





# Evolution of the Spider Homeobox Gene Repertoire by Tandem and Whole Genome Duplication

Madeleine E. Aase-Remedios <sup>1</sup>, Ralf Janssen <sup>2,\*</sup>, Daniel J. Leite <sup>1</sup>, Lauren Sumner-Rooney,<sup>3</sup> and Alistair P. McGregor <sup>1,\*</sup>

<sup>1</sup>Department of Biosciences, Durham University, Durham, DH1 3LE, United Kingdom

<sup>2</sup>Department of Earth Sciences, Uppsala University, Uppsala, 752 36, Sweden

<sup>3</sup>Museum für Naturkunde, Leibniz-Institut für Evolutions- und Biodiversitätsforschung, Berlin, 10115, Germany

\*Corresponding authors: E-mails: alistair.mcgregor@durham.ac.uk; ralf.janssen@geo.uu.se.

Associate editor: Patricia Wittkopp

## Abstract

Gene duplication generates new genetic material that can contribute to the evolution of gene regulatory networks and phenotypes. Duplicated genes can undergo subfunctionalization to partition ancestral functions and/or neofunctionalization to assume a new function. We previously found there had been a whole genome duplication (WGD) in an ancestor of arachnospulmonates, the lineage including spiders and scorpions but excluding other arachnids like mites, ticks, and harvestmen. This WGD was evidenced by many duplicated homeobox genes, including two Hox clusters, in spiders. However, it was unclear which homeobox paralogues originated by WGD versus smaller-scale events such as tandem duplications. Understanding this is a key to determining the contribution of the WGD to arachnospulmonate genome evolution. Here we characterized the distribution of duplicated homeobox genes across eight chromosome-level spider genomes. We found that most duplicated homeobox genes in spiders are consistent with an origin by WGD. We also found two copies of conserved homeobox gene clusters, including the Hox, NK, HRO, *lrx*, and SINE clusters, in all eight species. Consistently, we observed one copy of each cluster was degenerated in terms of gene content and organization while the other remained more intact. Focussing on the NK cluster, we found evidence for regulatory subfunctionalization between the duplicated NK genes in the spider *Parasteatoda tepidariorum* compared to their single-copy orthologues in the harvestman *Phalangium opilio*. Our study provides new insights into the relative contributions of multiple modes of duplication to the homeobox gene repertoire during the evolution of spiders and the function of NK genes.

**Key words:** evolution, development, homeobox, spider, whole genome duplication.

## Introduction

Whole genome duplications (WGDs) have occurred several times during animal evolution (Ohno 1970; Putnam et al. 2008; Flot et al. 2013; Kenny et al. 2016; Schwager et al. 2017; Nong et al. 2021). Duplicated genes are thought to be released from selective constraints, allowing paralogues to subfunctionalize, specialize, or neofunctionalize either by changes to their cis-regulatory controls or protein sequence while collectively maintaining the robustness of the ancestral functionality (Force et al. 1999; Jiménez-Delgado et al. 2009; Tinti et al. 2012; Espinosa-Cantú et al. 2015; Marlétaz et al. 2018). The two rounds of whole genome duplication (WGD) (2R WGD) in the gnathostome ancestor have been credited with facilitating the evolution of vertebrate novelties, however, the extent to which WGDs in other animal lineages have contributed to diversification is not well understood (Meyer and Schartl 1999; Shimeld and Holland 2000; Cañestro et al. 2013; Simakov et al. 2020; Aase-Remedios and Ferrier 2021).

Many retained ohnologues (genes duplicated via WGD) are transcription factors that regulate processes in

development, which further implicates the role of duplication in the evolution of phenotypes and the underlying developmental regulatory networks (Brunet et al. 2017). Homeobox genes are a superclass of transcription factors that expanded by tandem duplication early in animal evolution into eleven recognized classes encompassing more than 100 bilaterian gene families with diverse and wide-reaching roles in development (Table 1) (Holland et al. 2007; Larroux et al. 2008; Zhong et al. 2008; Kumar 2009; Butts et al. 2010; Bürglin 2011; Zhong and Holland 2011; Holland 2013; Holland et al. 2017). Certain homeobox gene expansions may have coincided with innovations that occurred at the evolutionary origins of animals and subsequently bilaterians (García-Fernández 2005; Holland 2013; Holland 2015). Tandem duplication of homeobox genes early in animal evolution resulted in clusters of related genes that have been conserved across animals (Pollard and Holland 2000; Larroux et al. 2007, 2008; Ferrier 2016a). Homeobox gene clusters (Table 1) are thought to be maintained by functional constraint on their genomic organization exerted by complex and

Received: May 26, 2023. Revised: October 02, 2023. Accepted: October 26, 2023

© The Author(s) 2023. Published by Oxford University Press on behalf of Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Open Access

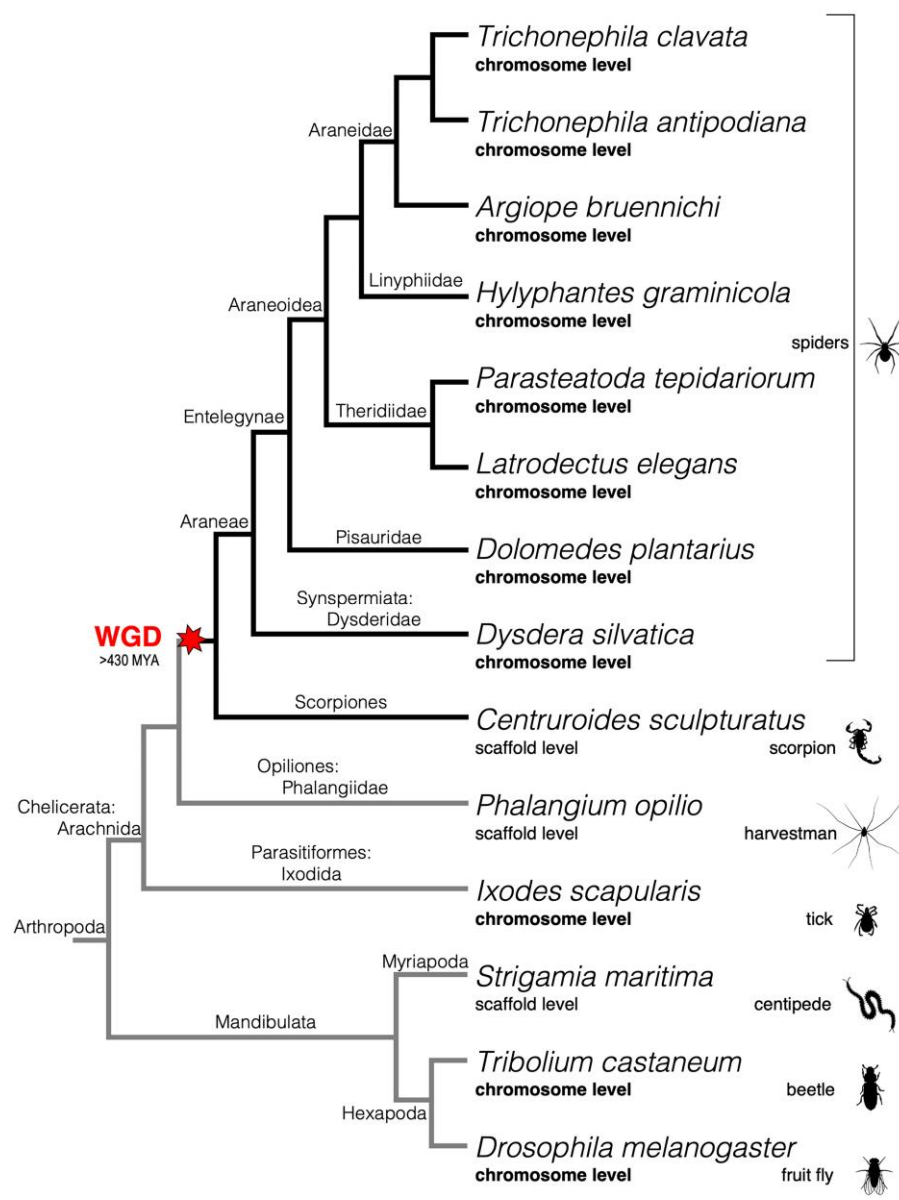
overlapping regulatory elements controlling the expression of adjacent genes (Irimia et al. 2012; Ferrier 2016b).

The best-known homeobox gene cluster, the Hox cluster, illustrates how the genomic organization of homeobox genes is important for their function. Most arthropod genomes contain one Hox cluster consisting of ten ANTP-class genes (Table 1), which are involved in patterning the antero-posterior axis of animals in the same relative order as the genomic organization of the genes within the cluster (Krumlauf 2018 and the references therein). The Hox cluster is highly conserved across bilaterians, both in genomic organization and function, though differences in Hox gene expression are known to underlie changes to animal body plans (Averof and Patel 1997; Abzhanov et al. 1999; Janssen et al. 2014;

Martin et al. 2016; Serano et al. 2016; Janssen and Pechmann 2023). Hox clusters have also been retained in duplicate following WGDs, as exhibited by the four vertebrate Hox clusters that arose via the 2R WGD (Holland and Garcia-Fernández 1996; McLysaght et al. 2002; Holland et al. 2007; Putnam et al. 2008; Cañestro et al. 2013; Holland 2013; Holland and Ocampo Daza 2018). The 2R WGD appears to have a relaxed constraint on vertebrate Hox cluster organization so that while the complement of genes from *Hox1* to *Hox14* is present in at least one copy, no cluster contains all fourteen genes (Hoegg and Meyer 2005). It is still not well understood to what extent the patterns of gene loss between duplicate clusters and the subfunctionalization observed for vertebrate homeobox gene clusters apply more widely to

**Table 1** Genes and *Drosophila melanogaster* synonyms in conserved bilaterian homeobox clusters. Bilaterian homeobox gene **families** contain genes descendant from a single inferred gene in the ancestor of all bilaterians, as previously defined by Ferrier (2016b) and the references therein. Families are grouped into **classes**, which themselves originated by duplication earlier in animal evolution. **Clusters** are composed of related genes that arose by tandem duplication and have been conserved across bilaterians. The Hox cluster contains the Hox genes, while the ancestral bilaterian Super-Hox cluster includes the related ANTP-class genes that are often found linked to the Hox cluster (Butts et al. 2010; Ferrier 2016b). The ParaHox cluster is the sister of the Hox cluster and consists of three bilaterian families; dashes for *Xlox/Pdx* indicate it is not found in arthropods (Brooke et al. 1998; Ferrier 2016a). The NK cluster also belongs to the ANTP class. Some ANTP-class NK2 cluster genes were first described as part of a “pharyngeal” cluster, containing NK2.1, NK2.2, and *Msx1x*, as well as two nonhomeobox genes, *Pax1/9* and *FoxA1/2* (Simakov and Kawashima 2017); linkage and sometimes clustering between NK2.1, NK2.2, and *Msx1x* is evident across bilaterians, and here is referred to as the NK2 cluster (Simakov et al. 2015; Ferrier 2016b). The PRD-class HRO cluster (Mazza et al. 2010) also includes the LIM-class gene *Isl* as they are often found linked (Ferrier (2016b)). The TALE-class *Lrx* cluster likely underwent independent duplications in different lineages; *ara* and *caup* are a fly-specific duplication of one *Lrx* family, often clustered with single arthropod *mirr* orthologues of a second *Lrx* family (Irimia et al. 2008; Kerner et al. 2009). The dash for *Lrx3* indicates it is not found in *D. melanogaster*. The SINE-class cluster (Gallardo et al. (1999) consists of three families and is conserved across bilaterians (Ferrier 2016b).

Cluster (Class)	Bilaterian Family	Arthropod Family	Synonyms	Cluster (Class)	Bilaterian Family	Arthropod Family	Synonyms
Super-Hox (ANTP)	<i>Hox1</i>	<i>lab</i>	<i>labial (lab)</i>	NK (ANTP)	<i>NK1</i>	<i>NK1</i>	<i>slouch (slou)</i>
	<i>Hox2</i>	<i>pb</i>	<i>proboscipedia (pb)</i>		<i>NK3</i>	<i>NK3</i>	<i>bagpipe (bap)</i>
	<i>Hox3</i>	<i>Hox3</i>	<i>zerknüllt (zen), zen2, bicoid (bcd)</i>		<i>NK4</i>	<i>NK4</i>	<i>tinman (tin)</i>
	<i>Hox4</i>	<i>Dfd</i>	<i>Deformed (Dfd)</i>		<i>NK5</i>	<i>NK5</i>	<i>H6-like homeobox (Hmx)</i>
	<i>Hox5</i>	<i>Scr</i>	<i>Sex combs reduced (Scr)</i>		<i>NK6</i>	<i>NK6</i>	<i>HGTX</i>
	<i>Hox6-8</i>	<i>ftz</i>	<i>fushi tarazu (ftz)</i>		<i>NK7</i>	<i>NK7</i>	<i>NK7.1</i>
		<i>Antp</i>	<i>Antennapedia (Antp)</i>		<i>Msx</i>	<i>Msx</i>	<i>Drop (Dr), msh</i>
		<i>Ubx</i>	<i>Ultrabithorax (Ubx)</i>		<i>Tlx</i>	<i>Tlx</i>	<i>C15</i>
		<i>abd-A</i>	<i>abdominal-A (abd-A)</i>		<i>Lbx</i>	<i>Lbx</i>	<i>ladybird early (lbe), ladybird late (lbi)</i>
	<i>Hox9-15</i>	<i>Abd-B</i>	<i>Abdominal-B (Abd-B)</i>		<i>Hhex</i>	<i>Hhex</i>	<i>HHEX</i>
	<i>Abox</i>	<i>Abox</i>	<i>CG34031</i>		<i>Noto</i>	<i>Noto</i>	<i>CG18599</i>
	<i>Dbx</i>	<i>Dbx</i>	<i>Dbx</i>		<i>Emx</i>	<i>Emx</i>	<i>ems, E5</i>
	<i>Dlx</i>	<i>Dlx</i>	<i>Distalless (Dll)</i>	NK2 (ANTP)	<i>NK2.1</i>	<i>NK2.1</i>	<i>scarecrow (scro)</i>
	<i>En</i>	<i>En</i>	<i>engrailed (en), invected (inv)</i>		<i>NK2.2</i>	<i>NK2.2</i>	<i>ventral nervous system defective (vnd)</i>
	<i>Evx</i>	<i>Evx</i>	<i>even skipped (eve)</i>		<i>Msx1x</i>	<i>Msx1x</i>	<i>CG15696</i>
	<i>Gbx</i>	<i>Gbx</i>	<i>unplugged (unpg)</i>		<i>Hlx</i>	<i>Hlx</i>	<i>Homeodomain protein 2.0 (H2.0)</i>
	<i>Meox</i>	<i>Meox</i>	<i>buttonless (btn)</i>	<i>Lrx</i> (TALE)	<i>Lrx</i>	<i>Lrx1</i>	<i>araucan (ara), caupolican (caup)</i>
	<i>Mnx</i>	<i>Mnx</i>	<i>extra-extra (exex)</i>			<i>Lrx2</i>	<i>mirror (mirr)</i>
	<i>Nedx</i>	<i>Nedx</i>	<i>lateral muscles scarcer (lms)</i>			<i>Lrx3</i>	–
ParaHox (ANTP)	<i>Ro</i>	<i>Ro</i>	<i>rough (ro)</i>	HRO + <i>Isl</i> (PRD + LIM)	<i>Hbn</i>	<i>Hbn</i>	<i>homeobrain (hbn)</i>
	<i>Xlox/Pdx</i>	–	–		<i>Rax</i>	<i>Rax</i>	<i>Retinal Homeobox (Rx)</i>
	<i>Gsx</i>	<i>Gsx</i>	<i>intermediate neuroblasts defective (ind)</i>		<i>Otp</i>	<i>Otp</i>	<i>orthopedia (otp)</i>
	<i>Cdx</i>	<i>Cdx</i>	<i>caudal (cad)</i>	SINE (SINE)	<i>Isl</i>	<i>Isl</i>	<i>tailup (tup)</i>
					<i>Six1/2</i>	<i>Six1/2</i>	<i>sine oculis (so)</i>
					<i>Six3/6</i>	<i>Six3/6</i>	<i>optix</i>
					<i>Six4/5</i>	<i>Six4/5</i>	<i>Six4</i>



**Fig. 1.** Cladogram of arthropods used in this study. The red star indicates the timing of the WGD; preduplicate lineages have gray lines, and duplicate lineages have black lines. Genome assembly level is indicated below the species binomial, in bold for the chromosome-level assemblies. Higher phylogenetic classifications are labeled on phylogeny nodes. The topology was drawn based on published molecular phylogenies (Ontano et al. 2021; Li et al. 2022). Silhouettes were taken from the PhyloPic database [https://www.phylopic.org/].

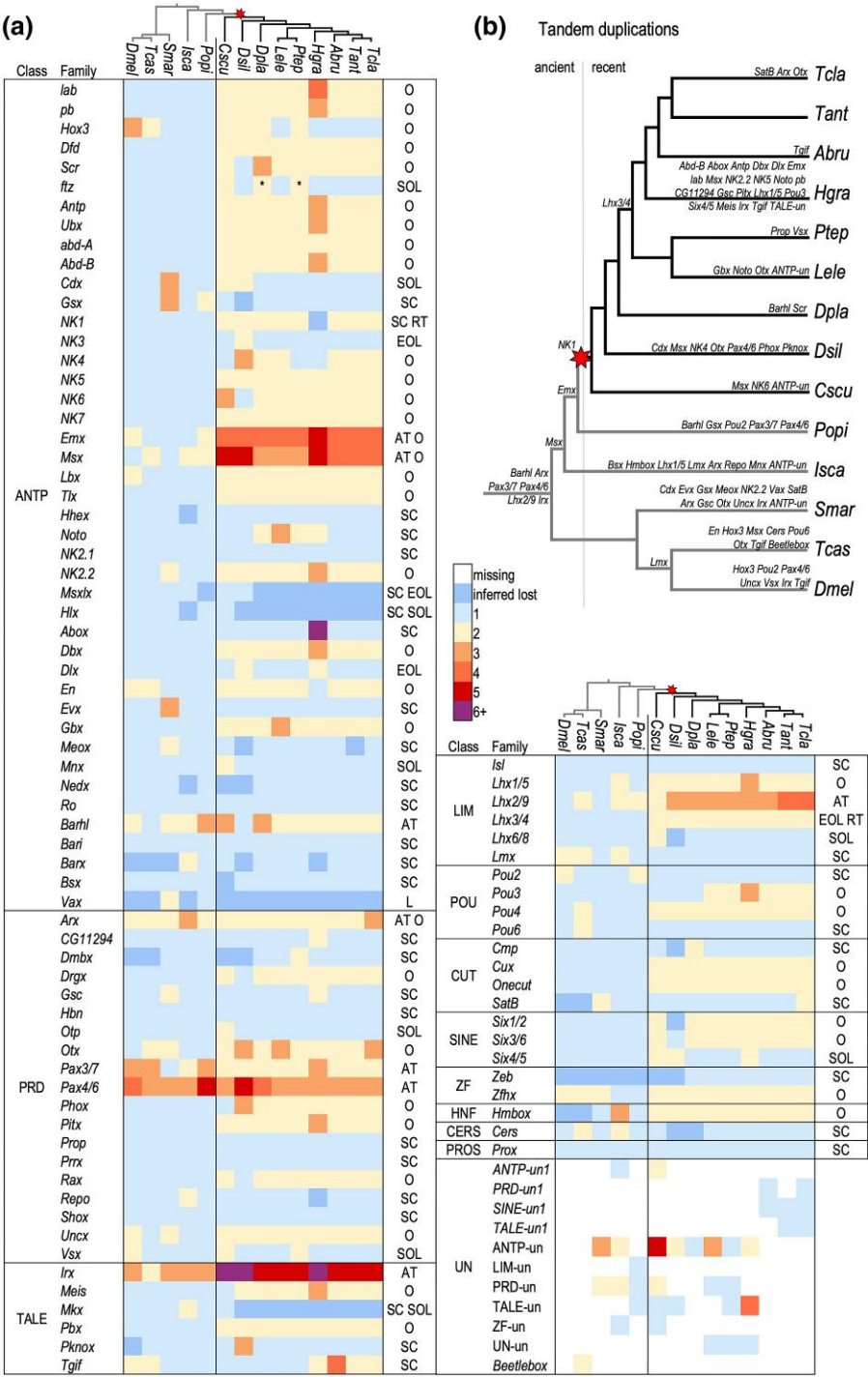
other homeobox gene clusters (Table 1) and to other animal lineages that have undergone WGD.

Besides gnathostomes, an independent WGD took place in the arachnospulmonate ancestor approximately 450 MYA (Fig. 1), resulting in nearly 60% of homeobox genes retained in duplicate in spiders and scorpions and their relatives, including two Hox clusters (Schwager et al. 2017; Leite et al. 2018; Harper et al. 2021). Within each Hox gene family, one ohnologue exhibits different spatial and temporal expression compared to the other, though overall, spatial and temporal collinearity is largely intact for each cluster (Schwager et al. 2007, 2017; Turetzek et al. 2022). Several of the other retained homeobox genes exhibit divergent expression consistent with both sub- and neofunctionalization in *Parasteatoda tepidariorum*, showing that duplication of these genes has played a role in the evolution of spider development (Leite et al. 2018).

The lack of chromosomal-level genome assemblies for spiders has so far hindered a full understanding of the

relationships among duplicated homeobox genes and the relative impact of the WGD compared to tandem duplication. It was not possible to determine whether clusters were fully intact or had undergone degradation, beyond one of the Hox clusters (Schwager et al. 2017). Leite et al. (2018) described the clustering of a few genes, each belonging to several of the other homeobox gene clusters besides the Hox, including fragments of the NK, HRO, *Irx*, and SINE clusters, some of which were found in duplicate. However, it remained impossible to determine whether these duplicates were ohnologues, indicating the retention of ohnologue clusters, or if they resulted from tandem duplications occurring before or after the WGD. Furthermore, it could not be determined if these duplicates were shared with other spider species because of the lack of comparative genomic resources.

We surveyed the homeobox gene complement and synteny across eight recently released chromosome-level genome assemblies available for spiders including *P. tepidariorum*



**Fig. 2.** Many homeobox genes were retained in duplicate in scorpion and spider genomes. a) Heatmap of homeobox genes by family. The relationships among duplicates (except for species-specific recent tandem duplications or family/species-specific losses) are indicated to the right. O: ohnologue retained; SOL: spider ohnologue lost; EOL: entelegyne ohnologue lost; AT: ancient tandem duplication; RT: recent tandem duplication; SC: single copy retained. \*The second *ftz* ohnologues in *D. plantarius* and *P. tepidariorum* are likely pseudogenes as there are stop codons in the homeodomains. b) The inferred timing of ancient and recent tandem duplications among the included arthropod lineages. Gene families are listed above branches (as well as below for the root and *H. graminicola*). Species abbreviations are as in Table 2.

(Fig. 1, see Materials and Methods). Our analysis allowed us to distinguish WGD ohnologues from paralogues that arose via tandem duplication for the first time among spiders. We found that many spider homeobox genes and clusters were retained in duplicate following the WGD, within conserved patterns of whole-genome synteny. We also characterized both ancient and recent tandem duplications and determined the timing of these relative to the WGD and the divergence of major lineages within arthropods. We found that the NK cluster was retained in duplicate and largely conserved across spiders. Complementing the previous analysis of *Emx* and *Msx*

family genes (Leite et al. 2018), we described the expression of the rest of the NK cluster genes during embryogenesis in the harvestman *Phalangium opilio*, an outgroup to the arachnophiles that did not undergo the WGD, and their orthologues and ohnologues from the spider *P. tepidariorum*. For these NK cluster genes, we detected no evidence of selection on coding sequences in our comparisons between paralogues or between single copy *P. opilio* genes and their *P. tepidariorum* ohnologues, but for each family with retained ohnologues in *P. tepidariorum*, we found distinct expression patterns for each ohnologue during



**Table 2** Homeobox genes across arthropods are largely stable, except for an increase in the arachnopulmonates. Homeobox families: the number of described bilaterian homeobox gene families with at least one gene present in a genome (Holland et al. 2007; Zhong et al. 2008; Zhong and Holland 2011; Ferrier 2016b). Tandem duplications are divided into ancient (predating the WGD) and recent (after the WGD). Species to the right of the vertical line had an ancestral WGD. Species abbreviations: *Dmel* *Drosophila melanogaster*, *Tcas* *Tribolium castaneum*, *Smar* *Strigamia maritima*, *Isca* *Ixodes scapularis*, *Popi* *Phalangium opilio*, *Cscu* *Centruroides sculpuratus*, *Dsil* *Dysdera silvatica*, *Dpla* *Dolomedes plantarius*, *Lele* *Latrodectus elegans*, *Ptep* *Parasteatoda tepidariorum*, *Hgra* *Hylyphantes graminicola*, *Abru* *Argiope bruennichi*, *Tant* *Trichonephila antipodiana*, *Tcla* *Trichonephila clavata*.

	<i>Dmel</i>	<i>Tcas</i>	<i>Smar</i>	<i>Isca</i>	<i>Popi</i>	<i>Cscu</i>	<i>Dsil</i>	<i>Dpla</i>	<i>Lele</i>	<i>Ptep</i>	<i>Hgra</i>	<i>Abru</i>	<i>Tant</i>	<i>Tcla</i>
Homeobox genes	103	104	114	110	108	161	143	150	152	151	176	149	147	152
Homeobox families	83	85	90	89	91	90	79	87	89	91	86	88	87	89
Families with paralogues from ancient TD(s)	5	5	6	7	8	7	7	8	8	8	8	8	8	8
Families with paralogues from recent TD(s)	9	9	13	8	5	4	8	3	6	4	23	3	2	5
Families with only one gene	71	71	74	76	81	41	35	41	46	44	42	45	44	45
Families with retained ohnologues	–	–	–	–	–	46	41	44	41	44	37	40	40	40

embryogenesis, providing evidence for regulatory sub-functionalization. Combined, these results provide new insight into both the patterns of ohnologue retention and their consequences in spider evolution.

## Results

### Homeobox Gene Repertoires in Arthropods

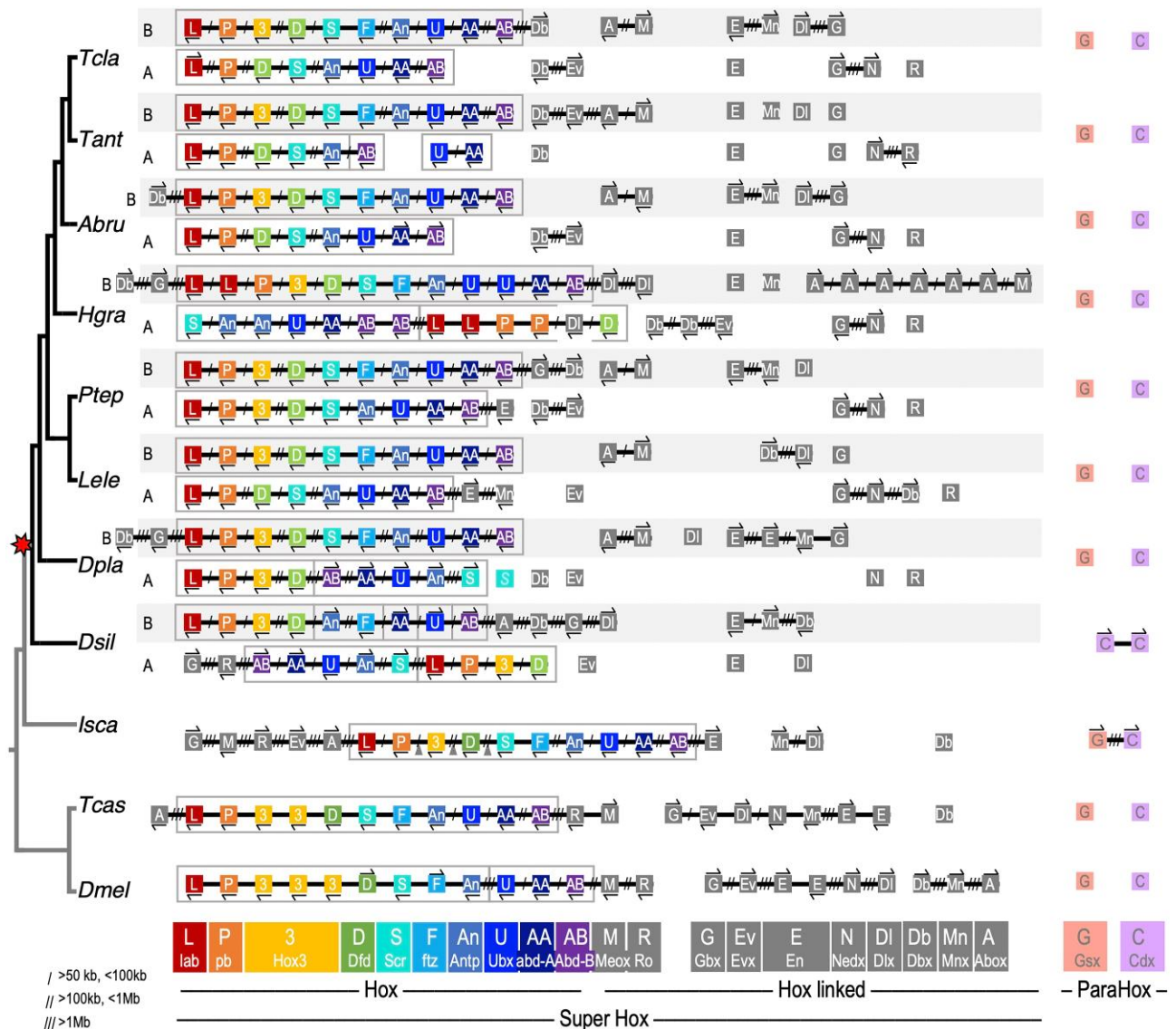
We surveyed the homeobox genes of eight spider species with chromosome-level genome assemblies, representing seven entelegyne and one synspermiatan species and compared these to the scorpion *Centruroides sculpuratus*, the harvestman *P. opilio*, and the tick *Ixodes scapularis*, as well as three mandibulates (Figs. 1 and 2 and Table 2). We found that spiders and scorpions have similar numbers of homeobox genes, and on average, have 1.4 times more homeobox genes than outgroup species that did not have an ancestral WGD (Table 2). The spiders and scorpions had an average of 150 homeobox genes; the most are found in *Hylyphantes graminicola* with 172 and the fewest in *Dysdera silvatica* with 140. Mandibulates and nonarachnopulmonate arachnids had on average 107 homeobox genes, with the fewest in the insects (103) and the most in *Strigamia maritima* (113). Previous surveys found 96 and 69 homeobox genes in *I. scapularis* and *P. opilio*, respectively, 145 in *P. tepidariorum*, and 156 in *C. sculpuratus* (Schwager et al. 2017; Leite et al. 2018), but we identified between 1% and 55% more in each of these species (Table 2). For *P. opilio*, searching both the recent genome assembly (Gainett et al. 2021) as well as the developmental transcriptome (Sharma et al. 2012) provided a more complete dataset and enabled the identification of more homeobox genes than previously described.

In total, 90 described homeobox families were represented in our arthropod species set (Table 2), though some were absent in different species, such that no species had a gene from every family (Fig. 2a, supplementary table S1, Supplementary Material online). In addition, we also found several new homeobox genes. A few of these may represent new lineage-specific unnamed (“un”) families, temporarily called *ANTP-un1*, *PRD-un1*, *SINE-un1*, and *TALE-un1* based on the classes to which they likely belong. *ANTP-un1* is found only in the chelicerates, while *PRD-un1* and *SINE-un1* were found only in the araneids, and *TALE-un1* only in the *Trichonephila* spp. (Fig. 2a, supplementary table S1, Supplementary

Material online). The “1” is simply a placeholder to indicate that these genes have an orthologue in at least two species and to distinguish them from other unnamed or unclassified homeodomain sequences that have no similarity to described families and were found only in one species, sometimes in several copies.

Comparing this taxonomic breadth of species in concert with genomic synteny enabled us to distinguish ancient and recent tandem duplications from likely ohnologues and determine the timing of duplications relative to major radiations in arthropod evolution. We have defined ancient tandem duplicates as paralogues that are often linked or clustered in most spider genomes and date to a duplication node in an ancestor before the arachnopulmonate WGD, while recent tandem duplications are also linked or clustered, but have occurred either after the WGD or in the outgroups. Six homeobox gene families, *Arx*, *Barhl*, *Irx*, *Pax3/7*, *Pax4/6*, and *Lhx2/9*, underwent ancient tandem duplication events that are shared by other arthropods and may have ancestry even deeper in the bilaterian phylogeny (Fig. 2b). Only three of these ancient tandem paralogues have been subsequently lost; just one *Lhx2/9* gene is found in *Drosophila melanogaster*, and only one *Barhl* gene is found in *Tribolium castaneum* and *D. silvatica*. Within the chelicerates, but predating the arachnopulmonate WGD, two further tandem duplications occurred. *Msx* likely duplicated in the common ancestor of ticks, harvestmen, and arachnopulmonates, and *Emx* likely duplicated in the ancestor of harvestmen and arachnopulmonates. *NK1* duplicated in tandem, likely just after the WGD, while the other ohnologue was probably lost. Besides these duplications, the majority of tandem duplications in spiders were more recent and lineage-specific (Fig. 2b). In particular, *H. graminicola* has the highest number of gene families with recent tandem duplications (23 families) (Fig. 2b), resulting in this species having the highest number of homeobox genes identified (172). This could be an assembly artifact; however, overall the *H. graminicola* genome did not exhibit an exceptionally large number of duplicated genes (Zhu et al. 2022).

In contrast to tandem duplicates, dispersed paralogues consistently found on different chromosomes in surveyed spider genomes were inferred to be ohnologues. There were between 37 and 46 families with retained ohnologues in arachnopulmonates, and the families that were retained



**Fig. 3.** Two ohnologous Hox clusters and a disintegrated arthropod ParaHox cluster are found in spiders. Genes are represented by colored boxes; chromosomes are black lines. Gene orientation is indicated by arrows above or below genes, and distances are indicated by slashes between genes. Gene families are denoted by a one- or two-letter code, and by color for Hox and ParaHox genes, or in gray for Hox-linked genes. ParaHox genes are colored based on their relationship to Hox genes, *Gsx* with anterior genes, and *Cdx* with posterior genes; *Xlox* was not found in any arthropod genome. Species abbreviations are as in Table 2.

were largely consistent across species. All the Hox genes besides *Hox3* and *ftz* were retained as ohnologues, as were a few Super-Hox genes, namely *Dbx*, *En*, and *Gbx* (Fig. 2a). Many of the NK cluster genes were also retained as ohnologues, including *NK4*, *NK5*, *NK6*, *NK7*, *Lbx*, *Tlx*, as well as the ohnologues of ancient tandem paralogues of *Emx* and *Msx* (Fig. 2a).

Ohnologues were likely lost from several families following the WGD (Fig. 2a). Thirty-two families are found in single copies in scorpions and spiders (discounting subsequent lineage-specific tandem duplications of single-copy genes), indicating that an ohnologue was lost between the WGD and the divergence of scorpions and spiders. These include many of the Super-Hox families (*Abox*, *Dlx*, *Evx*, *Meox*, *Mnx*, *Nedx*, and *Ro*), the ParaHox family *Gsx*, most of the NK2 cluster families (*Hlx*, *Msx1x*, and *NK2.1*) and the HRO cluster family *Hbn*. Six

families, including the Hox family *ftz*, the ParaHox family *Cdx* and the HRO cluster family *Otp*, were found in single copies in spiders, while both ohnologues were retained in the scorpion, indicating the loss of an ohnologue in a spider ancestor (Fig. 2a). Within spiders, an ohnologue was lost from four families in the entelegyne ancestor: *Dlx*, *Six4/5*, *Msx1x*, and *NK3*. *D. silvatica*, the sole synspermiatan representative, has lost the largest number of ohnologues from eighteen families, though the extent to which this applies across synspermiatans is not known. Only a few families underwent losses multiple times, namely *Hox3* in *Latrodectus elegans* and the araneids; *Tgif* in the theridiids, *Dolomedes plantarius*, *D. silvatica*, and *C. sculpturatus*; and *Barx* in both *H. graminicola* and *D. silvatica*. Only one family, *Vax*, was potentially lost before the WGD, or both ohnologues were lost

in the arachnoplumonate ancestor (Fig. 2a). Though it was not reported previously (Leite et al. 2018), we identified a single *Vax* gene in the harvestman genome (Fig. 2a, supplementary table S1, Supplementary Material online).

## Homeobox Gene Clusters in Spider Genomes

### The Hox Cluster

Each of the spider genomes studied here contains two Hox clusters (Fig. 3), consistent with previous descriptions for *P. tepidariorum* and other spiders (Schwager et al. 2007, 2017; Fan et al. 2021; Sheffer et al. 2021; Wang et al. 2022). With the new *P. tepidariorum* chromosome-level assembly, we confirmed that each cluster is found on a different chromosome. Cluster B is typically intact, organized, and contains orthologues of all ten arthropod Hox genes, while cluster A often lacks *Hox3* and *ftz*, and is sometimes rearranged (Fig. 3). There are some exceptions to this, however. For one, there have been several tandem duplications of *H. graminicola* Hox genes resulting in four *lab* paralogues, and three *pb*, *Antp*, *Ubx*, and *Abd-B* paralogues (Fig. 3). While these duplications have occurred in both *H. graminicola* Hox clusters, cluster B is ordered, and contains the single orthologues of *Hox3* and *ftz*, while cluster A lacks those genes and has been rearranged (Fig. 3). Furthermore, the Super-Hox gene *Dlx-A* is between *pb-A-2* and *Dfd-A* separated by 124 kb from *pb-A-2* and 380 kb from *Dlx-A* and *Dfd-A*. In *D. plantarius* there is a new gene, identified with highest similarity in the homeodomain to *Scr*, but not linked to either Hox cluster, possibly representing a divergent *Scr* paralogue (blue “S” in Fig. 3). In *D. silvatica*, both Hox clusters are rearranged, though one does contain both *Hox3* and *ftz*, while the other has lost only *ftz* (Fig. 3). Pseudogenes of *ftz* were found in *D. plantarius* and *P. tepidariorum*, identified by the homeodomain disrupted by a stop codon. In *P. tepidariorum*, the *ftz*-pseudogene is not linked to either Hox cluster, while in *D. plantarius*, it is adjacent to *ftz-B* (supplementary table S1 and fig. S3F, Supplementary Material online).

The Super-Hox genes consist of the Hox cluster as well as several ANTP-class genes consistently found linked to each other and/or the Hox cluster across bilaterians (Table 1) (Butts et al. 2008; Ferrier 2016b). While many Super-Hox genes are still linked to one another and/or to the Hox cluster in the tick and insects, Super-Hox genes have dispersed in spider genomes and many Super-Hox families have also returned to single copy (Figs. 2a and 3). Amid the overall disintegration of the Super-Hox cluster, *Abox* and *Meox* are consistently found clustered in all the entelegyne species. Other remnants of linkage between Super-Hox genes include *Dbx* and *Gbx* linked to Hox cluster B in *H. graminicola*, *P. tepidariorum*, and *D. plantarius* (Fig. 3). *En* is also often linked to *Mnx*, *Gbx* to *Nedx*, and *Dbx* to *Evx*, though these have more frequently been disrupted and dispersed, unlike the consistency of *Abox* and *Meox*. Synteny of Super-Hox genes may be conserved by the constraint of regulatory elements or may result from the passive maintenance of ancestral linkage. While

the Hox cluster has largely been retained in duplicate in spiders, most of the Super-Hox families have returned to the single copy and linkage has dispersed following the WGD.

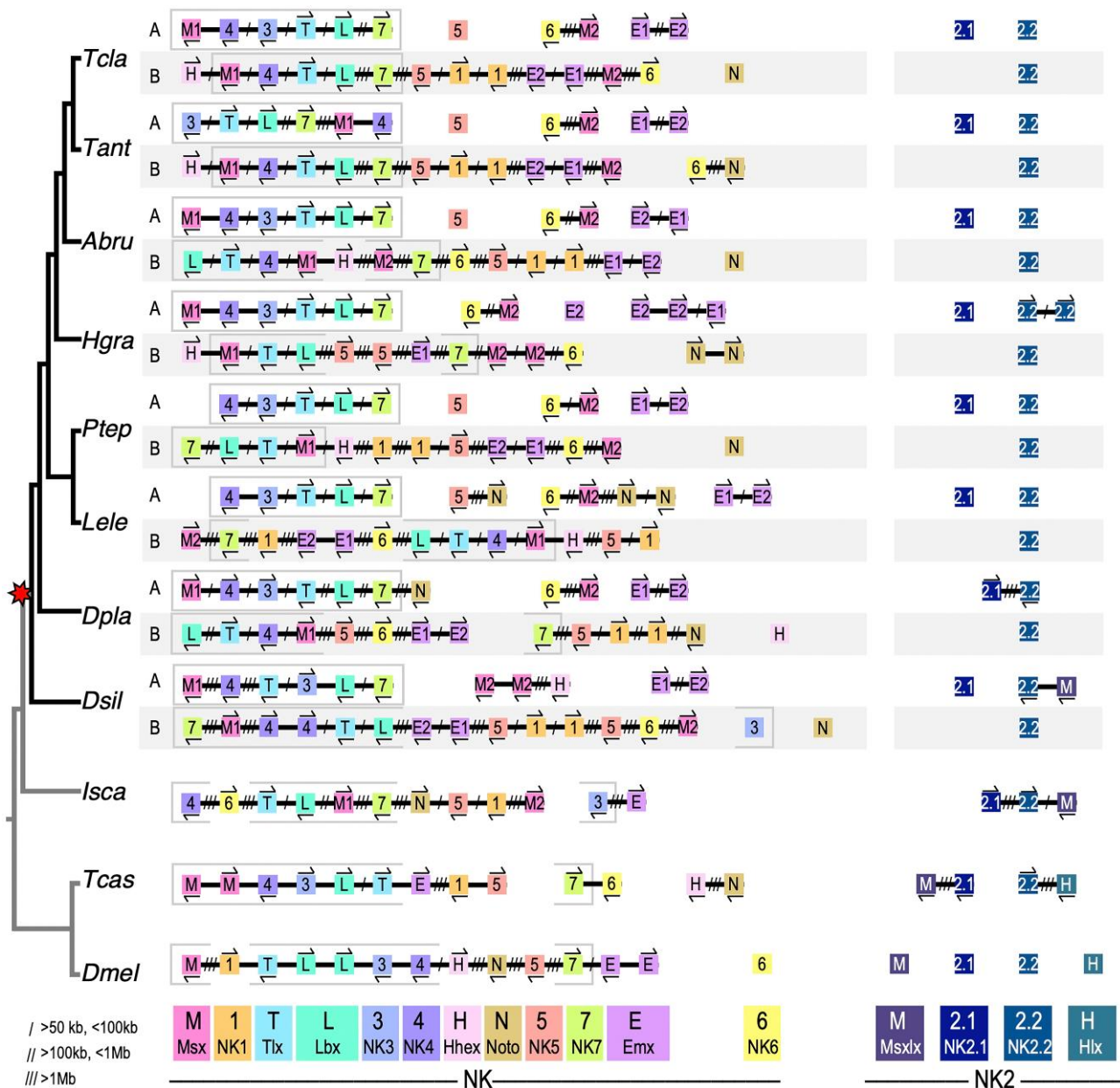
The ParaHox cluster has largely disintegrated in spiders. Neither ParaHox gene is linked to the other, and both have returned to single copy (Fig. 3). The two *Cdx* paralogues in *D. silvatica* likely arose in a recent lineage-specific tandem duplication, as they form a two-gene cluster in only this species. In the insects *D. melanogaster* and *T. castaneum*, *Gsx* (*intermediate neuroblasts defective* (*ind*)) and *Cdx* (*caudal* (*cad*)) are unlinked, while in the tick they are on the same chromosome but separated by a large distance (28 Mb) (Fig. 3). This is consistent with the ancestral breakdown of the ParaHox cluster in an arthropod ancestor, and the further separation of the genes onto separate chromosomes in insects and spiders (Fig. 3). It is not clear whether the retained ParaHox genes belong to the same ohnology group or not (i.e. both A or both B versus one A and one B), as there is only one ohnologue retained in any of the species surveyed.

### The NK Cluster

We found two ohnologous NK clusters in all spider genomes (Fig. 4). One consists of *Msx1-A*, *NK4-A*, *NK3-A*, *Tlx-A*, *Lbx-A*, and *NK7-A* clustered on one chromosome, while *Msx1-B*, *Tlx-B*, *Lbx-B*, and *NK7-B* are found on a different chromosome, also with *NK3-B* and *NK4-B* in the species that have retained these genes (core NK clusters are delineated with gray boxes in Fig. 4). Cluster B is less organized; *NK3-B* has likely been lost in entelegynes (estimated divergence 280 to 310 MYA) (Fernández, Kallal, et al. 2018; Shao and Li 2018) and *NK4-B* has likely been lost in *H. graminicola* and *P. tepidariorum*, independently. NK core cluster A is conserved across spiders, except that *Msx1-A* is not found in *L. elegans* or *P. tepidariorum*, which may indicate the loss of *Msx1-A* in a theridiid ancestor (est. 120 MYA) (Fernández, Kallal, et al. 2018). While NK cluster A is found in a conserved complement and order in almost all spider species analyzed, it is rarely linked to any other genes belonging to the larger ancestral bilaterian NK complement (Table 1). In contrast, NK cluster B has a varying gene complement, having undergone lineage-specific gene losses and more rearrangements, making it more degenerate and less conserved between species. This second cluster is, however, often linked to other ancestral NK cluster genes *Hhex*, *Noto*, *Emx1-B*, *Emx2-B*, *NK5-B*, and *NK1*, which has undergone a tandem duplication, and to *NK6-B* and *Msx2-B* in all spiders (Fig. 4). *NK6-A* is found linked to *Msx2-A* in all the entelegyne species, and *NK6* and *Msx2* genes can be found in cluster A and/or B in all spiders, suggesting this two-gene *NK6-Msx2* cluster predates the WGD and was retained in duplicate some species, while others have lost either *Msx2* or *NK6* ohnologues.

The NK2 cluster has largely broken down in spiders, with successive gene losses: first *Hlx* from chelicerates, and then *Msx1x* from entelegynes, though further lineage sampling may change this pattern (Fig. 4). In all spiders,





**Fig. 4.** Two ohnologous NK clusters are retained, and a disintegrated NK2 cluster is found in spiders. Genes are represented by colored boxes; chromosomes are black lines. Gene orientation is indicated by arrows above or below genes, and distances are indicated by slashes between genes. Gene families are denoted by color and a one- or two-letter code. The genes making up the spider core NK cluster are delineated by gray boxes. Species abbreviations are as in Table 2.

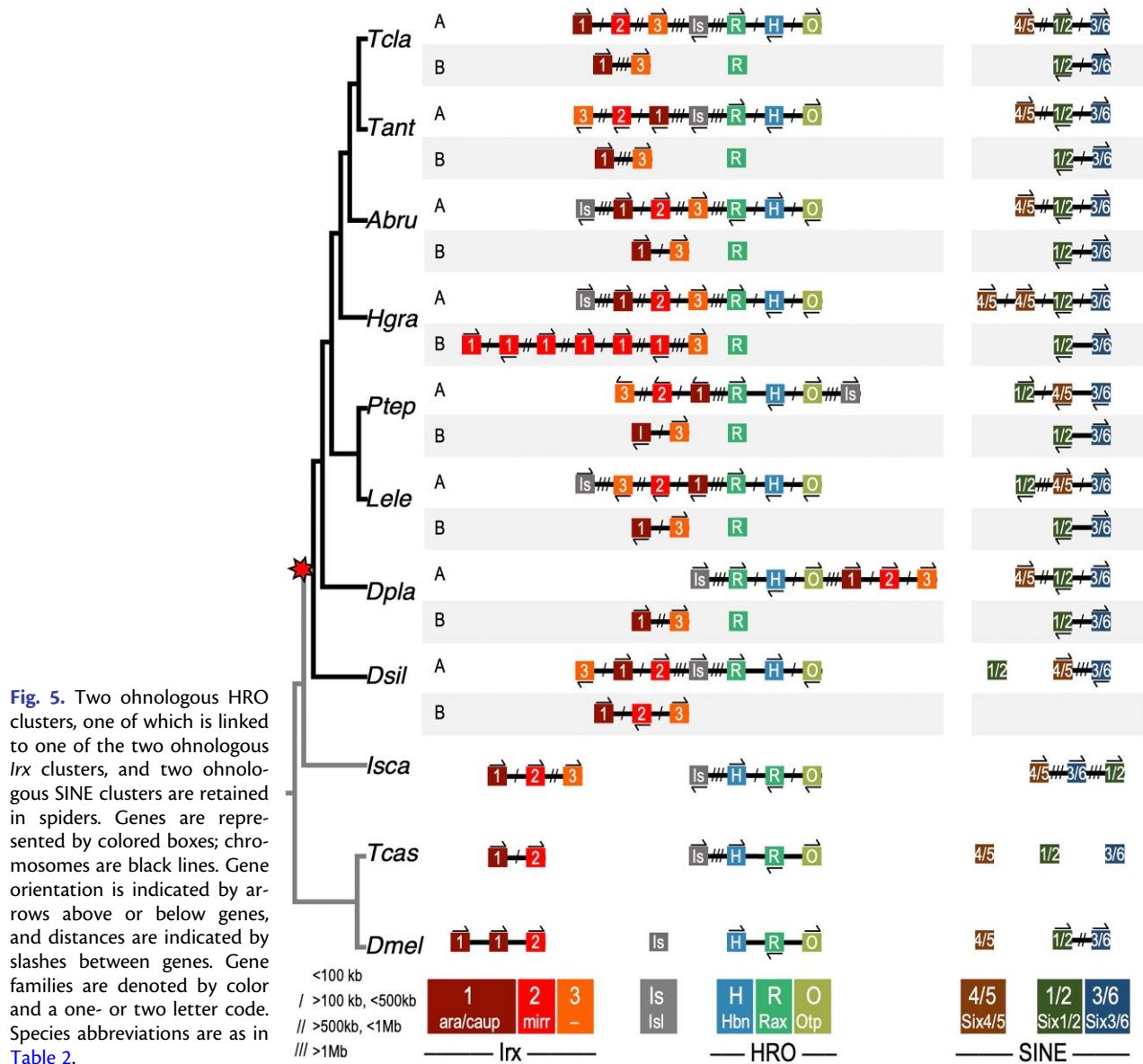
NK2.2 is retained in duplicate on two separate chromosomes, indicating an ohnologous relationship, while NK2.1 has returned to the single copy (Fig. 4). NK2.1 and NK2.2 are unlinked in all spiders except for *D. plantarius*, indicating either that the cluster has separated twice, or alternatively, that different ohnologies have been lost in different lineages. Phylogenetics were not able to determine which ohnologue of NK2.1 was retained in different species because no examined genome contains both ohnologies.

#### The SINE Cluster

There are five SINE-class genes in two clusters in spiders, consistent with a duplication in the WGD followed by

the loss of one *Six4/5* ohnologue. Spider SINE clusters likely underwent different rearrangements in different lineages (Fig. 5). *D. silvatica* has retained only one copy of each gene, while the entelegynes have retained B ohnologies of both *Six1/2* and *Six3/6* in a conserved cluster. In araneids, in cluster A all three genes are linked, though *Six4/5-A* is separated by a larger distance than that separating *Six1/2-A* from *Six3/6-A*, which remain tightly clustered. There has also been an inversion which may have occurred ancestrally among the theridiids that brought *Six4/5-A* closer to *Six3/6-A* than *Six1/2-A*. This indicates that the separation of *Six4/5-A* from the other two genes in araneids occurred after the divergence of the theridiids and that the three





genes making up SINE cluster A were likely still tightly clustered in their shared ancestor. Among the outgroups, all three SINE genes are linked in the tick, and only two are linked in the fly, while none are linked in the beetle, indicating the spider SINE clusters have been retained more intact relative to other arthropod lineages.

#### The HRO Cluster

The PRD-class HRO cluster was so named because the order of the genes in the fly genome is *homeobrain* (*hbn*), *Retial homeobox* (*Rx*), then *orthopedia* (*otp*); this order is conserved in insects and the tick genome, indicating that it is likely the ancestral arthropod arrangement (Fig. 5). In spiders, the HRO cluster is found intact, but reordered: *Rax*, *Hbn*, then *Otp* (Fig. 5). There have also been several inversions of the orientations of the genes within the cluster, but the R-H-O order is conserved across spiders, suggesting this rearrangement occurred once ancestrally. It likely did not predate the WGD but occurred in a spider

ancestor because the scorpion HRO cluster A order is *Hbn*, *Rax*, then *Otp*, as in insects and the tick (supplementary table S1, Supplementary Material online). As well as one conserved intact HRO cluster, spiders retained only *Rax* in duplicate, while there are two *Rax* and two *Otp* ohnologues retained in the scorpion (Figs. 2 and 5).

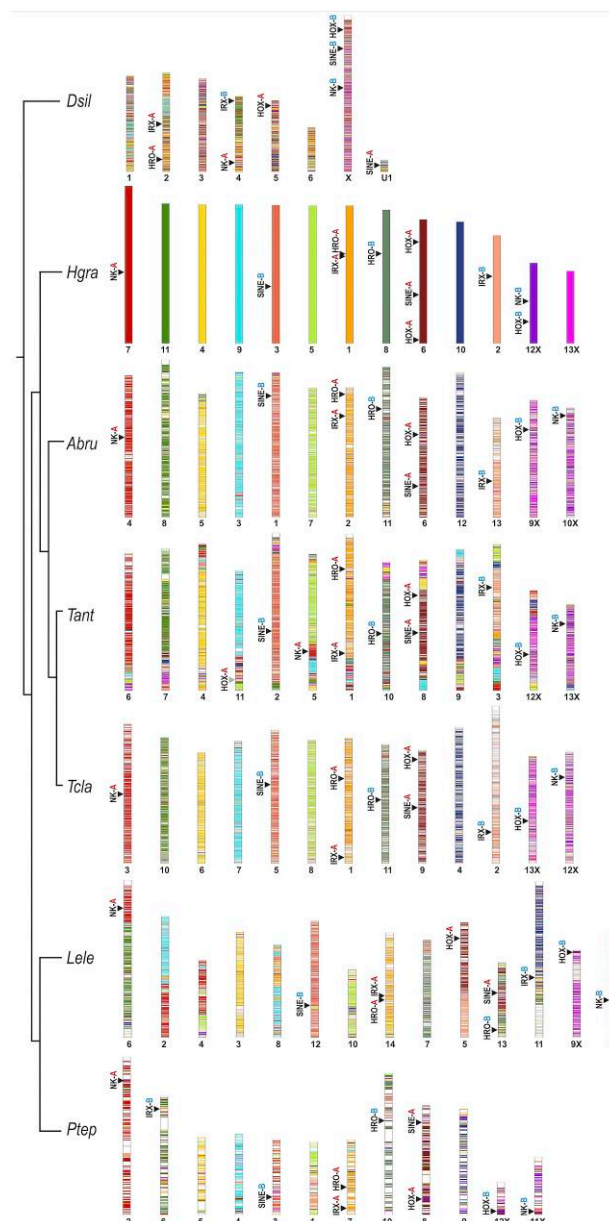
The LIM-class gene *Isl* is consistently found linked to the HRO cluster (Fig. 5). Ferrier (2016) suggested this pattern is evidence for the ancestral linkage between the LIM-class and PRD-class genes in the Giga-cluster of homeobox progenitor genes in an animal ancestor. The linkage of *Isl* to the HRO cluster has been observed across species with a most recent common ancestor in the ancestor of all animals. The consistent linkage pattern between these genes in spiders, the tick, and beetle genomes supports this association. This broad phylogenetic distribution makes this between-class linkage unlikely to be a convergent secondary

association; rather it is more likely to be a conserved ancestral trait.

### The *lrx* Cluster

*Lrx* genes have duplicated independently several times (see Table 1), giving rise to six genes in spiders. Each of the ohnologous *lrx* clusters contains one of each gene type, *lrx1*, *lrx2*, and *lrx3*, though *lrx2-B* has been lost from one of the clusters in the entelegynes (Fig. 5). Consistent with the findings of Kerner et al. (2009), insect *lrx* genes fall into two clades, one containing orthologues of *D. melanogaster* mirror (*mirr*) and the other with orthologues of *araucan* (*ara*) and *caupolican* (*caup*) (Table 1 and supplementary fig. S1, Supplementary Material online). Spider and scorpion *lrx1-A* and *lrx1-B* group with *ara* and *caup*, while spider and scorpion *lrx2-A* and *lrx2-B* group with *mirr* (supplementary fig. S1, Supplementary Material online). The exact relationships between *lrx3* and other arthropod genes remain somewhat unclear. There are three *lrx* genes in the centipede *S. maritima* and the millipede *Trigoniulus corallinus* genomes (Chipman et al. 2014; So et al. 2022), one of which is not contained within either the *lrx1/ara/caup* or *lrx2/mirr* clade, and is resolved as sister to the *lrx3* clade in the *lrx* gene phylogeny, albeit with very low bootstrap support (supplementary fig. S1, Supplementary Material online). The potential *S. maritima* *lrx3* is clustered with *lrx1* and *lrx2* in the same relative position as spider *lrx3* (Figs. 5 and supplementary S3C, Supplementary Material online). There is a clear orthologue of each of the three arachnoplumonate *lrx* gene families in both the harvestman and tick genomes, indicating the three gene types originated in or before the chelicerate ancestor, and possibly *lrx3* genes were lost from an insect ancestor after the divergence of Myriapoda (est. 520 MYA) (Fernández, Edgecombe, et al. 2018). According to Kerner et al. (2009), *lrx* genes duplicated independently in spirarians, ecdysozoans, and deuterostomes, suggesting that all paralogues arose from a single ancestral bilaterian *lrx* gene. From our phylogeny (supplementary fig. S1, Supplementary Material online), we can tentatively infer that the duplication that gave rise to fly paralogues is ancestral to the arthropods and that there was a three-gene cluster in the ancestral chelicerate that further duplicated in the arachnoplumonate WGD resulting in up to six spiders *lrx* paralogues in two ohnologous clusters.

In spiders, one of the two ohnologous *lrx* clusters is also linked to the HRO + *Isl* cluster (Fig. 5). Since this association occurs only to *lrx* cluster A, while *lrx* cluster B is consistently not linked to the second *Rax* paralogue (Fig. 5, supplementary fig. S3E-L, Supplementary Material online), a likely explanation is that one *lrx* cluster translocated to be linked to HRO + *Isl* following the WGD. We infer it occurred secondarily, i.e. on only one ancestral arachnoplumonate chromosome, rather than deeper in animal ancestry due to the consistent lack of linkage between HRO + *Isl* with *lrx* genes in any other species examined here, in contrast to the linkage between HRO and *Isl* found across bilaterians.



**Fig. 6.** Conserved macrosynteny of spider chromosomes and associated locations of homeobox gene clusters. Vertical bars show stacks of genes on each chromosome representing a macrosynteny block, with *H. graminicola* as the reference genome. Chromosomes of other species were ordered to show the orthologous relationships to *H. graminicola*. Chromosome numbers are denoted below, with "X" indicating sex chromosomes. Locations of HOX, NK, *lrx*, SINE and HRO clusters are shown with black arrows, along with the assignment for cluster A (red) versus cluster B (blue). The gray triangle in *T. antipodiana* indicates the partial translocation of Hox cluster A.

### Conserved Macrosynteny of Spider Genomes

To contextualize our observations of homeobox cluster conservation and rearrangements, we explored macrosynteny across spider chromosomes. We used *H. graminicola* as the reference because the *D. silvatica* genome is highly rearranged compared to all entelegynes in this study, and *H. graminicola* represents an early-branching lineage relative to the majority of other entelegynes (Fig. 6, supplementary fig. S4A, Supplementary Material online).

Across entelegynes, we detected clear patterns of orthology between chromosomes, and consistency in the synteny blocks in which certain homeobox clusters are found, but these patterns were not as evident when compared to the sole synspermiatan representative *D. silvatica* (Fig. 6, supplementary fig. S4A, Supplementary Material online). Depictions of locations of all homeobox genes across genomes can be found in supplementary fig. S3, Supplementary Material online.

In each araneid analyzed here, there are thirteen chromosomes; however, in the theridiids, there are twelve in *P. tepidariorum* and fourteen in *L. elegans* (Fig. 6, supplementary fig. S4A, Supplementary Material online). In *P. tepidariorum*, the synteny block corresponding to chromosome 2 in *H. graminicola* (coral pink in Fig. 6) appears to have been mostly lost or distributed throughout the genome, and *lrx* cluster B which belongs to this block has been relocated (Fig. 6, supplementary fig. S4A, Supplementary Material online). In *L. elegans* there appear to have been large fission and fusion events on chromosomes 2, 4, 5, 6, 8, 11, and 13. Such events in *L. elegans* are consistent with Hox cluster A and SINE cluster A being found on separate chromosomes, as well as the rearrangement of *lrx* cluster B (Figs. 6 and supplementary S3G and fig. S4A, Supplementary Material online).

Sex chromosomes have been identified in several spider species (Fig. 6 and supplementary fig. S4A, Supplementary Material online) (Sheffer et al. 2021; Zhu et al. 2022; Miller et al. 2023). Interestingly, Hox cluster B and NK cluster B are both found on the same sex chromosome in *H. graminicola* (12) and *D. silvatica* (X), but on separate sex chromosomes in other entelegynes, whereas Hox cluster A is located on an autosome (Fig. 6, supplementary fig. S4A, Supplementary Material online). Across entelegynes, NK cluster A is found in the same conserved synteny block corresponding to a full chromosome in all species except for *T. antipodiana*. In this species, a rearrangement brought NK cluster A onto another chromosome, and this is the only entelegyne species in which the NK cluster A is rearranged (Fig. 4). Overall, this analysis illustrates the conservation of macrosynteny blocks and chromosomes across entelegyne genomes and shows several chromosomal rearrangements that have affected homeobox gene clusters.

### Analysis of Selection of NK Cluster Genes

We next sought to understand how homeobox genes have evolved following WGD and tandem duplications. In the NK cluster, we identified families with single-copy orthologues, ancient tandem paralogues, and WGD ohnologues. To determine if genes with different relationships were under different selective pressures, we compared the aBSREL estimated  $\omega$  values calculated across coding sequence phylogenies for seven NK cluster families with different inferred duplication histories: *Msx*, *NK4*, *NK3*, *Tlx*, *Lbx*, *NK7*, and *Emx*. While none of the focal ohnologues or tandem paralogues was found to be under positive selection, certain genes from the preduplicate outgroups were. *NK7* was found to be under positive selection in

*D. melanogaster* and *I. scapularis*, *Tlx* also in *I. scapularis*, and *NK3* in *T. castaneum*. However, certain ohnologues did have longer branches separating them from other ohnologues or the single-copy outgroup genes in the aBSREL-generated phylogenies, indicating a higher substitution rate. These include *Msx1* in arachnoplumonates and the harvestman, *NK3-A* in the entelegynes, and *Lbx-B* in the scorpion. No consistent pattern was detected between A versus B ohnologues, despite NK cluster A being more organized, nor was there a clear pattern between ohnologues versus tandem duplicates. Supplementary Material online underpinning these results is available on GitHub (see Materials and Methods).

In the one-to-one comparisons between *P. tepidariorum* and *P. opilio* genes, and between *P. tepidariorum* ohnologues, we calculated dN/dS ratios with codeML to detect positive selection. For all NK cluster genes, none approached 1; all were very low, ranging from 0.0011 (between *P. opilio Msx2* and *P. tepidariorum Msx2-A*) to 0.1423 (between *P. tepidariorum Emx1-A* and *Emx1-B*) (Tables 3 and supplementary S3, Supplementary Material online). There are some patterns among gene families, e.g. lower dN/dS ratios for *Tlx* ohnologues compared to a consistently higher rate for *NK7* ohnologues. The *Emx1* and *Emx2* and *Msx1* and *Msx2* ohnologues had consistently low dN/dS values when compared to their *P. opilio* orthologues but varied when compared between ohnologues in *P. tepidariorum*. Overall, these low dN/dS ratios likely reflect the highly conserved nature of homeobox genes.

### Developmental Expression of NK Cluster Genes in *P. tepidariorum* and *P. opilio*

It was previously shown that several ohnologues of spider homeobox genes had undergone sub- and/or neofunctionalization with respect to their single-copy orthologues in the harvestman (Leite et al. 2018). To better understand the roles of homeobox genes following WGD in the arachnoplumonate ancestor, we assayed the expression of NK cluster genes in the spider *P. tepidariorum* in comparison to the harvestman *P. opilio*.

#### Single-Copy Genes

Only one copy each of *NK3* and *NK4* was found in both *P. opilio* and *P. tepidariorum*, indicating the loss of an ohnologue of each from the spider following the WGD. The single spider *NK3* is strongly expressed in the developing heart and is weakly expressed around the stomodaeum and in the head (Fig. 7a to e). This is very similar to the single spider *NK4*, which is also expressed strongly in the heart (Fig. 7f to j). Unlike *NK3*, *NK4* is also expressed in a dot-like fashion in the ventral nervous system of the opisthosoma, and at the anterior rim of the head lobes (Fig. 7 asterisks and double arrowheads in Fig. 7f to i). We did not detect any *NK3* expression in the developing heart of the harvestman; instead, *NK3* is expressed in a segmental fashion ventral to the base of the appendages and in similar short transverse segmental stripes in the opisthosomal segments (Fig. 7k



**Table 3** The ratio of nonsynonymous to synonymous substitution rates between *P. tepidariorum* and *P. opilio* NK cluster genes shows no evidence of positive selection among paralogues. Pairwise dN/dS values were calculated with codeML for each *P. opilio* and *P. tepidariorum* NK cluster gene. The bars at the right show the dN/dS value relative to the black bar (length of 1) at the top. The relationship between each gene pair is indicated in the third column, where 1 to 1 indicates direct orthology (the loss of an ohnologue), 1 to 2 indicates the retention of two ohnologues in the spider, O is the comparison between the two spider ohnologues, and T is the comparison between tandem paralogues. For families with ancient tandem duplications (*Msx* and *Emx*), gene names have been updated to reflect the inferred relationships between paralogues in *P. opilio* and subsequent ohnologues in *P. tepidariorum*, such that, e.g. *P. opilio* *Emx1* is the pro-orthologue of *P. tepidariorum* *Emx1-A* and *Emx1-B*, and the same for *P. opilio* *Emx2* and *P. tepidariorum* *Emx2-A* and *Emx2-B*. Raw substitutions (S and N) and substitution rates (dS and dN) as calculated with codeML can be found in [supplementary table S3, Supplementary Material](#) online.

Popi	Ptep		dN/dS	
NK4	NK4	1 to 1	0.0673	
NK3	NK3	1 to 1	0.0584	
Tlx	Tlx-B	1 to 2	0.0060	
	Tlx-A	1 to 2	0.0094	
	Tlx-A x Tlx-B	O	0.0042	
Lbx	Lbx-B	1 to 2	0.0421	
	Lbx-A	1 to 2	0.0299	
	Lbx-A x Lbx-B	O	0.0076	
NK7	NK7-B	1 to 2	0.0642	
	NK7-A	1 to 2	0.0430	
	NK7-A x NK7-B	O	0.0575	
Msx1	Msx1-B	1 to 1	0.0104	
Msx2	Msx2-A	1 to 2	0.0011	
	Msx2-B	1 to 2	0.0106	
	Msx2-A x Msx2-B	O	0.0051	
	Msx1-B x Msx2-B	T	0.0845	
Emx1	Emx1-A	1 to 2	0.0123	
	Emx1-B	1 to 2	0.0128	
Emx2	Emx2-A	1 to 2	0.0120	
	Emx2-B	1 to 2	0.0094	
	Emx1-A x Emx1-B	O	0.1423	
	Emx2-A x Emx2-B	O	0.0118	
	Emx1-A x Emx2-A	T	0.0188	
	Emx1-B x Emx2-B	T	0.0124	

to n). The expression of harvestman NK3 is thus more like the expression of its orthologue in *D. melanogaster* (Azpiazu and Frasch 1993) and the millipede *Glomeris marginata* (supplementary fig. S5, Supplementary Material online) than that of the spider. Expression of NK4 in the harvestman is restricted to the developing dorsal vessel at late developmental stages (Fig. 7o and p), a pattern that is conserved among panarthropods (Azpiazu and Frasch 1993; Janssen and Damen 2008; Treffkorn et al. 2018).

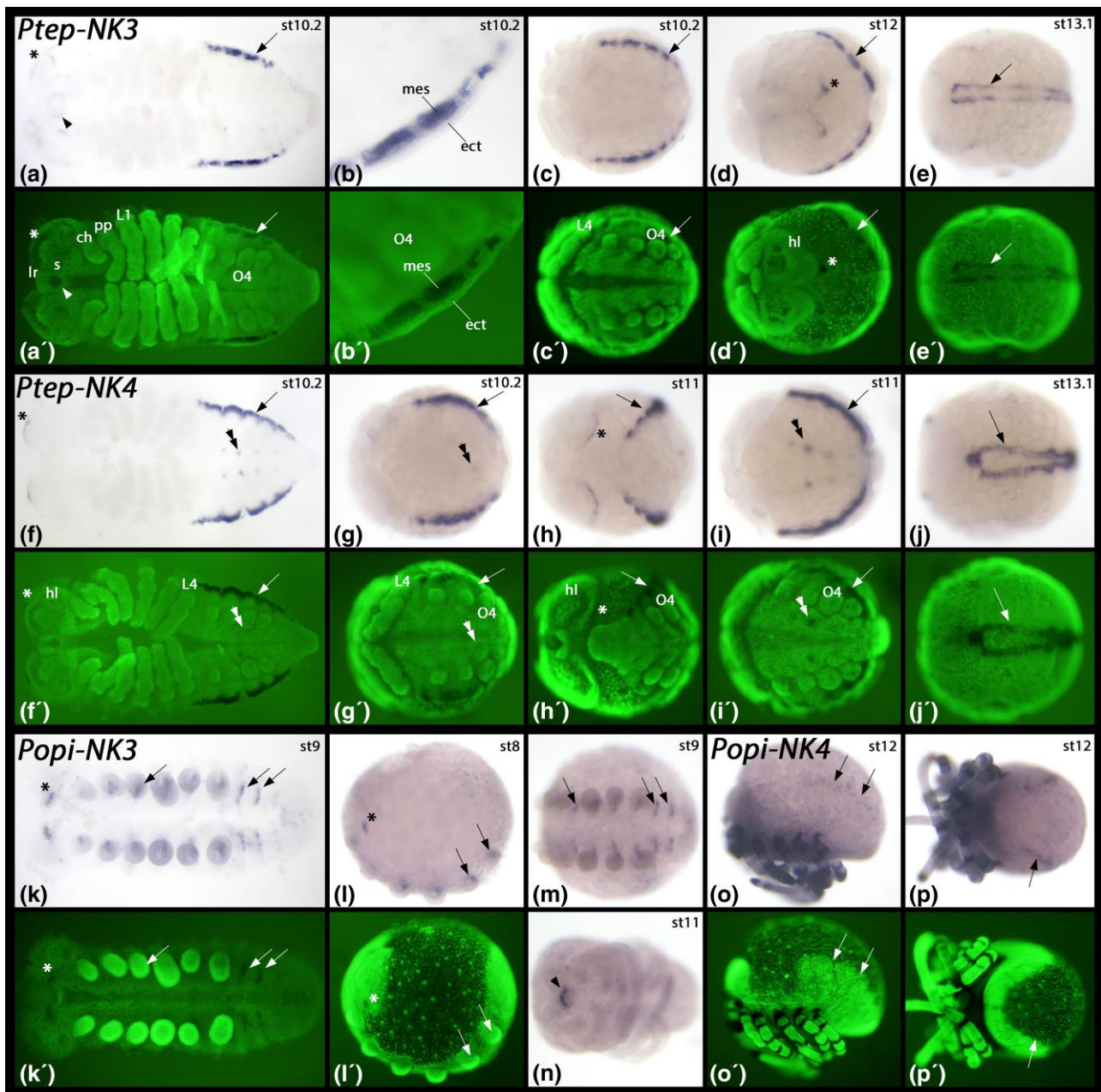
#### Ancient Tandem Paralogues

Both *Emx* and *Msx* families have likely undergone an ancient tandem duplication, resulting in two paralogues of each gene in the harvestman genome, with four *Emx* and three *Msx* paralogues retained in *P. tepidariorum*. In previous transcriptome-based analyses of *P. opilio*, the highly similar sequences for each *Emx* paralogue were merged into a single transcript model, and only a partial sequence was found for *Msx1*, split across two different transcript models (Sharma et al. 2012), but the genome revealed distinct loci for each paralogue. In *P. tepidariorum*, Leite et al. (2018) described

*Emx1-A* and *Emx2-A* (previously designated *Emx4* and *Emx3*, respectively) expression in the precheliceral region and in patches in each segment along the ventral midline, and *Emx1-B* and *Emx2-B* expression (previously designated *Emx2* and *Emx1*) in the anterior of each opisthosomal segment, with *Emx1-B* also expressed at the base of the prosomal appendages. *P. opilio* *Emx2* (previously designated *Emx*) was detected at the base of the appendages and in patches along the ventral midline in stage 10 to 11 embryos (Leite et al. 2018).

Very similar to its paralogue *Emx2*, *P. opilio* *Emx1* is expressed in all posterior segments including the cheliceral segment (Fig. 8). The sharp anterior border of expression is conserved among other arthropods, including spiders (Walldorf and Gehring 1992; Simonnet et al. 2006; Schinko et al. 2008; Birkan et al. 2011; Janssen 2017; Leite et al. 2018) and an onychophoran (Janssen 2017). Indeed, the overall expression pattern of *P. opilio* *Emx1* is identical to that of the previously reported *Emx2*, with the exception that expression of *Emx1* is detected earlier.

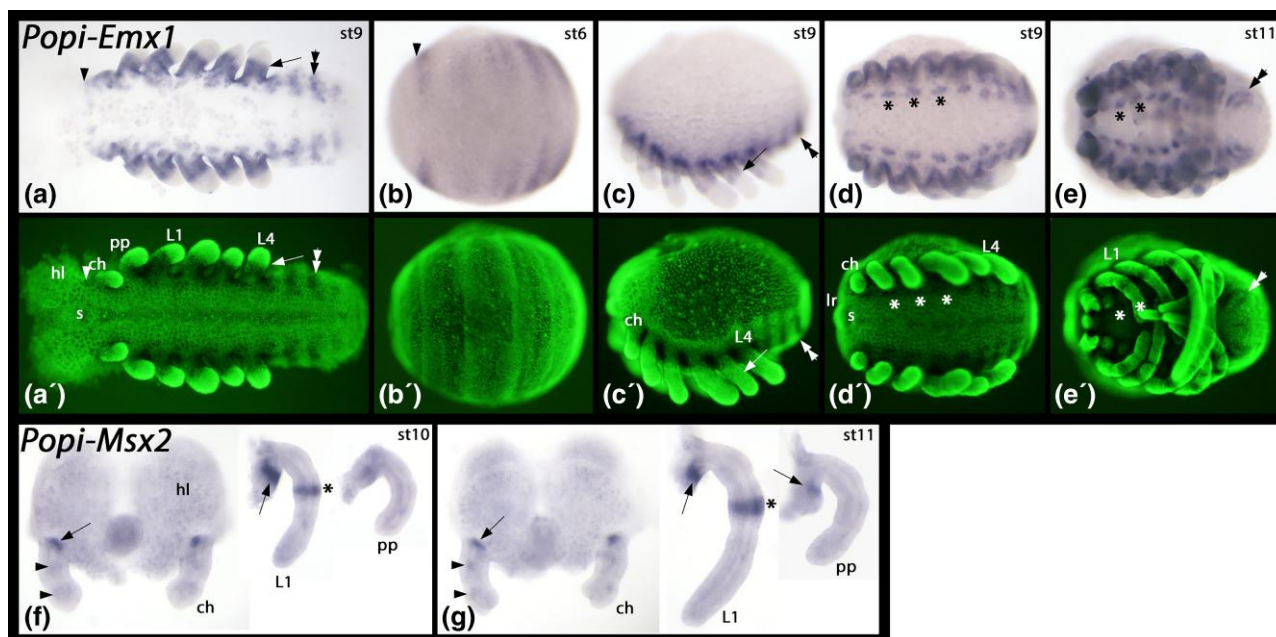
According to Leite et al. (2018), *P. tepidariorum* *Msx1-B* (previously designated *Msx3*) is expressed in the base of



**Fig. 7.** Expression of *P. tepidariorum* and *P. opilio* NK3 and NK4. In all panels, anterior is to the left. Panels a to c, f, g, i, k, m, and n show ventral views. Panels d and h show anterior-dorsal views. Panels e, and j show dorsal views. Panels l and o show lateral views and dorsal up. Panel p shows a posterior-dorsal view. Embryos in panels a, f, and k are flat-mounted. Panel b shows a magnification of the posterior region of a flat-mounted embryo. Panels a' to l', o', and p' show SYBR green stained embryos as shown in panels a to l, o, and p. Developmental stages of *P. tepidariorum* (after Mittmann and Wolff (2012)) and *P. opilio* (after Gainett et al. (2022)) are indicated in the upper right corner of each panel with a bright field image. Arrows in panels a, c to j, o, and p mark expression of NK3 and NK4 in the developing heart of *P. tepidariorum*, and expression of NK4 in the developing heart of *P. opilio*. The arrowhead in panel a points to expression in the labrum. The asterisk in panels a and d mark expression in the head. In panel b, expression in the mesoderm is marked (note absence of expression in the overlying ectoderm). Double arrowheads in panels f, g, and i point to dot-like expression of NK4 in the opisthosoma of the spider. Asterisks in panels f and h mark stripes of expression in the anterior of the head. Asterisks in panels k and l mark two domains of NK3 expression in the head of the harvestman. Arrows in panels k to m mark short segmental stripes of NK3 expression. The arrowhead in panel n marks expression in the labrum at late developmental stages. Abbreviations: ch, chelicera; ect, ectoderm; hl, head lobe; L1 to L4, first to fourth walking leg; lr, labrum; mes, mesoderm; O, opisthosomal segment; pp, pedipalp; s, stomodaeum.

the prosomal appendages at stage 11, while *Msx2-B* (previously designated *Msx2*) is expressed in the chelicerae at stage 12, and *Msx2-A* (previously designated *Msx1*) is expressed in each segment along the ventral midline in stages

8 and 10 (Akiyama-Oda and Oda 2020). In *P. opilio*, *Msx1* (previously designated *Msx*) was detected in a stripe in each segment at stage 7 and along the ventral midline in each segment at stage 13 (Leite et al. 2018). In contrast,



**Fig. 8.** Expression of *P. opilio* *Emx1* and *Msx2*. In panels a to e, anterior is to the left and ventral views (except panel c, lateral view and dorsal up). Panels f and g each show dissected head lobes and chelicerae, a walking leg and a pedipalp. The dissected heads are shown with anterior up, ventral views. The dissected legs and pedipalps present lateral views, dorsal to the right. Panels a'–e' show SYBR green stained embryos as shown in panels a–e. Developmental stages after Gainett et al. (2022) are indicated in the upper right corner of each panel with a bright field image. Arrowheads in panels a and b mark the sharp anterior border of *Emx1* expression anterior to the cheliceral segment. The arrows in panels a and c mark sharp distal border of expression in the appendages. Asterisks in panel d mark expression in the ventral nervous system, ventral to the base of the appendages. Double arrowheads in panels a, c, and e mark expression in the opisthosoma. Arrows in panels f and g point to expression at the base of the appendages. Arrowheads in panels f and g point to dot-like expression in the chelicerae. Asterisks in panels f and g mark a ring-like expression in the legs. Abbreviations: ch, chelicera; hl, head lobe; L1–L4, first to fourth walking leg; lr, labrum; pp, pedipalp.

we detected *P. opilio* *Msx2* expression only at late developmental stages and in a very restricted dot-like patch at the base of the appendages, and a ring-like domain in the legs (Fig. 8f and g). Thus, *P. opilio* *Msx1* expression in the segments and along the ventral midline is similar to *P. tepidariorum* *Msx2-A* expression, while *P. opilio* *Msx2* expression at the base of appendages is similar to *P. tepidariorum* *Msx1-B* expression. *P. opilio* *Msx2* expression was also detected in a small region of the chelicerae, matching the expression of *Msx2-B* in stage 12 *P. tepidariorum* embryos that was observed by Leite et al. (2018) and interpreted as a novel expression domain. Our description of a matching expression pattern from a previously unknown harvestman gene, *Msx2*, means that the inferred neofunctionalization is more likely a case of subfunctionalization between the tandem *Msx* paralogues.

### Ohnologues

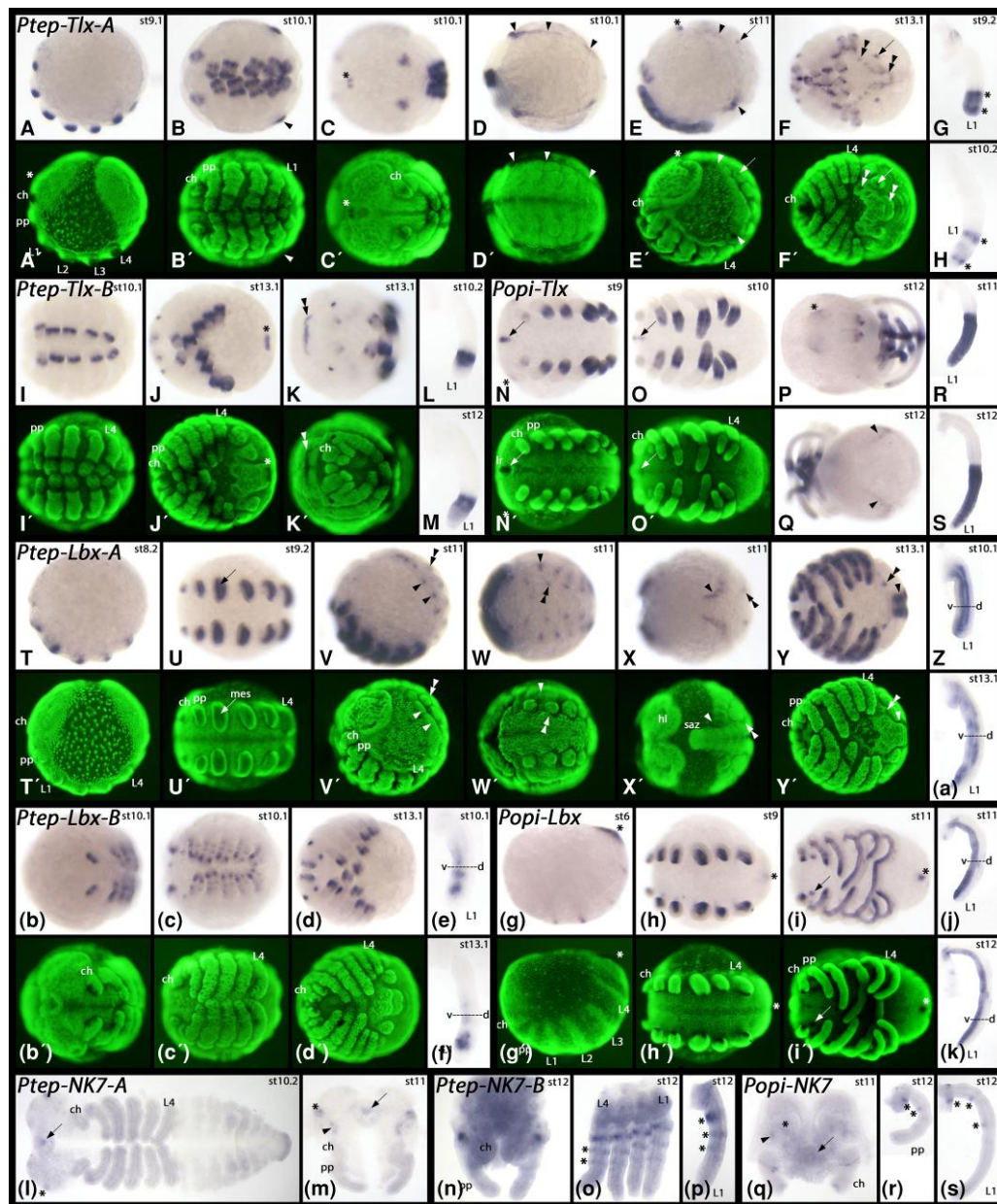
Three other NK families, NK7, *Tlx*, and *Lbx*, are found in the SC in *P. opilio* and have two retained ohnologues in *P. tepidariorum* (Fig. 4). The two ohnologues of *P. tepidariorum* *Tlx* are both expressed in the distal region of the legs, but in complementary patterns (Fig. 9A to M). *Tlx-A* is first expressed in the complete distal region of the appendages and restricted into two rings of stronger expression around stage 10 (Fig. 9G and H). *Tlx-B* expression only begins after this point and is strongest in the region between the two rings of *Tlx-A* expression (Fig. 9M). Apart from its

expression in the developing appendages, *Tlx-A* is also expressed in the dorsal tissue of the opisthosoma in a pattern that suggests a role in heart development (Fig. 9B, D, and E), and at later stages, in the ventral opisthosoma (Fig. 9F). In contrast, *Tlx-B* is only expressed in the opisthosoma at the very posterior at late developmental stages, after segment addition is complete (Fig. 9J). Unlike *Tlx-A*, *Tlx-B* is expressed in the form of a transverse stripe in the developing head at late developmental stages (double arrowheads in Fig. 9K) at the anterior edge of non-neurogenic ectoderm that overgrows the neurogenic ectoderm of the head lobe from the anterior.

The single *Tlx* gene of the harvestman shares many aspects with one or both spider ohnologues, indicating the spider *Tlx* ohnologues have subfunctionalized. *P. opilio* *Tlx* is expressed in the distal half of the developing appendages (Fig. 9R and S), reflecting the combined expression of both spider *Tlx* ohnologues (Fig. 9C and H, L and M). Like *P. tepidariorum* *Tlx-A*, *P. opilio* *Tlx* is expressed in dorsal tissue of the opisthosoma and in the labrum (Fig. 9N and O), and like *P. tepidariorum* *Tlx-B*, *P. opilio* *Tlx* is expressed at the anterior edges of the head lobes (asterisk in Fig. 9P).

Expression of *P. tepidariorum* *Lbx-A* first appears in the mesoderm of the outgrowing prosomal limb buds (Fig. 9T and U). Later, expression in the appendages is also seen in the ventral ectoderm, but the dorsal ectoderm remains free of expression (Fig. 9W and Y). Additional expression of *Lbx-A* is found in the dorsal ectoderm of the opisthosoma





**Fig. 9.** Expression of *P. tepidariorum* and *P. opilio* *Tlx*, *Lbx* and *NK7* ohnologues. In all panels showing whole embryos, anterior is to the left. Dissected heads in panels m, n, and q, are anterior up. Dissected appendages are shown by lateral views, dorsal to the right. Panels A, E, T, V, and g show lateral views and dorsal up. Panels B, C, D, F, I, J, N, O, U, W, Y, b, c, d, h, and i show ventral views. Panels K and P show anterior views. Panel Q shows a posterior view. Panel X shows a dorsal view. The embryo in panel I is flat-mounted. Panels A'-F, I-K', N', O', T'-Y', b'-d', and g'-i' show SYBR green stained embryos as shown in panels with corresponding bright field panels. Developmental stages of *P. tepidariorum* (after Mittmann and Wolff (2012)) and *P. opilio* (after Gainett et al. (2022)) are indicated in the upper right corner of each panel with a bright field image. Arrowheads in panels B, D, and E mark dorsal expression in the opisthosoma. The asterisks in panels C and E mark expression in the labrum. The arrows in panels E and F mark expression in the posterior spinneret, and the double arrowhead marks ventral expression in the opisthosoma. The asterisk in panel J marks expression of *Tlx-B* in the very posterior of the embryo. The double arrowhead in panel K marks a transverse stripe of expression in the head. The arrows in panels N and O mark expression of *Tlx* in the labrum of the harvestman. The asterisk in panel P marks expression in the head lobes. The arrow in panel U points to expression in the mesoderm of the leg. Arrowheads in panels V-Y point to dorsal expression in the opisthosoma and double arrowheads point to expression ventral to the base of the opisthosomal appendages. The asterisks in panels g to i mark expression in the very posterior of the embryos. The arrows in panels l, m, and q point to expression in the labrum. Asterisks and arrowheads in panels m and q mark expression of *NK7* in the head of the spider and the harvestman. Asterisks in panels o, p, r, and s mark expression in the appendages. Abbreviations: ch, chelicera; L1-L4, first to fourth walking leg; mes, mesoderm; pp, pedipalp; saz, segment addition zone.

and in segmental patches in the opisthosoma (Fig. 9V and Y). This latter expression is possibly associated with the ventral-most tissue of the opisthosomal limb buds. Compared to *Lbx-A*, expression of *P. tepidariorum Lbx-B* appears later and is restricted to the prosomal appendages (Fig. 9b to d). *Lbx-B* expression within the developing appendages is similar, but not identical, to that of *Lbx-A*; *Lbx-B* is expressed in fewer ventral cells than *Lbx-A* (Fig. 9e and f). Expression of the single *P. opilio Lbx* gene is similar to that of *P. tepidariorum Lbx-A*, appearing early during development in the mesoderm of the prosomal appendages (Fig. 9g and h), and later extending into ventral ectodermal cells of the limbs (Fig. 9j and k). Unlike the two *P. tepidariorum Lbx* ohnologues, *P. opilio Lbx* is also expressed at the posterior of the developing embryo (asterisks in Fig. 9h and i).

Expression of both ohnologues of *P. tepidariorum NK7* were only detected in relatively late stages of embryogenesis. *NK7-A* is exclusively expressed in the stomodaeum and in a few patches in the head lobes (Fig. 9l and m). *NK7-B* is expressed in three rings in the developing walking legs (Fig. 9n to p). The expression of *P. opilio NK7* combines the expression of both *P. tepidariorum NK7* genes as it is expressed in the head lobes and in three ventral domains in the legs (Fig. 9q to s), again suggesting the subfunctionalization of the spider *NK7* genes.

## Discussion

### Evolution of the Repertoire and Organization of Spider Homeobox Genes Before and After the WGD

We detected widespread conserved patterns of homeobox gene content and clusters across spider genomes that were impacted both by tandem duplications and the WGD. Leite et al. (2018) previously discovered that up to 59% of spider homeobox families were duplicated in arachnoplumonates and that most of these were dispersed duplicates rather than closely linked tandem duplicates. However, the few fragmented spider genomes available at the time meant that whether these were ohnologues or tandem duplicates could not reliably be determined. Here we showed that most duplicated homeobox genes are likely ohnologues retained after WGD, and the majority have been retained in all spiders surveyed (Fig. 2a, Table 2). The arachnoplumonate WGD has therefore played a prominent role in shaping the homeobox gene repertoires of spiders.

Duplicated homeobox gene clusters seem to have evolved asymmetrically following the WGD. Between the two ohnologous Hox clusters as well as the two NK, *lrx*, and SINE clusters, one cluster is more conserved in gene content, orientation, and older than the other. This asymmetry could be a result of mechanisms of regulatory subfunctionalization (e.g. complementary degenerating mutations to regulatory elements (Force et al. 1999)) acting not only at the level of individual genes but also on homeobox gene clusters.

Contextualizing the organization of the conserved duplicated homeobox gene clusters across entire spider genomes, we found that synteny is largely conserved among the

entelegynes (Fig. 6). We also observed a few rearrangements of homeobox clusters associated with chromosomal rearrangements (supplementary fig. S4B and C, Supplementary Material online). In *H. graminicola*, Hox cluster B and NK cluster B are both on chromosome 12, one of the two sex chromosomes, while in the other entelegynes, the synteny blocks that make up the *H. graminicola* sex chromosomes are found on two distinct chromosomes, one with Hox cluster B and the other with NK cluster B. This could have arisen from a rearrangement among the sex chromosomes in *H. graminicola* or independent rearrangements in theridiids and araneids. Given the  $X_1X_2$  male and  $X_1X_1X_2X_2$  female sex determination in most entelegynes including the species shown here (Maddison and Leduc-Robert 2013), differential rates of purifying selection on sex chromosomes versus autosomes may impact the extent to which the synteny of these chromosomes is conserved, thus affecting the homeobox clusters on them (Kořínková and Král 2013; Charlesworth 2017; Cordellier et al. 2020).

There seems to be a disconnect between the relative organization of spider Hox clusters and their regulation. Hox genes exhibit spatial and temporal collinearity (Ferrier 2019 and the references therein). The regulatory elements responsible for temporal collinearity constrain Hox gene organization, thus it has been hypothesized that Hox cluster integrity is essential for the mechanism of temporal collinearity (Duboule 2007; Krumlauf 2018; Ferrier 2019). While spatial collinearity is exhibited by both spider Hox clusters, temporal collinearity is more evident in cluster A, though this cluster is less organized (Schwager et al. 2017; Turetzek et al. 2022). Hox cluster organization and regulation vary between animals, for example, in mammals, each Hox cluster is condensed and is regulated globally, while in some invertebrates the Hox cluster is organized and expressed in subclusters (Kmita and Duboule 2003; Deschamps and Duboule 2017; Wang et al. 2017). This necessitates a better understanding of the relationship between the organization and regulation of the ohnologous spider Hox clusters.

The conservation of other homeobox clusters indicates that they may similarly be constrained by regulatory elements. ANTP-class Super-Hox genes arose with the Hox cluster, and many are still linked to one another and the Hox cluster across bilaterians (Butts et al. 2008; Ferrier 2016b). However, in spiders, the only consistent pattern among Super-Hox genes was the conserved clustering of *Abox* and *Meox*, though these two genes were not linked to either Hox cluster. This suggests that amid many rearrangements and gene losses, there may be a functional constraint on this two-gene cluster in spiders that would not have been recognized without comparisons across this taxonomic breadth.

Our comparison of NK cluster genes in spiders with other bilaterians revealed a pattern of conserved subclusters. In spiders, we found a core NK cluster consisting of a subset of the bilaterian NK complement largely conserved in gene order (Fig. 4). Its ohnologous cluster, NK cluster B, is less conserved but is linked to more of the bilaterian

NK complement (Fig. 4). Some NK gene organization patterns seen in spiders are similar to those in other arthropods. While the NK genes in different drosophilid species have undergone a series of rearrangements, fragmenting and then reuniting the NK cluster, certain gene pairs or triplets are consistently clustered and move as units referred to as “contiguities” (Chan et al. 2015). In spiders, *NK3* and *NK4* are adjacent, as in *D. melanogaster*, *T. castaneum*, *Bombyx terrestris*, many butterflies, and *Anopheles gambiae*, and *Tlx* and *Lbx* are adjacent, as in *D. melanogaster*, *T. castaneum*, several butterflies, *A. gambiae*, and *I. scapularis* (Fig. 4) (Luke et al. 2003; Chan et al. 2015; Mulhair et al. 2022). Though *NK1* has undergone a recent tandem duplication in spiders, *NK5* and *NK1* are also adjacent in *T. castaneum*, *B. terrestris*, *A. gambiae*, and *I. scapularis*, but not *D. melanogaster* or butterflies (Fig. 4) (Chan et al. 2015; Mulhair et al. 2022).

These NK subclusters are even conserved across other bilaterians (Luke et al. 2003; Larroux et al. 2007; Wotton et al. 2009; Hui et al. 2012; Mulhair et al. 2022). In the annelid *Platynereis dumerilii* and the invertebrate chordate amphioxus *Branchiostoma floridae*, *NK3* and *NK4* are clustered, as are *Lbx* and *Tlx* (Hui et al. 2012), while in human, one *NK4* ohnologue clusters with *Tlx* and *Lbx* ohnologues (Luke et al. 2003; Wotton et al. 2009). In contrast, the NK subclusters of tardigrades consist of different subgroups (Treffkorn et al. 2018), which most likely is a result of rearrangements in this lineage. Overall, these patterns suggest that while the entire NK cluster has undergone many rearrangements in different animal lineages, these events have repeatedly occurred between these subclusters, and far less often within them. This semiflexible organization distinguishes the NK cluster from the Hox cluster, which has a more consistent gene complement and order, but each of the NK subclusters may be maintained by a conserved regulatory mechanism similar to the Hox cluster.

### NK Cluster Genes Show Conserved and Subfunctionalized Expression

We observed that the genes in NK subclusters are expressed in similar patterns. NK cluster genes are broadly involved in patterning the mesoderm (Jagla et al. 2001; Saudemont et al. 2008; Treffkorn et al. 2018). The expression of *NK3* and *NK4* in both arachnids is similar to the *D. melanogaster* orthologues *tin* and *bap* in the heart and visceral mesoderm (Azpiazu and Frasch 1993). The fly *Tlx* orthologue, *C15*, and *Lbx* paralogues *lbe* and *lbi* are also expressed in legs like their orthologues in both *P. tepidarium* and *P. opilio* (Jagla et al. 2001; Campbell 2005; Maqbool et al. 2006). This suggests these genes share conserved cis-regulatory elements that may not only preserve their clustering but also result in their conserved co-expression across arthropods. It would be interesting to identify these regulatory elements and to investigate the extent of co-expression within NK subclusters more broadly across bilaterians.

Though a few examples of potential neofunctionalization have been described, subfunctionalization seems to

be the more common fate for duplicated homeobox genes (Schwager et al. 2017; Leite et al. 2018), and other genes (Schomburg et al. 2015; Janssen et al. 2021) in spiders. Using a more thorough annotation of NK genes from the harvestman, we were able to determine the relative functional consequences of tandem duplication versus the WGD, which was particularly important for families with ancient tandem duplication events like *Emx* and *Msx*. Expression of *Emx* tandem paralogues was similar while A and B ohnologues have undergone temporal subfunctionalization (Leite et al. 2018), despite the *Emx* tandem duplication now estimated to predate the WGD. Furthermore, the two *Emx* paralogues in *P. opilio* are expressed in near-identical patterns (Fig. 8) (Leite et al. 2018), which suggests that while the tandem duplication gave rise to two conserved paralogues, *Emx* genes did not subfunctionalize until after the WGD. For *Msx*, we found that *P. opilio* *Msx2* exhibits a similar expression pattern to *Msx2-B*, which now seems more likely to have undergone subfunctionalization instead of neofunctionalization. This stresses the importance of understanding the full evolutionary history of a gene family before inferring sub- and/or neofunctionalization. Furthermore, the roles of the ancestral single copies of *Msx* and *Emx* are not yet known but could be determined with comparisons to earlier-branching lineages of chelicerates or arachnids.

It is thought that duplicated genes encoding developmental transcription factors like homeobox genes are more likely to undergo regulatory subfunctionalization, driven by the complementary loss of enhancers (Force et al. 1999; Jiménez-Delgado et al. 2009; Espinosa-Cantú et al. 2015; Marlétaz et al. 2018). Indeed, the clear patterns of subfunctionalization observed for the *P. tepidarium* NK ohnologues and single-copy orthologues from *P. opilio* despite no NK cluster ohnologues being found to be under positive selection suggest changes primarily occurred in their regulatory regions.

### Comparisons to the Outcomes of Other WGD Events

Our results suggest that there have been similar outcomes between the arachnoplumonate WGD and the likely independent WGDs in horseshoe crabs (Nossa et al. 2014; Kenny et al. 2016; Shingate et al. 2020; Nong et al. 2021). Here, many homeobox gene families are expanded to four or more copies and there are multiple duplicated homeobox clusters, including up to six Hox clusters, six NK clusters, and seven SINE clusters as a result of up to three WGDs (Nong et al. 2021). Like spiders, in horseshoe crabs some ohnologous clusters are more intact than others; for instance, two of the six Hox clusters are intact and ordered, while the others have lost between two and eight genes (Shingate et al. 2020). However, in the absence of additional chromosome-level assemblies, the extent to which this applies to other horseshoe crabs is not yet clear.

In comparison to vertebrate homeobox gene clusters resulting from the 2R WGD, the ohnologous clusters of



spiders are more conserved in terms of gene content. Across the four human Hox clusters, complementary gene losses resulted in no cluster containing all gene families, but in spiders, one Hox cluster is intact while the other has lost only two genes (Fig. 3). The four ohnologous human NK clusters are spread across five chromosomes, and most genes are retained in two or maximum three copies (Luke et al. 2003; Wotton et al. 2009). While in spiders the HRO cluster is intact, *Hbn* is not found in chordates, and *Otp* is not linked to *Rax* or *Rax2* in the human genome (Mazza et al. 2010). However, for SINE genes, the clusters have evolved similarly between vertebrates and spiders following WGD (Kawakami et al. 2000; Ferrier 2016b). In both of these lineages, linkage of all three SINE genes is conserved on one chromosome, while the ohnologous cluster(s) have undergone losses and rearrangements. Therefore, the pattern of one copy being retained intact while the other has degenerated, as seen for most ohnologous spider homeobox gene clusters, is similar for only some human homeobox clusters. Still, these few similarities illustrate how duplication can lead to asymmetric degradation of otherwise conserved gene clusters.

Besides conserved clusters, some homeobox genes have returned to single copy in both humans and spiders or scorpions, namely, *Hlx*, *Hhex*, *Mnx*, *Dmbx*, *Otp*, *Prop*, and *Mkx* (Holland et al. 2007; Leite et al. 2018). Genes often found in single copy in both spiders and horseshoe crabs are *Bari*, *Barx*, CG11294, *Dmbx*, *Hhex*, *Meox*, *Mkx*, *Mxslx*, and *NK4* (Nong et al. 2021). This suggests that certain genes may be constitutive singletons, disadvantageous in multiple copies, or unable to subfunctionalize and thus lost. However, these comparisons are only between three distantly related species; wider sampling of WGD lineages may reveal clearer patterns of genes that tend to return to the single copy and the underlying reasons.

WGDs are hypothesized to have had important evolutionary consequences, primarily based on the finding that 2R WGD occurred along the vertebrate stem. However, the impact of WGDs in other animal lineages remains unclear. We still do not know to what extent the evolution of the morphological novelties and species diversity of spiders was driven by the arachnoplumonate WGD, though our work contributes to a growing body of evidence detailing widespread subfunctionalization of developmental ohnologues. The homeobox genes we have characterized here as well as other developmental genes provide candidates for future investigations, which, in concert with future studies of other chelicerate lineages, advance our aim to understand the impact of the arachnoplumonate WGD and WGDs more widely across animals.

## Conclusions

Analyses of WGDs in arachnoplumonates, horseshoe crabs, and vertebrates show that homeobox genes, as well as other regulatory genes are commonly retained in duplicate (Cañestro et al. 2013; Berthelot et al. 2014;

Leite et al. 2016; Harper et al. 2021; Janssen et al. 2021). Across eight spiders, representing three major clades we found that the majority of duplicated homeobox genes likely originated in the arachnoplumonate WGD and that often, cluster or subcluster integrity has been maintained in at least one ohnologous copy. Comparisons of expression patterns of selected homeobox genes support extensive subfunctionalization among spider ohnologues. Taken together, our results help us understand the impact of WGD and tandem duplication on the evolution of homeobox genes and genomes in spiders and provide an important comparison to WGD in other animals.

## Materials and Methods

### Homeobox Gene Annotation

We classified the homeobox gene complement of the publicly available chromosome-level genomes from eight spider species, *D. silvatica* (Sánchez-Herrero et al. 2019; Escuer et al. 2022), *D. plantarius* (Blaxter et al. 2022), *L. elegans* (Wang et al. 2022), *P. tepidariorum* (Zhu et al. 2023), *H. graminicola* (Zhu et al. 2022), *A. bruennichi* (Sheffer et al. 2021), *T. antipodiana* (Fan et al. 2021), and *T. clavata* (Hu et al. 2022). We examined the previous homeobox gene classification of the scorpion *C. sculpturatus*, which shared the ancestral WGD, as a outgroup to the spiders (Schwager et al. 2017; Leite et al. 2018). As non-WGD outgroups, we included two arachnids, the harvestman *P. opilio* (Gainett et al. 2021), and the tick *I. scapularis* (De et al. 2023), as well as three mandibulates: the centipede *S. maritima* (Chipman et al. 2014; So et al. 2022), the fruit fly *D. melanogaster* (BDGP6.32), in which many of these homeobox genes and families were first described (Zhong et al. 2008), and the beetle *T. castaneum* (Herndon et al. 2020), which was used by Butts et al. (2008) to describe the Super-Hox cluster. Only chromosome-level assemblies were used for synteny analyses, but homeobox genes were counted and classified in all genomes. Genome accessions are shown in Table 4.

We used tBLASTn searches (Altschul et al. 1990) with the contents of HomeoDB (accessed July 2020) (Zhong et al. 2008; Zhong and Holland 2011), as well as the previous classification of *P. tepidariorum* homeobox genes (Schwager et al. 2017; Leite et al. 2018) as the query set to retrieve homeodomain sequences from all eight spider genomes, and re-annotated the homeobox genes for the harvestman, tick, and centipede. Homeodomain sequences were then curated by tBLASTn searches against annotated gene models or transcriptomes for all species except *D. plantarius*, for which an annotation is not available, and classified based on sequence similarity and diagnostic residues of the homeodomain for all families. For some gene families with several duplication events, e.g. *Irx*, *Pax4/6*, *Pax3/7*, as well as for the NK cluster genes, phylogenies were constructed (see below). All homeodomain sequences can be found in the supplementary information (supplementary table S1, Supplementary Material online).

**Table 4** Genome assemblies used in this study and accessions. NCBI accession numbers or sequence database DOIs are given for genomes with homeobox genes annotated in this study for the first time. Citations are given for genomes with published homeobox gene annotations. Genome size, N50, and Benchmarking Universal Single-Copy Orthologues (BUSCO) scores are shown for each genome as calculated for the annotated chromosome-level assemblies with the arthropoda\_odb10 database using BUSCO v5.4.7 or were taken from NCBI and the genome papers (cited in the text above). Abbreviations: C: completeness; S: single copy; D: duplicated. The first-listed *A. bruennichi* genome, sequenced by the Darwin Tree of Life project, had a higher N50 and larger size, so was used for homeobox annotation, however it has not been annotated, so an older version, the second-listed, was used for the whole genome synteny comparison as it has a corresponding annotation. A homeobox annotation was also conducted on this older version (supplementary table S1, Supplementary Material online). The older *P. tepidariorum* assembly was not to chromosome-level so homeobox genes were re-annotated on the newer chromosome-level genome despite existing annotations on the previous version.

Species	Accession and Citation	Size	N50	BUSCO
<i>T. clavata</i>	<a href="https://spider.bioinfotoolkits.net/base/download/-1">https://spider.bioinfotoolkits.net/base/download/-1</a>	2.6 Gb	202.1 Mb	C:83.8% [S:80.8%, D:3.0%]
<i>T. antipodiana</i>	<a href="http://dx.doi.org/10.5524/100868">http://dx.doi.org/10.5524/100868</a>	2.3 Gb	172.9 Mb	C:96.3% [S:92.5%, D:3.8%]
<i>A. bruennichi</i> (DTOL)	GCF_947563725.1	1.8 Gb	139.1 Mb	unannotated
<i>A. bruennichi</i> (OLD)	GCA_015342795.1	1.7 Gb	124.2 Mb	C:87.0% [S:81.5%, D:5.5%]
<i>H. graminicola</i>	GCA_023701765.1	936.1 Mb	77.1 Mb	C:96.0% [S:92.2%, D:3.8%]
<i>P. tepidariorum</i> (OLD)	GCF_000365465.2; annotated by Schwager et al. 2017, Leite et al. 2018	1.4 Gb	4.1 Mb	98% completeness reported
<i>P. tepidariorum</i> (NEW)	<a href="https://doi.org/10.11922/sciencedb.o00019.00014">https://doi.org/10.11922/sciencedb.o00019.00014</a>	1.1 Gb	93.9 Mb	C:95.2% [S:92.4%, D:2.8%]
<i>L. elegans</i>	<a href="http://dx.doi.org/10.5524/102210">http://dx.doi.org/10.5524/102210</a>	1.6 Gb	114.3 Mb	C:58.3% [S:55.2%, D:3.1%]
<i>D. plantarius</i>	GCA_907164885.2	2.8 Gb	216.7 Mb	unannotated
<i>D. silvatica</i>	GCA_006491805.2	1.4 Gb	174.2 Mb	C:77.2% [S:74.7%, D:2.5%]
<i>C. sculpturatus</i>	GCA_000671375.2; annotated by Schwager et al. 2017, Leite et al. 2018	925.5 Mb	537.5 kb	96.8% completeness reported
<i>P. opilio</i>	GCA_019434445.1	576.9 Mb	211 kb	95.1% reported
<i>I. scapularis</i>	GCA_016920785.2; annotated by Schwager et al. 2017, Leite et al. 2018	2.2 Gb	132.1 Mb	C:98.8% [S:93.4%, D:5.4%]
<i>S. maritima</i>	GCA_000239455.1; annotated by Chipman et al. 2014, Leite et al. 2018	176.2 Mb	139.5 kb	93.8% completeness reported
<i>T. castaneum</i>	GCA_000002335.3; HomeoDB; Zhong et al. 2008, Zhong and Holland 2011	165.9 Mb	4.5 Mb	C:98.4% [S:97.7%, D:0.7%]
<i>D. melanogaster</i>	GCA_000001215.4; HomeoDB; Zhong et al. 2008, Zhong and Holland 2011	143.7 Mb	25.3 Mb	C:92.9% [S:92.3%, D:0.6%]

## Phylogenetics

To determine the timing of duplication of the arachnopulmonate *lrx* genes relative to preduplicate paralogues, we made a full-length amino acid sequence alignment from a subset of our spider species, the scorpion, and a transcriptome from another synspermiatan spider, *Segestria senoculata* (Gene Expression Omnibus accession GSE241073). Full sequences for genomes with poor annotations were curated from GenBank searches and BLASTp (Altschul et al. 1990) searches against nr (Benson et al. 2018). These sequences were aligned with MUSCLE set to the protein alignment defaults (Edgar 2004) and manually curated and trimmed to remove gaps. We inferred a Maximum-Likelihood phylogeny with IQ-TREE (Nguyen et al. 2015), determined the best substitution model with ModelFinder (Kalyaanamoorthy et al. 2017), and used 1,000 bootstrap replicates to optimize topology support. The phylogeny was visualized with FigTree (FigTree 2018). The full and trimmed alignments and the phylogeny can be found in the supplementary information on GitHub (supplementary fig. S1, Supplementary Material online, [https://github.com/madeleineaaseremedios/SpiderHomeoboxSequences]). Other family phylogenies were

made using only the homeodomain sequence and aligned manually, but inferred as for the *lrx* family tree, except for the NK cluster and NK-linked genes phylogeny, which used full protein sequences and was aligned with MUSCLE set to the protein default (Edgar 2004) (supplementary fig. S2, Supplementary Material online).

## Whole Genome Synteny Comparison

We performed whole genome synteny comparisons to explore chromosomal rearrangements that might underlie trends in homeobox placement in spider genomes. We used all spider species mentioned above, except for *D. plantarius* due to the lack of genome annotation, and compared these to nonspider species *I. scapularis*, *T. castaneum*, and *D. melanogaster*. Hierarchical orthogroups (HOGs) of all non-homeobox genes were identified using Orthofinder (Emms and Kelly 2019) with default settings, specifying the species tree: ((Dmel, Tcas)N1, (Isca, (Dsil, ((Ptep, Lele)N5, (Hgra, (Abru, (Tcla, Tant)N8)N7)N6)N4)N3)N2)N0). These were combined with HOGs of our homeobox orthology and used to reconstruct ancestral gene orders with Agora (Muffato et al. 2023), using the basic pipeline. However, we used Agora

to illustrate chromosomes with macrosynteny and orthology relationships, rather than the ancestral synteny blocks, termed Contiguous Ancestral Regions (CARs) detected by Agora, because it resolved a better signal of chromosomal rearrangements across the whole genome compared to the smaller coverage of CARs. We also used MCScanX (Wang et al. 2012) to capture microsynteny relationships between spider genomes and used Circos (Krzywinski et al. 2009) to plot synteny relationships and homeobox positions in the genome.

### Coding Sequence Evolutionary Analyses

For core NK cluster genes, we curated full coding sequences for a subset of species with the best annotations: *P. tepidariorum*, *A. bruennichi*, *D. silvatica*, *D. melanogaster*, *T. castaneum*, and *I. scapularis*, as well as four species not included in the synteny comparison: the centipede *S. maritima*, the harvestman *P. opilio*, the scorpion *C. sculpturatus*, and the onychophoran *Euperipatoides rowelli*, for which NK homeobox genes have been described, but which lack chromosomal-level genome assemblies (Chipman et al. 2014; Schwager et al. 2017; Leite et al. 2018; Treffkorn et al. 2018). The NK cluster genes contained families with different duplication histories, including WGD ohnologues in arachnospulmonates, tandem duplications both before and after the WGD, and families with only single genes represented after the WGD, allowing us to compare possible signatures of subfunctionalization between the different evolutionary trajectories and different types of duplicates.

We used aBSREL v2.3 (Smith et al. 2015) to detect selection on each branch of the phylogeny for each of the spider NK core cluster gene families (*Msx*, *NK3*, *NK4*, *Tlx*, *Lbx*, and *NK7*) and *Emx*. We used sequences from three spiders, *P. tepidariorum*, *D. silvatica*, and *A. bruennichi*, and the scorpion, harvestman, tick, fly, and beetle, as above, and rooted the trees with sequences from the onychophoran (Treffkorn et al. 2018). We created and trimmed codon alignments with PAL2NAL (Suyama et al. 2006) from coding sequences and protein sequence alignments inferred with MUSCLE set to the protein alignment default (Edgar 2004). We ran aBSREL using hyphy v2.5.46 and the outputs were visualized with HyPhy Vision at [<http://vision.hyphy.org/>] (Pond et al. 2005; Kosakovsky Pond et al. 2020). We also conducted pairwise comparisons of dN/dS ratios between *P. opilio* and *P. tepidariorum* for NK cluster genes calculated with codeML (Goldman and Yang 1994) from codon alignments inferred as above for each gene pair. Sequence data and aBSREL outputs can be found in the supplementary information on GitHub at [<https://github.com/madeleineaaserremedios/NKseqevoldata>] and visualized with HyPhy Vision as above.

### Animal Husbandry, Embryo Collection, Polymerase Chain Reaction, Riboprobe Synthesis, and In Situ Hybridization

Embryos of the spider *P. tepidariorum* were collected from the colony established in Uppsala, Sweden, and

maintained as described in Prpic et al. (2008). Embryos of the harvestman *P. opilio* were collected from wild-caught specimens in Uppsala, Sweden. Adult specimens and embryos of *P. opilio* were treated as described in Janssen et al. (2021). We applied the staging system of Mittmann and Wolff (2012) for *P. tepidariorum*, and Gainett et al. (2022) for *P. opilio*. We investigated all developmental stages of *P. tepidariorum* and *P. opilio* from the formation of the early germ band to dorsal closure. Total RNA from *P. tepidariorum* and *P. opilio* was isolated from a mix of embryonic stages using TRIzol (Invitrogen). mRNA was isolated from total RNA using the Dynabeads mRNA Purification System (Invitrogen) and reverse transcribed into cDNA using the SuperScriptII system (Invitrogen). Polymerase chain reactions were performed by applying sets of gene-specific primers. For all investigated genes, reverse primers were ordered with additional 5-prime T7-promotor sequences (gggTAATACGACTCACTATAG) for subsequent antisense RNA probe synthesis (David and Wedlich 2001). In situ hybridizations were performed as per Janssen et al. (2018). Primer sequences can be found in supplementary table S2, Supplementary Material online.

### Supplementary Material

Supplementary material is available at *Molecular Biology and Evolution* online.

### Acknowledgments

We would like to thank Julio Rozas and Paula Escuer for early access to the *D. silvatica* genome. This project was funded in by a NERC grant (NE/T006854/1) to APM and LSR.

### Author Contributions

A.P.M. and L.S.R. conceived of the project. M.E.A.R. annotated the homeobox genes, conducted the phylogenetics and protein sequence analyses, and drafted the manuscript. R.J. characterized the developmental expression of NK genes. D.J.L. conducted the whole genome macrosynteny analysis. All authors contributed to writing and approved the final version of the manuscript.

### Data Availability

All data underpinning this study is available through the links above or in the supplementary information.

### References

- Aase-Remedios ME, Ferrier DEK. Improved understanding of the role of gene and genome duplications in chordate evolution with new genome and transcriptome sequences. *Front Ecol Evol.* 2021;9:429. <https://doi.org/10.3389/fevo.2021.703163>.



- Abzhonov A, Popadic A, Kaufman TC. Chelicerate Hox genes and the homology of arthropod segments. *Evol Dev.* 1999;**1**(2):77–89. <https://doi.org/10.1046/j.1525-142x.1999.99014.x>.
- Akiyama-Oda Y, Oda H. Hedgehog signaling controls segmentation dynamics and diversity via *msx1* in a spider embryo. *Sci Adv.* 2020;**6**(37):eaba7261. <https://doi.org/10.1126/sciadv.aba7261>.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol.* 1990;**215**(3):403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2).
- Averof M, Patel NH. Crustacean appendage evolution associated with changes in Hox gene expression. *Nature.* 1997;**388**(6643):682–686. <https://doi.org/10.1038/41786>.
- Azpiazu N, Frasch M. Tinman and bagpipe: two homeo box genes that determine cell fates in the dorsal mesoderm of *Drosophila*. *Genes Dev.* 1993;**7**(7b):1325–1340. <https://doi.org/10.1101/gad.7.7b.1325>.
- Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Ostell J, Pruitt KD, Sayers EW. Genbank. *Nucleic Acids Res.* 2018;**46**(D1):D41–D47. <https://doi.org/10.1093/nar/gkx1094>.
- Berthelot C, Brunet F, Chalopin D, Juanchich A, Bernard M, Noël B, Bento P, Da Silva C, Labadie K, Alberti A, et al. The rainbow trout genome provides novel insights into evolution after whole-genome duplication in vertebrates. *Nat Commun.* 2014;**5**(1):3657. <https://doi.org/10.1038/ncomms4657>.
- Birkan M, Schaeper ND, Chipman AD. Early patterning and blastodermal fate map of the head in the milkweed bug *Oncopeltus fasciatus*. *Evol Dev.* 2011;**13**(5):436–447. <https://doi.org/10.1111/j.1525-142X.2011.00497.x>.
- Blaxter M, Mieszkowska N, Di Palma F, Holland P, Durbin R, Richards T, Berriman M, Kersey P, Hollingsworth P, Wilson W, et al. Sequence locally, think globally: the Darwin tree of life project. *Proc Natl Acad Sci.* 2022;**119**:e2115642118. <https://doi.org/10.1073/pnas.2115642118>.
- Brooke NM, Garcia-Fernández J, Holland PWH. The ParaHox gene cluster is an evolutionary sister of the Hox gene cluster. *Nature.* 1998;**392**(6679):920–922. <https://doi.org/10.1038/31933>.
- Brunet F, Lorin T, Bernard L, Haftek-Terreau Z, Galiana D, Scharl M, Volff JN. *Evolutionary biology: self/nonself evolution, species and complex traits evolution, methods and concepts*. Cham: Springer International Publishing; 2017. p. 369–396.
- Bürglin TR. Homeodomain subtypes and functional diversity. In: Hughes TR, editors. *A handbook of transcription factors. Subcellular biochemistry*. Dordrecht: Springer Netherlands; 2011. p. 95–122.
- Butts T, Holland PWH, Ferrier DEK. The urbilaterian super-Hox cluster. *Trends Genet.* 2008;**24**(6):259–262. <https://doi.org/10.1016/j.tig.2007.09.006>.
- Butts T, Holland PWH, Ferrier DEK. Ancient homeobox gene loss and the evolution of chordate brain and pharynx development: deductions from amphioxus gene expression. *Proc R Soc B Biol Sci.* 2010;**277**(1699):3381–3389. <https://doi.org/10.1098/rspb.2010.0647>.
- Campbell G. Regulation of gene expression in the distal region of the *Drosophila* leg by the Hox11 homolog, C15. *Dev Biol.* 2005;**278**(2):607–618. <https://doi.org/10.1016/j.ydbio.2004.12.009>.
- Cañestro C, Albalat R, Irimia M, Garcia-Fernández J. Impact of gene gains, losses and duplication modes on the origin and diversification of vertebrates. *Semin Cell Dev Biol.* 2013;**24**(2):83–94. <https://doi.org/10.1016/j.semcdb.2012.12.008>.
- Chan C, Jayasekera S, Kao B, Páramo M, von Grotthuss M, Ranz JM. Remodelling of a homeobox gene cluster by multiple independent gene reunions in *Drosophila*. *Nat Commun.* 2015;**6**(1):6509. <https://doi.org/10.1038/ncomms7509>.
- Charlesworth D. Evolution of recombination rates between sex chromosomes. *Philos Trans R Soc B Biol Sci.* 2017;**372**(1736):20160456. <https://doi.org/10.1098/rstb.2016.0456>.
- Chipman AD, Ferrier DEK, Brena C, Qu J, Hughes DST, Schröder R, Torres-Oliva M, Znassi N, Jiang H, Almeida FC, et al. The first myriapod genome sequence reveals conservative arthropod gene content and genome organisation in the centipede *Strigamia* *maritima*. *PLoS Biol.* 2014;**12**(11):e1002005. <https://doi.org/10.1371/journal.pbio.1002005>.
- Cordellier M, Schneider JM, Uhl G, Posnien N. Sex differences in spiders: from phenotype to genomics. *Dev Genes Evol.* 2020;**230**(2):155–172. <https://doi.org/10.1007/s00427-020-00657-6>.
- David R, Wedlich D. PCR-based RNA probes: a quick and sensitive method to improve whole mount embryo in situ hybridizations. *BioTechniques.* 2001;**30**(4):768–774. <https://doi.org/10.2144/01304st02>.
- De S, Kingan SB, Kitsou C, Portik DM, Foor SD, Frederick JC, Rana VS, Paulat NS, Ray DA, Wang Y, et al. A high-quality *Ixodes scapularis* genome advances tick science. *Nat Genet.* 2023;**55**(2):301–311. <https://doi.org/10.1038/s41588-022-01275-w>.
- Deschamps J, Duboule D. Embryonic timing, axial stem cells, chromatin dynamics, and the Hox clock. *Genes Dev.* 2017;**31**(14):1406–1416. <https://doi.org/10.1101/gad.303123.117>.
- Duboule D. The rise and fall of Hox gene clusters. *Development.* 2007;**134**(14):2549–2560. <https://doi.org/10.1242/dev.001065>.
- Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 2004;**32**(5):1792–1797. <https://doi.org/10.1093/nar/gkh340>.
- Emms DM, Kelly S. Orthofinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* 2019;**20**(1):238. <https://doi.org/10.1186/s13059-019-1832-y>.
- Escuer P, Pisarenco VA, Fernández-Ruiz AA, Vizueta J, Sánchez-Herrero JF, Arnedo MA, Sánchez-Gracia A, Rozas J. The chromosome-scale assembly of the Canary Islands endemic spider *Dysdera silvatica* (Arachnida, Araneae) sheds light on the origin and genome structure of chemoreceptor gene families in chelicerates. *Mol Ecol Resour.* 2022;**22**(1):375–390. <https://doi.org/10.1111/1755-0998.13471>.
- Espinosa-Cantú A, Ascencio D, Barona-Gómez F, De Luna A. Gene duplication and the evolution of moonlighting proteins. *Front Genet.* 2015;**6**:227. <https://doi.org/10.3389/fgene.2015.00227>.
- Fan Z, Yuan T, Liu P, Wang LY, Jin JF, Zhang F, Zhang ZS. A chromosome-level genome of the spider *Trichonephila antipodiana* reveals the genetic basis of its polyphagy and evidence of an ancient whole-genome duplication event. *GigaScience.* 2021;**10**(3):1–15. <https://doi.org/10.1093/GIGASCIENCE/GIAB016>.
- Fernández R, Edgecombe GD, Giribet G. Phylogenomics illuminates the backbone of the Myriapoda tree of life and reconciles morphological and molecular phylogenies. *Sci Rep.* 2018a;**8**(1):83. <https://doi.org/10.1038/s41598-017-18562-w>.
- Fernández R, Kallal RJ, Dimitrov D, Ballesteros JA, Arnedo MA, Giribet G, Hormiga G. Phylogenomics, diversification dynamics, and comparative transcriptomics across the spider tree of life. *Curr Biol.* 2018b;**28**(9):1489–1497.e5. <https://doi.org/10.1016/j.cub.2018.03.064>.
- Ferrier DEK. The origin of the Hox/ParaHox genes, the ghost locus hypothesis and the complexity of the first animal. *Brief Funct Genomics.* 2016a;**15**(5):333–341. <https://doi.org/10.1093/bfpg/elfv056>.
- Ferrier DEK. Evolution of homeobox gene clusters in animals: the giga-cluster and primary vs. secondary clustering. *Front Ecol Evol.* 2016b;**4**:1–13. <https://doi.org/10.3389/fevo.2016.00036>.
- Ferrier DEK. Space and time in Hox/ParaHox gene cluster evolution. In: Fusco G, editors. *Perspectives on evolutionary and developmental biology: essays for alessandro minelli*. Padova: Padova University Press; 2019. p. 245–274.
- FigTree. 2018. FigTree v1.4.4. <http://tree.bio.ed.ac.uk/software/figtree/>.
- Flot J-F, Hespels B, Li X, Noel B, Arkhipova I, Danchin EGJ, Hejnol A, Henrissat B, Koszul R, Aury J-M, et al. Genomic evidence for ameiotic evolution in the bdelloid rotifer *Adineta vaga*. *Nature.* 2013;**500**(7463):453–457. <https://doi.org/10.1038/nature12326>.
- Force A, Lynch M, Pickett FB, Amores A, Yan Y, Postlethwait J. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics.* 1999;**151**(4):1531–1545. <https://doi.org/10.1093/genetics/151.4.1531>.
- Gainett G, Crawford AR, Klementz BC, So C, Baker CM, Setton EVW, Sharma PP. Eggs to long-legs: embryonic staging of the harvestman

- Phalangium opilio (Opiliones), an emerging model arachnid. *Front Zool.* 2022;**19**(1):11. <https://doi.org/10.1186/s12983-022-00454-z>.
- Gainett G, González VL, Ballesteros JA, Setton EVW, Baker CM, Barolo Gargiulo L, Santibáñez-López CE, Coddington JA, Sharma PP. The genome of a daddy-long-legs (Opiliones) illuminates the evolution of arachnid appendages. *Proc R Soc B Biol Sci.* 2021;**288**(1956): 20211168. <https://doi.org/10.1098/rspb.2021.1168>.
- Gallardo ME, Lopez-Rios J, Fernaud-Espinosa I, Granadino B, Sanz R, Ramos C, Ayuso C, Seller MJ, Brunner HG, Bovolenta P, et al. Genomic cloning and characterization of the human homeobox gene SIX6 reveals a cluster of SIX genes in chromosome 14 and associates SIX6 hemizygosity with bilateral anophthalmia and pituitary anomalies. *Genomics.* 1999;**61**(1):82–91. <https://doi.org/10.1006/geno.1999.5916>.
- García-Fernández J. The genesis and evolution of homeobox gene clusters. *Nat Rev Genet.* 2005;**6**(12):881–892. <https://doi.org/10.1038/nrg1723>.
- Goldman N, Yang Z. A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol Biol Evol.* 1994;**11**(5): 725–736. <https://doi.org/10.1093/oxfordjournals.molbev.a040153>.
- Harper A, Baudouin Gonzalez L, Schöner A, Janssen R, Seiter M, Holzem M, Arif S, McGregor AP, Sumner-Rooney L. Widespread retention of ohnologs in key developmental gene families following whole-genome duplication in arachnospulmonates. *G3 (Bethesda).* 2021;**11**:jkab299. <https://doi.org/10.1093/g3journal/jkab299>.
- Herndon N, Shelton J, Gerischer L, Ioannidis P, Ninova M, Dönitz J, Waterhouse RM, Liang C, Damm C, Siemanski J, et al. Enhanced genome assembly and a new official gene set for *Tribolium castaneum*. *BMC Genomics.* 2020;**21**(1):47. <https://doi.org/10.1186/s12864-019-6394-6>.
- Hoegg S, Meyer A. Hox clusters as models for vertebrate genome evolution. *Trends Genet.* 2005;**21**(8):421–424. <https://doi.org/10.1016/j.tig.2005.06.004>.
- Holland PWH. Evolution of homeobox genes. *Wiley Interdiscip Rev Dev Biol.* 2013;**2**(1):31–45. <https://doi.org/10.1002/wdev.78>.
- Holland PWH. Did homeobox gene duplications contribute to the Cambrian explosion? *Zool Lett.* 2015;**1**(1):1. <https://doi.org/10.1186/s40851-014-0004-x>.
- Holland PWH, Booth HAF, Bruford EA. Classification and nomenclature of all human homeobox genes. *BMC Biol.* 2007;**5**(1):47. <https://doi.org/10.1186/1741-7007-5-47>.
- Holland PWH, García-Fernández J. Hoxgenes and chordate evolution. *Dev Biol.* 1996;**173**(2):382–395. <https://doi.org/10.1006/dbio.1996.0034>.
- Holland PWH, Marlétaz F, Maeso I, Dunwell TL, Paps J. New genes from old: asymmetric divergence of gene duplicates and the evolution of development. *Philos Trans R Soc B Biol Sci.* 2017;**372**(1713): 20150480. <https://doi.org/10.1098/rstb.2015.0480>.
- Holland LZ, Ocampo Daza D. A new look at an old question: when did the second whole genome duplication occur in vertebrate evolution? *Genome Biol.* 2018;**19**(1):209. <https://doi.org/10.1186/s13059-018-1592-0>.
- Hu W, Jia A, Ma S, Zhang G, Wei Z, Lu F, Luo Y, Zhang Z, Sun J, Yang T, et al. Molecular atlas reveals the tri-sectional spinning mechanism of spider dragline silk. *bioRxiv* 496984. <https://doi.org/10.1101/2022.06.21.496984>, 24 June 2022, preprint: not peer reviewed.
- Hui JHL, McDougall C, Monteiro AS, Holland PWH, Arendt D, Balavoine G, Ferrier DEK. Extensive chordate and annelid macro-synteny reveals ancestral homeobox gene organization. *Mol Biol Evol.* 2012;**29**(1):157–165. <https://doi.org/10.1093/molbev/msr175>.
- Irimia M, Maeso I, García-Fernández J. Convergent evolution of clustering of iroquois homeobox genes across metazoans. *Mol Biol Evol.* 2008;**25**(8):1521–1525. <https://doi.org/10.1093/molbev/msn109>.
- Irimia M, Tena JJ, Alexis MS, Fernández-Miñán A, Maeso I, Bogdanović O, de la Calle-Mustienes E, Roy SW, Gómez-Skarmeta JL, Fraser HB. Extensive conservation of ancient microsynteny across metazoans due to cis-regulatory constraints. *Genome Res.* 2012;**22**(12):2356–2367. <https://doi.org/10.1101/gr.139725.112>.
- Jagla K, Bellard M, Frasch M. A cluster of Drosophila homeobox genes involved in mesoderm differentiation programs. *BioEssays.* 2001;**23**(2):125–133. <https://doi.org/10.1002/1521-1878>.
- Janssen R. Comparative analysis of gene expression patterns in the arthropod labrum and the onychophoran frontal appendages, and its implications for the arthropod head problem. *EvoDevo.* 2017;**8**(1):1. <https://doi.org/10.1186/s13227-016-0064-4>.
- Janssen R, Andersson E, Betnér E, Bijl S, Fowler W, Höök L, Leyhr J, Mannelqvist A, Panara V, Smith K, et al. Embryonic expression patterns and phylogenetic analysis of panarthropod sox genes: insight into nervous system development, segmentation and gonadogenesis. *BMC Evol Biol.* 2018;**18**(1):88. <https://doi.org/10.1186/s12862-018-1196-z>.
- Janssen R, Damen WGM. Diverged and conserved aspects of heart formation in a spider. *Evol Dev.* 2008;**10**(2):155–165. <https://doi.org/10.1111/j.1525-142X.2008.00223.x>.
- Janssen R, Eriksson BJ, Tait NN, Budd GE. Onychophoran Hox genes and the evolution of arthropod Hox gene expression. *Front Zool.* 2014;**11**(1):22. <https://doi.org/10.1186/1742-9994-11-22>.
- Janssen R, Pechmann M. Expression of posterior Hox genes and opisthosomal appendage development in a mygalomorph spider. *Dev Genes Evol.* 2023. [Online ahead of print]. <https://doi.org/10.1007/s00427-023-00707-9>.
- Janssen R, Pechmann M, Turetzek N. A chelicerate Wnt gene expression atlas: novel insights into the complexity of arthropod Wnt-patterning. *EvoDevo.* 2021;**12**(1):12. <https://doi.org/10.1186/s13227-021-00182-1>.
- Jiménez-Delgado S, Pascual-Anaya J, García-Fernández J. Implications of duplicated cis-regulatory elements in the evolution of metazoans: the DDI model or how simplicity begets novelty. *Brief Funct Genomic Proteomic.* 2009;**8**(4):266–275. <https://doi.org/10.1093/bfpg/elp029>.
- Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermin LS. Modelfinder: fast model selection for accurate phylogenetic estimates. *Nat Methods.* 2017;**14**(6):587–589. <https://doi.org/10.1038/nmeth.4285>.
- Kawakami K, Sato S, Ozaki H, Ikeda K. Six family genes-structure and function as transcription factors and their roles in development. *BioEssays.* 2000;**22**(7):616–626. [https://doi.org/10.1002/1521-1878\(200007\)22:7<616::AID-BIES4>3.0.CO;2-R](https://doi.org/10.1002/1521-1878(200007)22:7<616::AID-BIES4>3.0.CO;2-R).
- Kenny NJ, Chan KW, Nong W, Qu Z, Maeso I, Yip HY, Chan TF, Kwan HS, Holland PWH, Chu KH, et al. Ancestral whole-genome duplication in the marine chelicerate horseshoe crabs. *Heredity (Edinb).* 2016;**116**(2):190–199. <https://doi.org/10.1038/hdy.2015.89>.
- Kerner P, Ikmi A, Coen D, Vervoort M. Evolutionary history of the iroquois/lrx genes in metazoans. *BMC Evol Biol.* 2009;**9**(1):74. <https://doi.org/10.1186/1471-2148-9-74>.
- Kmita M, Duboule D. Organizing axes in time and space; 25 years of colinear tinkering. *Science.* 2003;**301**(5631):331–333. <https://doi.org/10.1126/science.1085753>.
- Kořínková T, Král J. Karyotypes, sex chromosomes, and meiotic division in spiders. In: Nentwig W, editors. *Spider ecophysiology*. Berlin, Heidelberg: Springer; 2013. p. 159–171.
- Kosakovsky Pond SL, Poon AFY, Velazquez R, Weaver S, Hepler NL, Murrell B, Shank SD, Magalis BR, Bouvier D, Nekrutenko A, et al. Hyphy 2.5—a customizable platform for evolutionary hypothesis testing using phylogenies. *Mol Biol Evol.* 2020;**37**(1): 295–299. <https://doi.org/10.1093/molbev/msz197>.
- Krumlauf R. Hox genes, clusters and collinearity. *Int J Dev Biol.* 2018;**62**(11–12):659–663. <https://doi.org/10.1387/ijdb.180330rr>.
- Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. Circos: an information aesthetic for comparative genomics. *Genome Res.* 2009;**19**(9):1639–1645. <https://doi.org/10.1101/gr.092759.109>.
- Kumar JP. The sine oculis homeobox (SIX) family of transcription factors as regulators of development and disease. *Cell Mol Life Sci.* 2009;**66**(4):565–583. <https://doi.org/10.1007/s00018-008-8335-4>.
- Larroux C, Fahey B, Degnan SM, Adamski M, Rokhsar DS, Degnan BM. The NK homeobox gene cluster predates the origin of

- Hox genes. *Curr Biol.* 2007;**17**(8):706–710. <https://doi.org/10.1016/j.cub.2007.03.008>.
- Larroux C, Luke GN, Koopman P, Rokhsar DS, Shimeld SM, Degnan BM. Genesis and expansion of metazoan transcription factor gene classes. *Mol Biol Evol.* 2008;**25**(5):980–996. <https://doi.org/10.1093/molbev/msn047>.
- Leite DJ, Baudouin-Gonzalez L, Iwasaki-Yokozawa S, Lozano-Fernandez J, Turetzek N, Akiyama-Oda Y, Prpic N-M, Pisani D, Oda H, Sharma PP, et al. Homeobox gene duplication and divergence in arachnids. *Mol Biol Evol.* 2018;**35**(9):2240–2253. <https://doi.org/10.1093/molbev/msy125>.
- Leite DJ, Ninova M, Hilbrant M, Arif S, Griffiths-Jones S, Ronshaugen M, McGregor AP. Pervasive microRNA duplication in chelicerates: insights from the embryonic microRNA repertoire of the spider parasteatoda tepidariorum. *Genome Biol Evol.* 2016;**8**(7):2133–2144. <https://doi.org/10.1093/gbe/evw143>.
- Li M, Chen W-T, Zhang Q-L, Liu M, Xing C-W, Cao Y, Luo F-Z, Yuan M-L, et al. Mitochondrial phylogenomics provides insights into the phylogeny and evolution of spiders (Arthropoda: Araneae). *Zool Res.* 2022;**43**:566–584. <https://doi.org/10.24272/j.issn.2095-8137.2021.418>.
- Luke GN, Castro LFC, McLay K, Bird C, Coulson A, Holland PWH. Dispersal of NK homeobox gene clusters in amphioxus and humans. *Proc Natl Acad Sci.* 2003;**100**(9):5292–5295. <https://doi.org/10.1073/pnas.0836141100>.
- Maddison WP, Leduc-Robert G. Multiple origins of sex chromosome fusions correlated with chiasma localization in habronattus jumping spiders (Araneae: Salticidae). *Evolution.* 2013;**67**(8):2258–2272. <https://doi.org/10.1111/evo.12109>.
- Maqbool T, Soler C, Jagla T, Daczewska M, Lodha N, Palliyil S, VijayRaghavan K, Jagla K. Shaping leg muscles in Drosophila: role of ladybird, a conserved regulator of appendicular myogenesis. *PLoS One.* 2006;**1**(1):e122. <https://doi.org/10.1371/journal.pone.0000122>.
- Marlétaz F, Fribas PN, Maeso I, Tena JJ, Bogdanovic O, Perry M, Wyatt CDR, de la Calle-Mustienes E, Bertrand S, Burguera D, et al. Amphioxus functional genomics and the origins of vertebrate gene regulation. *Nature.* 2018;**564**(7734):64–70. <https://doi.org/10.1038/s41586-018-0734-6>.
- Martin A, Serano JM, Jarvis E, Bruce HS, Wang J, Ray S, Barker CA, O'Connell LC, Patel NH. CRISPR/Cas9 mutagenesis reveals Versatile roles of Hox genes in crustacean limb specification and evolution. *Curr Biol.* 2016;**26**(1):14–26. <https://doi.org/10.1016/j.cub.2015.11.021>.
- Mazza ME, Pang K, Reitzel AM, Martindale MQ, Finnerty JR. A conserved cluster of three PRD-class homeobox genes (homeobrain, rx and orthopedia) in the Cnidaria and Protostomia. *EvoDevo.* 2010;**1**(1):3. <https://doi.org/10.1186/2041-9139-1-3>.
- McLysaght A, Hokamp K, Wolfe KH. Extensive genomic duplication during early chordate evolution. *Nat Genet.* 2002;**31**(2):200–204. <https://doi.org/10.1038/ng884>.
- Meyer A, Scharl M. Gene and genome duplications in vertebrates: the one-to-four (-to-eight in fish) rule and the evolution of novel gene functions. *Curr Opin Cell Biol.* 1999;**11**(6):699–704. [https://doi.org/10.1016/S0955-0674\(99\)00039-3](https://doi.org/10.1016/S0955-0674(99)00039-3).
- Miller J, Zimin AV, Gordus A. Chromosome-level genome and the identification of sex chromosomes in Ulloborus diversus. *GigaScience.* 2023;**12**:giad002. <https://doi.org/10.1093/gigascience/giad002>.
- Mittmann B, Wolff C. Embryonic development and staging of the cobweb spider parasteatoda tepidariorum C. L. Koch, 1841 (syn.: achaearanea tepidariorum; araneomorphae; theridiidae). *Dev Genes Evol.* 2012;**222**(4):189–216. <https://doi.org/10.1007/s00427-012-0401-0>.
- Muffato M, Louis A, Nguyen NTT, Lucas J, Berthelot C, Roest Crollius H. Reconstruction of hundreds of reference ancestral genomes across the eukaryotic kingdom. *Nat Ecol Evol.* 2023;**7**(3):355–366. <https://doi.org/10.1038/s41559-022-01956-z>.
- Mulhair PO, Crowley L, Boyes DH, Harper A, Lewis OT, Holland PWH; Darwin Tree of Life Consortium. Diversity, duplication, and genomic organization of homeobox genes in lepidoptera. *Genome Res.* 2022;**33**(1):32–44. <https://doi.org/10.1101/gr.277118.122>.
- Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating Maximum-likelihood phylogenies. *Mol Biol Evol.* 2015;**32**(1):268–274. <https://doi.org/10.1093/molbev/msu300>.
- Nong W, Qu Z, Li Y, Barton-Owen T, Wong AYP, Yip HY, Lee HT, Narayana S, Baril T, Swale T, et al. Horseshoe crab genomes reveal the evolution of genes and microRNAs after three rounds of whole genome duplication. *Commun Biol.* 2021;**4**(1):83. <https://doi.org/10.1038/s42003-020-01637-2>.
- Nossa CW, Havlak P, Yue J-X, Lv J, Vincent KY, Brockmann HJ, Putnam NH. Joint assembly and genetic mapping of the Atlantic horseshoe crab genome reveals ancient whole genome duplication. *GigaScience.* 2014;**3**(1):9. <https://doi.org/10.1186/2047-217X-3-9>.
- Ohno S. *Evolution by gene duplication*. Berlin, Heidelberg: Springer Berlin Heidelberg; 1970.
- Ontano AZ, Gainett G, Aharon S, Ballesteros JA, Benavides LR, Corbett KF, Gavish-Regev E, Harvey MS, Monsma S, Santibáñez-López CE, et al. Taxonomic sampling and rare genomic changes overcome long-branch attraction in the phylogenetic placement of pseudoscorpions. *Mol Biol Evol.* 2021;**38**:2446–2467. <https://doi.org/10.1093/molbev/msab038>.
- Pollard SL, Holland PWH. Evidence for 14 homeobox gene clusters in human genome ancestry. *Curr Biol.* 2000;**10**(17):1059–1062. [https://doi.org/10.1016/S0960-9822\(00\)00676-X](https://doi.org/10.1016/S0960-9822(00)00676-X).
- Pond SLK, Frost SDW, Muse SV. Hyphy: hypothesis testing using phylogenies. *Bioinformatics.* 2005;**21**(5):676–679. <https://doi.org/10.1093/bioinformatics/bti079>.
- Prpic N-M, Schoppmeier M, Damen WGM. Collection and fixation of spider embryos. *Cold Spring Harb Protoc.* 2008;**2008**(10):pdb.prot5067. <https://doi.org/10.1101/pdb.prot5067>.
- Putnam NH, Butts T, Ferrier DEK, Furlong RF, Hellsten U, Kawashima T, Robinson-Rechavi M, Shoguchi E, Terry A, Yu J-K, et al. The amphioxus genome and the evolution of the chordate karyotype. *Nature.* 2008;**453**(7198):1064–1071. <https://doi.org/10.1038/nature06967>.
- Sánchez-Herrero JF, Frías-López C, Escuer P, Hinojosa-Alvarez S, Arnedo MA, Sánchez-Gracia A, Rozas J. The draft genome sequence of the spider Dysdera silvatica (Araneae, Dysderidae): a valuable resource for functional and evolutionary genomic studies in chelicerates. *GigaScience.* 2019;**8**(8):giz099. <https://doi.org/10.1093/gigascience/giz099>.
- Saudemont A, Dray N, Hudry B, Le Gouar M, Vervoort M, Balavoine G. Complementary striped expression patterns of NK homeobox genes during segment formation in the annelid platynereis. *Dev Biol.* 2008;**317**(2):430–443. <https://doi.org/10.1016/j.ydbio.2008.02.013>.
- Schinko JB, Kreuzer N, Offen N, Posnien N, Wimmer EA, Bucher G. Divergent functions of orthodenticle, empty spiracles and buttonhead in early head patterning of the beetle Tribolium castaneum (Coleoptera). *Dev Biol.* 2008;**317**(2):600–613. <https://doi.org/10.1016/j.ydbio.2008.03.005>.
- Schomburg C, Turetzek N, Schacht MI, Schneider J, Kirfel P, Prpic N-M, Posnien N. Molecular characterization and embryonic origin of the eyes in the common house spider Parasteatoda tepidariorum. *EvoDevo.* 2015;**6**(1):15. <https://doi.org/10.1186/s13227-015-0011-9>.
- Schwager EE, Schoppmeier M, Pechmann M, Damen WGM. Duplicated Hox genes in the spider Cupiennius salei. *Front Zool.* 2007;**4**(1):10. <https://doi.org/10.1186/1742-9994-4-10>.
- Schwager EE, Sharma PP, Clarke T, Leite DJ, Wierschin T, Pechmann M, Akiyama-Oda Y, Esposito L, Bechsgaard J, Bilde T, et al. The house spider genome reveals an ancient whole-genome duplication during arachnid evolution. *BMC Biol.* 2017;**15**(1):62. <https://doi.org/10.1186/s12915-017-0399-x>.
- Serano JM, Martin A, Liubicich DM, Jarvis E, Bruce HS, La K, Browne WE, Grimwood J, Patel NH. Comprehensive analysis of Hox gene expression in the amphipod crustacean Parhyale hawaiiensis. *Dev Biol.* 2016;**409**(1):297–309. <https://doi.org/10.1016/j.ydbio.2015.10.029>.
- Shao L, Li S. Early cretaceous greenhouse pumped higher taxa diversification in spiders. *Mol Phylogenet Evol.* 2018;**127**:146–155. <https://doi.org/10.1016/j.ympev.2018.05.026>.



- Sharma PP, Schwager EE, Extavour CG, Giribet G. Hox gene expression in the harvestman *Phalangium opilio* reveals divergent patterning of the chelicerate opisthosoma. *Evol Dev*. 2012;**14**(5): 450–463. <https://doi.org/10.1111/j.1525-142X.2012.00565.x>.
- Sheffer MM, Hoppe A, Krehenwinkel H, Uhl G, Kuss AW, Jensen L, Jensen C, Gillespie RG, Hoff KJ, Prost S. Chromosome-level reference genome of the European wasp spider *Argiope bruennichi*: a resource for studies on range expansion and evolutionary adaptation. *GigaScience*. 2021;**10**(1):1–12. <https://doi.org/10.1093/GIGASCIENCE/GIAA148>.
- Shimeld SM, Holland PWH. Vertebrate innovations. *Proc Natl Acad Sci U S A*. 2000;**97**(9):4449–4452. <https://doi.org/10.1073/pnas.97.9.4449>.
- Shingate P, Ravi V, Prasad A, Tay BH, Garg KM, Chattopadhyay B, Yap LM, Rheindt FE, Venkatesh B. Chromosome-level assembly of the horseshoe crab genome provides insights into its genome evolution. *Nat Commun*. 2020;**11**(1):1–13. <https://doi.org/10.1038/s41467-020-16180-1>.
- Simakov O, Kawashima T. Independent evolution of genomic characters during major metazoan transitions. *Dev Biol*. 2017;**427**(2): 179–192. <https://doi.org/10.1016/j.ydbio.2016.11.012>.
- Simakov O, Kawashima T, Marlétaz F, Jenkins J, Koyanagi R, Mitros T, Hisata K, Bredeson J, Shoguchi E, Gyoja F, et al. Hemichordate genomes and deuterostome origins. *Nature*. 2015;**527**(7579): 459–465. <https://doi.org/10.1038/nature16150>.
- Simakov O, Marlétaz F, Yue J, O'Connell B, Jenkins J, Brandt A, Calef R, Tung C, Huang T, Schmutz J, et al. Deeply conserved synteny resolves early events in vertebrate evolution. *Nat Ecol Evol*. 2020;**4**(6):820–830. <https://doi.org/10.1038/s41559-020-1156-z>.
- Simonnet F, Célérier M-L, Quéinnec E. Orthodenticle and empty spiracles genes are expressed in a segmental pattern in chelicerates. *Dev Genes Evol*. 2006;**216**(7-8):467–480. <https://doi.org/10.1007/s00427-006-0093-4>.
- Smith MD, Wertheim JO, Weaver S, Murrell B, Scheffler K, Kosakovsky Pond SL. Less is more: an adaptive branch-site random effects model for efficient detection of episodic diversifying selection. *Mol Biol Evol*. 2015;**32**(5):1342–1353. <https://doi.org/10.1093/molbev/msv022>.
- So WL, Nong W, Xie Y, Baril T, Ma H, Qu Z, Haimovitz J, Swale T, Gaitan-Espitia JD, Lau KF, et al. Myriapod genomes reveal ancestral horizontal gene transfer and hormonal gene loss in millipedes. *Nat Commun*. 2022;**13**(1):3010. <https://doi.org/10.1038/s41467-022-30690-0>.
- Suyama M, Torrents D, Bork P. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res*. 2006;**34**(Web Server):W609–W612. <https://doi.org/10.1093/nar/gkl315>.
- Tinti M, Johnson C, Toth R, Ferrier DEK, Mackintosh C. Evolution of signal multiplexing by 14-3-3-binding 2R-ohnologue protein families in the vertebrates. *Open Biol*. 2012;**2**(7):120103. <https://doi.org/10.1098/rsob.120103>.
- Treffkorn S, Kahnke L, Hering L, Mayer G. Expression of NK cluster genes in the onychophoran *Euperipatoides rowelli*: implications for the evolution of NK family genes in nephrozoans. *EvoDevo*. 2018;**9**(1):17. <https://doi.org/10.1186/s13227-018-0105-2>.
- Turetzek N, Pechmann M, Janssen R, Prpic N-M. Hox genes in spiders: their significance for development and evolution. *Semin Cell Dev Biol*. 2022;**152-153**:24–34. <https://doi.org/10.1016/j.semcdb.2022.11.014>.
- Walldorf U, Gehring WJ. Empty spiracles, a gap gene containing a homeobox involved in *Drosophila* head development. *EMBO J*. 1992;**11**(6):2247–2259. <https://doi.org/10.1002/j.1460-2075.1992.tb05284.x>.
- Wang Y, Tang H, DeBarry JD, Tan X, Li J, Wang X, Lee T, Jin H, Marler B, Guo H, et al. MCSscan: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res*. 2012;**40**(7):e49. <https://doi.org/10.1093/nar/gkr1293>.
- Wang S, Zhang J, Jiao W, Li J, Xun X, Sun Y, Guo X, Huan P, Dong B, Zhang L, et al. Scallop genome provides insights into evolution of bilaterian karyotype and development. *Nat Ecol Evol*. 2017;**1**(5): 1–12. <https://doi.org/10.1038/s41559-017-0120>.
- Wang Z, Zhu K, Li H, Gao L, Huang H, Ren Y, Xiang H. Chromosome-level genome assembly of the black widow spider *Latrodectus elegans* illuminates composition and evolution of venom and silk proteins. *GigaScience*. 2022;**11**:1–11. <https://doi.org/10.1093/GIGASCIENCE/GIAC049>.
- Wotton KR, Weierud FK, Juárez-Morales JL, Alvares LE, Dietrich S, Lewis KE. Conservation of gene linkage in dispersed vertebrate NK homeobox clusters. *Dev Genes Evol*. 2009;**219**(9-10): 481–496. <https://doi.org/10.1007/s00427-009-0311-y>.
- Zhong YF, Butts T, Holland PWH. HomeoDB: a database of homeobox gene diversity. *Evol Dev*. 2008;**10**(5):516–518. <https://doi.org/10.1111/j.1525-142X.2008.00266.x>.
- Zhong YF, Holland PWH. HomeoDB2: functional expansion of a comparative homeobox gene database for evolutionary developmental biology. *Evol Dev*. 2011;**13**(6):567–568. <https://doi.org/10.1111/j.1525-142X.2011.00513.x>.
- Zhu B, Jin P, Hou Z, Li J, Wei S, Li S. Chromosomal-level genome of a sheet-web spider provides insight into the composition and evolution of venom. *Mol Ecol Resour*. 2022;**22**(6):2333–2348. <https://doi.org/10.1111/1755-0998.13601>.
- Zhu B, Jin P, Zhang Y, Shen Y, Wang W, Li S. Genomic and transcriptomic analyses support a silk gland origin of spider venom glands. *BMC Biol*. 2023;**21**(1):82. <https://doi.org/10.1186/s12915-023-01581-7>.