in insects. *Evol. Dev.* 6: 310–324.
Nusse, R., and Varmus, H. E. 1992. *Wnt* genes. *Cell* 69: 1073–1087
Nüsslein-Volhard, C., and Wieschaus, E. 1980. Mutations affecting segment number and polarity in *Drosophila*. *Nature* 287: 795–801.
O'Donnell, B. C., and Jockusch, E. L. 2010. The expression of *wingless* and Engrailed in developing embryos of the mayfly *Ephoron leukon* (Ephemeroptera: Polymitarcyidae). *Dev. Genes Evol.* 220: 11–24.
Prpic, N. M., and Damen, W. G. M. 2005. Diversification of *nubbin* expression patterns in arthropods: data from another spider species *Cupiennius salei*. *Evol. Dev.* 7: 276–279.
Prpic, N.-M., and Damen, W. G. M. 2009. Notch-mediated segmentation of the appendages is a molecular phylotypic trait of the arthropods. *Dev. Biol.* 326: 262–271.
Rauskolb, C., and Irvine, K. D. 1999. *Notch*-mediated segmentation and growth control of the Drosophila leg. *Dev. Biol.* 210: 339–350.
Schwager, E. E., Schoppmeier, M., Pechmann, M., and Damen, W. G. M. 2007. Duplicated Hox genes in the spider *Cupiennius salei*. *Front. Zool.* 4: 10.
Sharma, P. P., Schwager, E. E., Extavour, C. G., and Giribet G. 2012. Hox gene expression in the harvestman *Phalangium opilio* reveals divergent patterning of the chelicerate opisthosoma. *Evol. Dev.* 14: 450–463.
Sharma, P. P., Schwager, E. E., Extavour, C. G., and Wheeler, W. C. 2014a. Hox gene duplications correlate with posterior heteronomy in scorpions. *Proc. R. Soc. B* 281: 20140661.
Sharma, P. P., et al. 2014b. Phylogenomic interrogation of Arachnida reveals systemic conflicts in phylogenetic signal. *Mol. Biol. Evol.* 31: 2963–2984.
Sharma, P. P., Fernández, R., Esposito, L. A., González-Santillán, E., and Monod, L. 2015. Phylogenomic resolution of scorpions reveals multilevel discordance with morphological phylogenetic signal. *Proc. R. Soc. B* 282: 20142953.
Sharma, P. P., and Wheeler, W. C. 2014. Cross-bracing uncalibrated nodes in molecular dating improves congruence of fossil and molecular age estimates. *Front. Zool*. 11: 57.
Simonnet, F., Deutsch, J., and Quéinnec, E. 2004. hedgehog is a segment polarity gene in a crustacean and a chelicerate. *Dev. Genes Evol.* 214: 537–545.
Sissom, W. D. 1990. Systematics, biogeography, and paleontology. In G. A. Polis (ed.). *The biology of scorpions*. Stanford University Press: Stanford (CA), pp. 64–160.
Stamatakis, A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22: 2688–2690.
Stamatakis, A., Hoover, P., and Rougemont, J. 2008. A rapid bootstrap algorithm for the RAxML Web servers. *Syst. Biol.* 57: 758–771.
Tavaré, S. 1986. Some probabilistic and statistical problems in the analysis of DNA sequences. *Lect. Math. Life Sci.* 17: 57–86.
Turchyn, N., Chesebro, J., Hrycaj, S., Couso, J. P., and Popadic, A. 2011. Evolution of *nubbin* function in hemimetabolous and holometabolous insect appendages. *Dev. Biol.* 357: 83–95.
Volschenk, E. S., Mattoni, C. I., and Prendini, L. 2008. Comparative anatomy of the mesosomal organs of scorpions (Chelicerata, Scorpiones), with implications for the phylogeny of the order. *Zool. J. Linn. Soc.* 154: 651–675
Wehe, A., Bansal, M. S., Burleigh, J. G., and Eulenstein, O. 2008. DupTree: a program for large-scale phylogenetic analyses using gene tree parsimony. *Bioinformatics* 24: 1540–1541.
Whelan, S., and Goldman, N. 2001. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol. Biol. Evol.* 18: 691–699.
Yang, Z. 1996. Among-site rate variation and its impact on phylogenetic analyses. *Trends. Ecol. Evol.* 11: 367–372.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article at the publisher's web-site.

File 1. Peptide alignments of all Hox genes and nucleotide alignments used to infer gene trees.

File 2. Treefile of maximum likelihood topology of the 68-amino acid alignment of conserved Hox region, using selected arthropod outgroup taxa with known Hox gene expression data ($\ln L = -2193.69$).