Evolution of the spider homeobox gene repertoire by tandem and whole genome duplication

Madeleine E. Aase-Remedios[1], Ralf Janssen[2]*, Daniel J. Leite[1], Lauren Sumner-Rooney[3], and Alistair P. McGregor[1]*

[1] Department of Biosciences, Durham University, Durham, DH1 3LE, UK.
[2] Department of Earth Sciences, Uppsala University, 752 36, Uppsala, Sweden.
[3] Museum für Naturkunde, Leibniz-Institut für Evolutions- und Biodiversitätsforschung, 10115 Berlin Germany.

*Corresponding authors: alistair.mcgregor@durham.ac.uk, ralf.janssen@geo.uu.se

ABSTRACT

Gene duplication generates new genetic material that can contribute to the evolution of gene regulatory networks and phenotypes. Duplicated genes can undergo subfunctionalisation to partition ancestral functions and/or neofunctionalisation to assume a new function. We previously found there had been a whole genome duplication (WGD) in an ancestor of arachnopulmonates, 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 key to determining the contribution of the WGD to arachnopulmonate genome evolution. Here we characterised 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, *Irx*, and SINE clusters, in all eight species. Consistently, we observed one copy of each cluster was degenerated in terms of gene content and organisation while the other remained more intact. Focussing on the NK cluster, we found evidence for regulatory subfunctionalisation 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.

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 subfunctionalise, specialise, or neofunctionalise 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 WGD (2R WGD) in the gnathostome ancestor has 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 recognised 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 (Garcia-Fernàndez 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; Larroux et al. 2008; Ferrier 2016a). Homeobox gene clusters (Table 1) are thought to be maintained by functional constraint on their genomic organisation

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

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 *Msxlx*, as well as two non-homeobox genes, *Pax1/9* and *FoxA1/2* (Simakov and Kawashima 2017); linkage and sometimes clustering between *NK2.1*, *NK2.2*, and *Msxlx* 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 *Irx* cluster likely underwent independent duplications in different lineages; *ara* and *caup* are a fly-specific duplication of one *Irx* family, often clustered with single arthropod *mirr* orthologues of a second *Irx* family (Irimia et al. 2008; Kerner et al. 2009). The dash for *Irx3* 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
Super-Hox (ANTP)	*Hox1*	*lab*	*lab*
	Hox2	*pb*	*pb*
	Hox3	*Hox3*	*zen, zen2, bcd*
	Hox4	*Dfd*	*Dfd*
	Hox5	*Scr*	*Scr*
	Hox6-8	*ftz*	*ftz*
		Antp	*Antp*
		Ubx	*Ubx*
		abd-A	*abd-A*
	Hox9-15	*Abd-B*	*Abd-B*
	Abox	*Abox*	*CG34031*
	Dbx	*Dbx*	*Dbx*
	Dlx	*Dlx*	*Dll*
	En	*En*	*en, inv*
	Evx	*Evx*	*eve*
	Gbx	*Gbx*	*unpg*
	Meox	*Meox*	*btn*
	Mnx	*Mnx*	*exex*
	Nedx	*Nedx*	*lms*
	Ro	*Ro*	*ro*
ParaHox (ANTP)	*Xlox/Pdx*	–	–
	Gsx	*Gsx*	*ind*
	Cdx	*Cdx*	*cad*

Cluster (Class)	Bilaterian family	Arthropod family	Synonyms
NK (ANTP)	*NK1*	*NK1*	*slou*
	NK3	*NK3*	*bap*
	NK4	*NK4*	*tin*
	NK5	*NK5*	*Hmx*
	NK6	*NK6*	*Hgtx*
	NK7	*NK7*	*NK7.1*
	Msx	*Msx*	*Dr, msh*
	Tlx	*Tlx*	*C15, cll*
	Lbx	*Lbx*	*lbe, lbl*
	Hhex	*Hhex*	*HHEX*
	Noto	*Noto*	*Not*
	Emx	*Emx*	*ems, E5*
NK2 (ANTP)	*NK2.1*	*NK2.1*	*scro*
	NK2.2	*NK2.2*	*vnd*
	Msxlx	*Msxlx*	*CG15696*
	Hlx	*Hlx*	*H2.0*
IRX (TALE)	*Irx*	*Irx1*	*ara, caup*
		Irx2	*mirr*
		Irx3	–
HRO+Isl (PRD + LIM)	*Hbn*	*Hbn*	*hbn*
	Rax	*Rax*	*Rx*
	Otp	*Otp*	*otp*
	Isl	*Isl*	*tup*
SINE (SINE)	*Six1/2*	*Six1/2*	*so*
	Six3/6	*Six3/6*	*optix*
	Six4/5	*Six4/5*	*Six4*

83 The best known homeobox gene cluster, the Hox cluster, illustrates how the genomic
84 organisation of homeobox genes is important for their function. Most arthropod
85 genomes contain one Hox cluster consisting of ten ANTP-class genes (Table 1), which
86 are involved in patterning the antero-posterior axis of animals in the same relative order
87 as the genomic organisation of the genes within the cluster (Krumlauf 2018 and the
88 references therein). The Hox cluster is highly conserved across bilaterians, both in
89 genomic organisation and function, though differences in Hox gene expression are
90 known to underlie changes to animal body plans (Averof and Patel 1997; Abzhanov et
91 al. 1999; Janssen et al. 2014; Martin et al. 2016; Serano et al. 2016; Janssen and
92 Pechmann 2023). Hox clusters have also been retained in duplicate following WGDs, as
93 exhibited by the four vertebrate Hox clusters that arose via the 2R WGD (Holland and
94 Garcia-Fernàndez 1996; McLysaght et al. 2002; Holland et al. 2007; Putnam et al.
95 2008; Cañestro et al. 2013; Holland 2013; Holland and Ocampo Daza 2018). The 2R
96 WGD appears to have relaxed constraint on vertebrate Hox cluster organisation so that
97 while the complement of genes from *Hox1* to *Hox14* are present in at least one copy, no
98 cluster contains all 14 genes (Hoegg and Meyer 2005). It is still not well understood to
99 what extent the patterns of gene loss between duplicate clusters and the
100 subfunctionalisation observed for vertebrate homeobox gene clusters apply more widely
101 to other homeobox gene clusters (Table 1) and to other animal lineages that have
102 undergone WGD.

103 Besides gnathostomes, an independent WGD took place in the arachnopulmonate
104 ancestor approximately 450 MYA (Figure 1), resulting in nearly 60% of homeobox
105 genes retained in duplicate in spiders and scorpions and their relatives, including two
106 Hox clusters (Schwager et al. 2017; Leite et al. 2018; Harper et al. 2021). Within each
107 Hox gene family, one ohnologue exhibits different spatial and temporal expression
108 compared to the other, though overall, spatial and temporal collinearity is largely intact
109 for each cluster (Schwager et al. 2007; Schwager et al. 2017; Turetzek et al. 2022).
110 Several of the other retained homeobox genes exhibit divergent expression consistent
111 with both sub- and neofunctionalisation 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 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 ohnologous clusters, or if they resulted from tandem

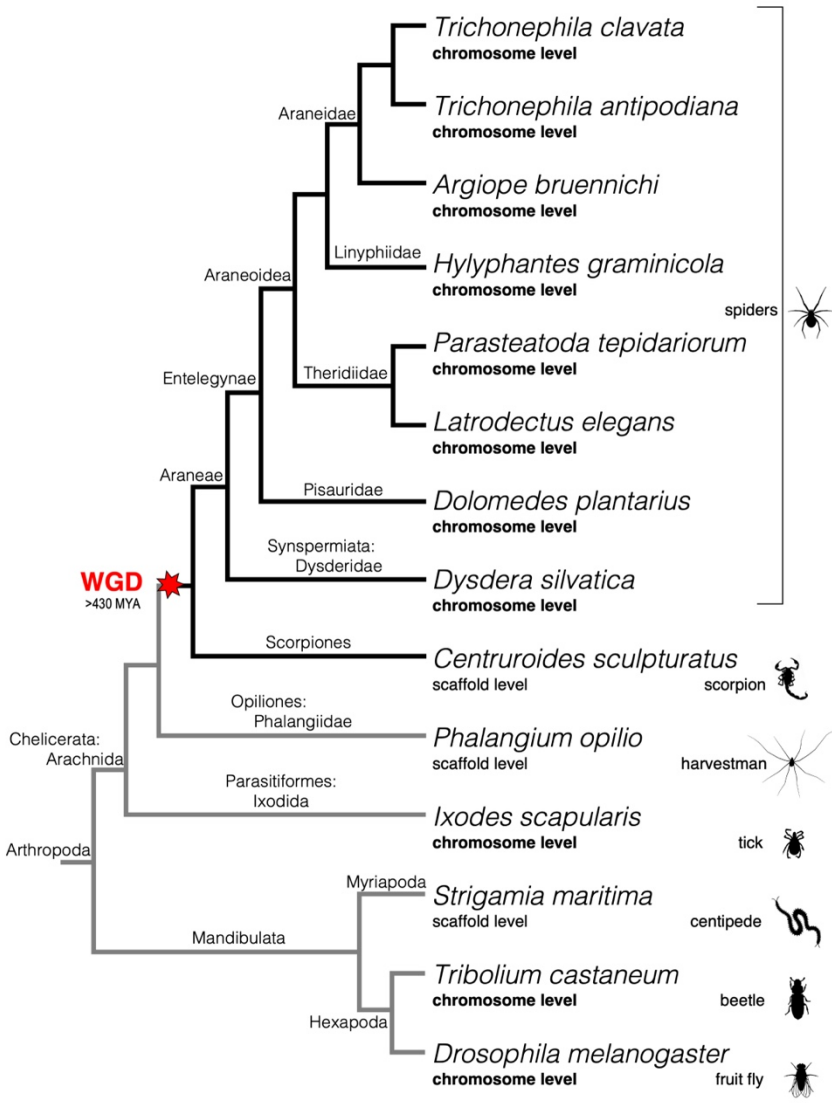


Figure 1 – Cladogram of arthropods used in this study. The red star indicates the timing of the whole genome duplication (WGD); pre-duplicate lineages have grey lines, duplicate lineages have black lines. Genome assembly level is indicated below the species binomial, in bold for the chromosome-level assemblies. Higher phylogenetic classifications are labelled 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/].

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* (Figure 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 characterised 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 *P. opilio*, an outgroup to the arachnopulmonates 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 *Phalangium 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 embryogenesis, providing evidence for regulatory subfunctionalisation. Combined, these results provide new insight into both the patterns of ohnologue retention and their consequences in spider evolution.

166 **RESULTS**

167 *Homeobox gene repertoires in arthropods*

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

185 In total, 90 described homeobox families were represented in our arthropod species set
186 (Table 2), though some were absent in different species, such that no species had a
187 gene from every family (Figure 2 A, Table S1). In addition, we also found several new
188 homeobox genes. A few of these may represent new lineage-specific unnamed ("un")
189 families, temporarily called *ANTP-un1*, *PRD-un1*, *SINE-un1*, and *TALE-un1* based on
190 the classes to which they likely belong. *ANTP-un1* is found only in the chelicerates,
191 while *PRD-un1* and *SINE-un1* were found only in the araneids, and *TALE-un1* only in
192 the *Trichonephila* spp. (Figure 2 A, Table S1). The '1' is simply a placeholder to indicate
193 that these genes have an orthologue in at least two species and to distinguish them

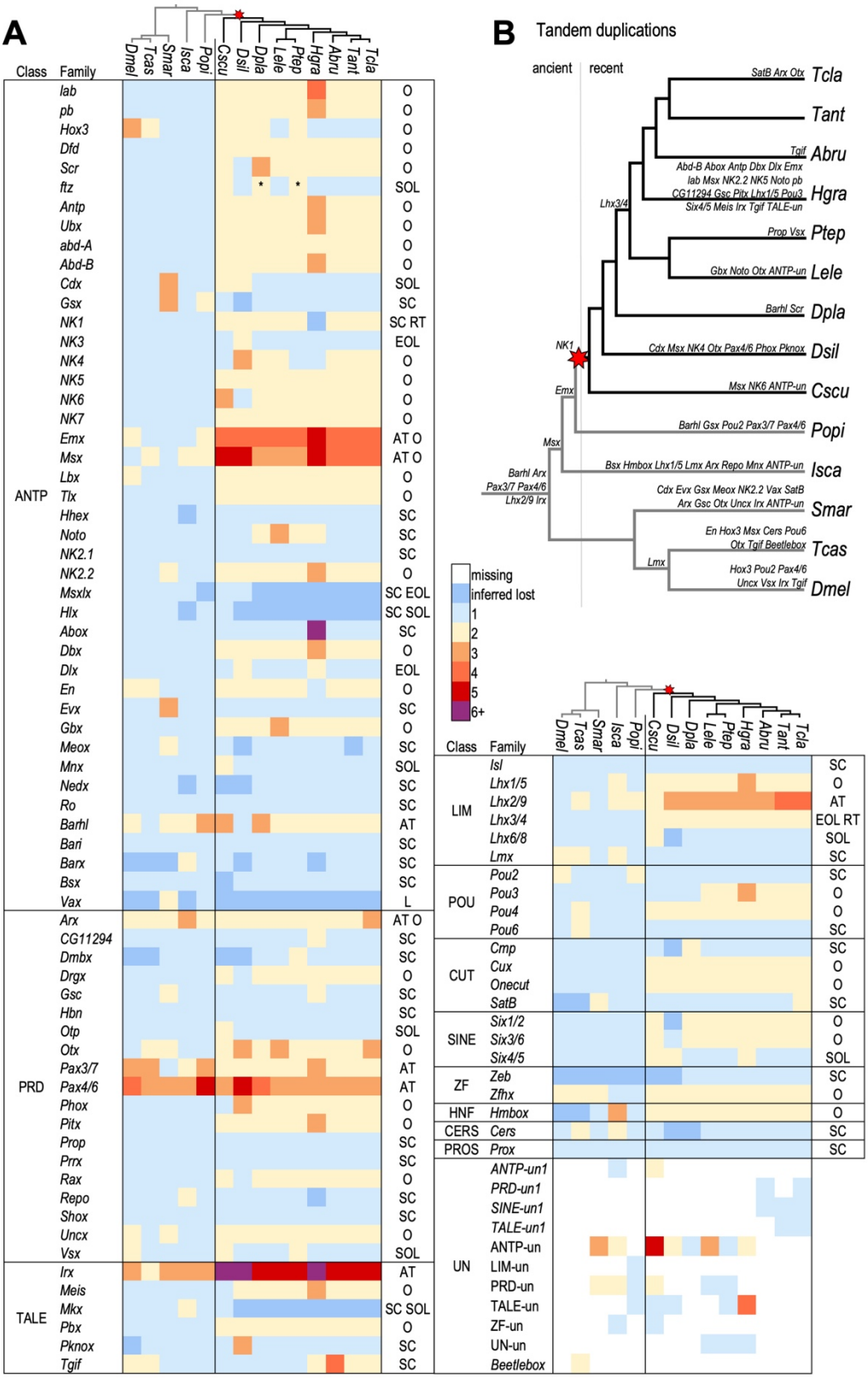
194 from other unnamed or unclassified homeodomain sequences that have no similarity to
195 described families and were found only in one species, sometimes in several copies.

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

	Dmel	*Tcas*	*Smar*	*Isca*	*Popi*	*Cscu*	*Dsil*	*Dpla*	*Lele*	*Ptep*	*Hgra*	*Abru*	*Tant*	*Tcla*
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

205

206 Comparing this taxonomic breadth of species in concert with genomic synteny enabled
207 us to distinguish ancient and recent tandem duplicates from likely ohnologues and
208 determine the timing of duplications relative to major radiations in arthropod evolution.
209 We have defined ancient tandem duplicates as paralogues that are often linked or
210 clustered in most spider genomes and date to a duplication node in an ancestor before
211 the arachnopulmonate WGD, while recent tandem duplicates are also linked or
212 clustered, but have occurred either after the WGD or in the outgroups. Six homeobox
213 gene families, *Arx*, *Barhl*, *Irx*, *Pax3/7*, *Pax4/6*, and *Lhx2/9*, underwent ancient tandem
214 duplication events that are shared by other arthropods and may have ancestry even
215 deeper in the bilaterian phylogeny (Figure 2 B). Only three of these ancient tandem
216 paralogues have been subsequently lost; just one *Lhx2/9* gene is found in *D.*
217 *melanogaster*, and only one *Barhl* gene is found in *T. castaneum* and *D. silvatica*.
218 Within the chelicerates, but predating the arachnopulmonate WGD, two further tandem
219 duplications occurred. *Msx* likely duplicated in the common ancestor of ticks,
220 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 (Figure 2 B). In particular, *H. graminicola* has the highest number of gene families with recent tandem duplications (23 families) (Figure 2 B), resulting in this species having the highest number of homeobox genes identified (172). This

Figure 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.

247 could be an assembly artefact; however, overall the *H. graminicola* genome did not
248 exhibit an exceptionally large number of duplicated genes (Zhu et al. 2022).

249 In contrast to tandem duplicates, dispersed paralogues consistently found on different
250 chromosomes in surveyed spider genomes were inferred to be ohnologues. There were
251 between 37 and 46 families with retained ohnologues in arachnopulmonates, and the
252 families that were retained were largely consistent across species. All the Hox genes
253 besides *Hox3* and *ftz* were retained as ohnologues, as previously reported for two
254 spider species (*P. tepidariorum* and *Cupiennius salei*) (Schwager et al. 2017), as well as
255 a few Super-Hox genes, namely *Dbx*, *En*, and *Gbx* (Figure 2 A). Many of the NK cluster
256 genes were also retained as ohnologues, including *NK4*, *NK5*, *NK6*, *NK7*, *Lbx*, *Tlx*, as
257 well as the ohnologues of ancient tandem paralogues of *Emx* and *Msx* (Figure 2 A).

258 Ohnologues were likely lost from several families following the WGD (Figure 2 A).
259 Thirty-two families are found in single copies in scorpions and spiders (discounting
260 subsequent lineage-specific tandem duplications of single-copy genes), indicating that
261 an ohnologue was lost between the WGD and the divergence of scorpions and spiders.
262 These include many of the Super-Hox families (*Abox*, *Dlx*, *Evx*, *Meox*, *Mnx*, *Nedx*, and
263 *Ro*), the ParaHox family *Gsx*, most of the NK2 cluster families (*Hlx*, *Msxlx*, and *NK2.1*)
264 and the HRO cluster family *Hbn*. Six families, including the Hox family *ftz*, the ParaHox
265 family *Cdx* and the HRO cluster family *Otp*, were found in single copies in spiders, while
266 both ohnologues were retained in the scorpion, indicating the loss of an ohnologue in a
267 spider ancestor (Figure 2 A). Within spiders, an ohnologue was lost from four families in
268 the entelegyne ancestor: *Dlx*, *Six4/5*, *Msxlx*, and *NK3*. *D. silvatica*, the sole
269 synspermiatan representative, has lost the largest number of ohnologues from eighteen
270 families, though the extent to which this applies across synspermiatans is not known.
271 Only a few families underwent losses multiple times, namely *Hox3* in *C. sculpturatus*, *D.*
272 *plantarius*, *Latrodectus elegans*, and the araneids; *Tgif* in the theridiids, *Dolomedes*
273 *plantarius*, *D. silvatica* and *C. sculpturatus*; and *Barx* in both *H. graminicola* and *D.*
274 *silvatica*. Only one family, *Vax*, was potentially lost before the WGD, or both ohnologues
275 were lost in the arachnopulmonate ancestor (Figure 2 A). Though it was not reported

276 previously (Leite et al. 2018), we identified a single *Vax* gene in the harvestman
277 genome (Figure 2 A, Table S1).

278 *Homeobox gene clusters in spider genomes*

279 The Hox cluster

280 Each of the spider genomes studied here contains two Hox clusters (Figure 3),
281 consistent with previous descriptions for *P. tepidariorum* and other spiders (Schwager et
282 al. 2007; Schwager et al. 2017; Fan et al. 2021; Sheffer et al. 2021; Wang et al. 2022).
283 With the new *P. tepidariorum* chromosome-level assembly, we confirmed that each
284 cluster is found on a different chromosome. Cluster B is typically intact, organised, and
285 contains orthologues of all ten arthropod Hox genes, while cluster A often lacks *Hox3*
286 and *ftz*, and is sometimes rearranged (Figure 3). There are some exceptions to this,
287 however. For one, there have been several tandem duplications of *H. graminicola* Hox
288 genes resulting in four *lab* paralogues, and three *pb*, *Antp*, *Ubx*, and *Abd-B* paralogues
289 (Figure 3). While these duplications have occurred in both *H. graminicola* Hox clusters,
290 cluster B is ordered, and contains the single orthologues of *Hox3* and *ftz*, while cluster A
291 lacks those genes and has been rearranged (Figure 3). Furthermore, the Super-Hox
292 gene *Dlx-A* is between *pb-A-2* and *Dfd-A* separated by 124 kb from *pb-A-2* and 380 kb
293 from *Dlx-A* and *Dfd-A*. In *D. plantarius* there is a new gene, identified with highest
294 similarity in the homeodomain to *Scr*, but not linked to either Hox cluster, possibly
295 representing a divergent *Scr* paralogue (blue *'S'* in Figure 3). In *D. silvatica*, both Hox
296 clusters are rearranged, though one does contain both *Hox3* and *ftz*, while the other has
297 lost only *ftz* (Figure 3). Pseudogenes of *ftz* were found in *D. plantarius* and *P.*
298 *tepidariorum*, identified by the homeodomain disrupted by a stop codon. In *P.*
299 *tepidariorum*, the *ftz*-pseudogene is not linked to either Hox cluster, while in *D.*
300 *plantarius,* it is adjacent to *ftz*-B (Table S1, Figure S3 F).

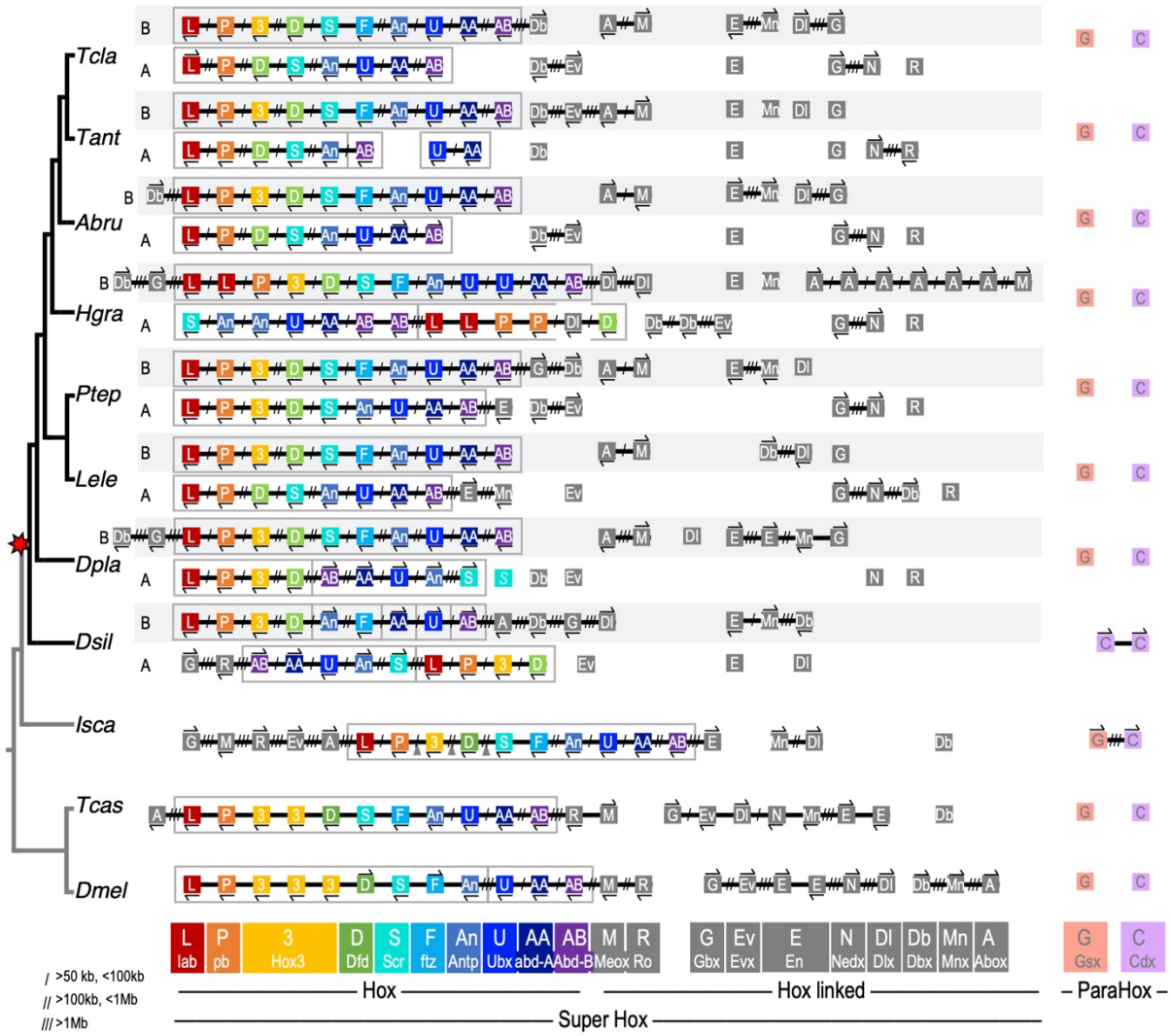


Figure 3: Two ohnologous Hox clusters and a disintegrated arthropod ParaHox cluster are found in spiders. Genes are represented by coloured boxes; chromosomes are black lines. Gene orientation is indicated by arrows above or below genes, and distances are indicated by slashes between genes. Gene families are denoted by a one- or two-letter code, and by colour for Hox and ParaHox genes, or in grey for Hox-linked genes. ParaHox genes are coloured 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.

301 The Super-Hox genes consist of the Hox cluster as well as several ANTP-class genes
302 consistently found linked to each other and/or the Hox cluster across bilaterians (Table
303 1) (Butts et al. 2008; Ferrier 2016b). While many Super-Hox genes are still linked to one
304 another and/or to the Hox cluster in the tick and insects, Super-Hox genes have
305 dispersed in spider genomes and many Super-Hox families have also returned to the
306 single copy (Figure 2 A, Figure 3). Amid the overall disintegration of the Super-Hox

307 cluster, *Abox* and *Meox* are consistently found clustered in all the entelegyne species.
308 Other remnants of linkage between Super-Hox genes include *Dbx* and *Gbx* linked to
309 Hox cluster B in *H. graminicola*, *P. tepidariorum*, and *D. plantarius* (Figure 3). *En* is also
310 often linked to *Mnx*, *Gbx* to *Nedx*, and *Dbx* to *Evx*, though these have more frequently
311 been disrupted and dispersed, unlike the consistency of *Abox* and *Meox*. Synteny may
312 be conserved by the constraint of regulatory elements or may result from the passive
313 maintenance of ancestral linkage. While the Hox cluster has largely been retained in
314 duplicate in spiders, most of the Super-Hox families have returned to the single copy
315 and linkage has dispersed following the WGD.

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

327 The NK cluster

328 We found two ohnologous NK clusters in all spider genomes (Figure 4). One consists of
329 *Msx1-A*, *NK4-A*, *NK3-A*, *Tlx-A*, *Lbx-A*, and *NK7-A* clustered on one chromosome, while
330 *Msx1-B*, *Tlx-B*, *Lbx-B*, and *NK7-B* are found on a different chromosome, also with *NK3-*
331 *B* and *NK4-B* in the species that have retained these genes (core NK clusters are
332 delineated with grey boxes in Figure 4). Cluster B is less organised; *NK3-B* has likely
333 been lost in entelegynes (estimated divergence 280-310 MYA) (Fernández, Kallal, et al.
334 2018; Shao and Li 2018) and *NK4-B* has likely been lost in *H. graminicola* and *P.*
335 *tepidariorum*, independently. NK core cluster A is conserved across spiders, except that

336 *Msx1-A* is not found in *L. elegans* or *P. tepidariorum*, which may indicate the loss of
337 *Msx1-A* in a theridiid ancestor (est. 120 MYA) (Fernández, Kallal, et al. 2018). While NK
338 cluster A is found in a conserved complement and order in almost all spider species
339 analysed, it is rarely linked to any other genes belonging to the larger ancestral
340 bilaterian NK complement (Table 1). In contrast, NK cluster B has a varying gene
341 complement, having undergone lineage-specific gene losses and more rearrangements,

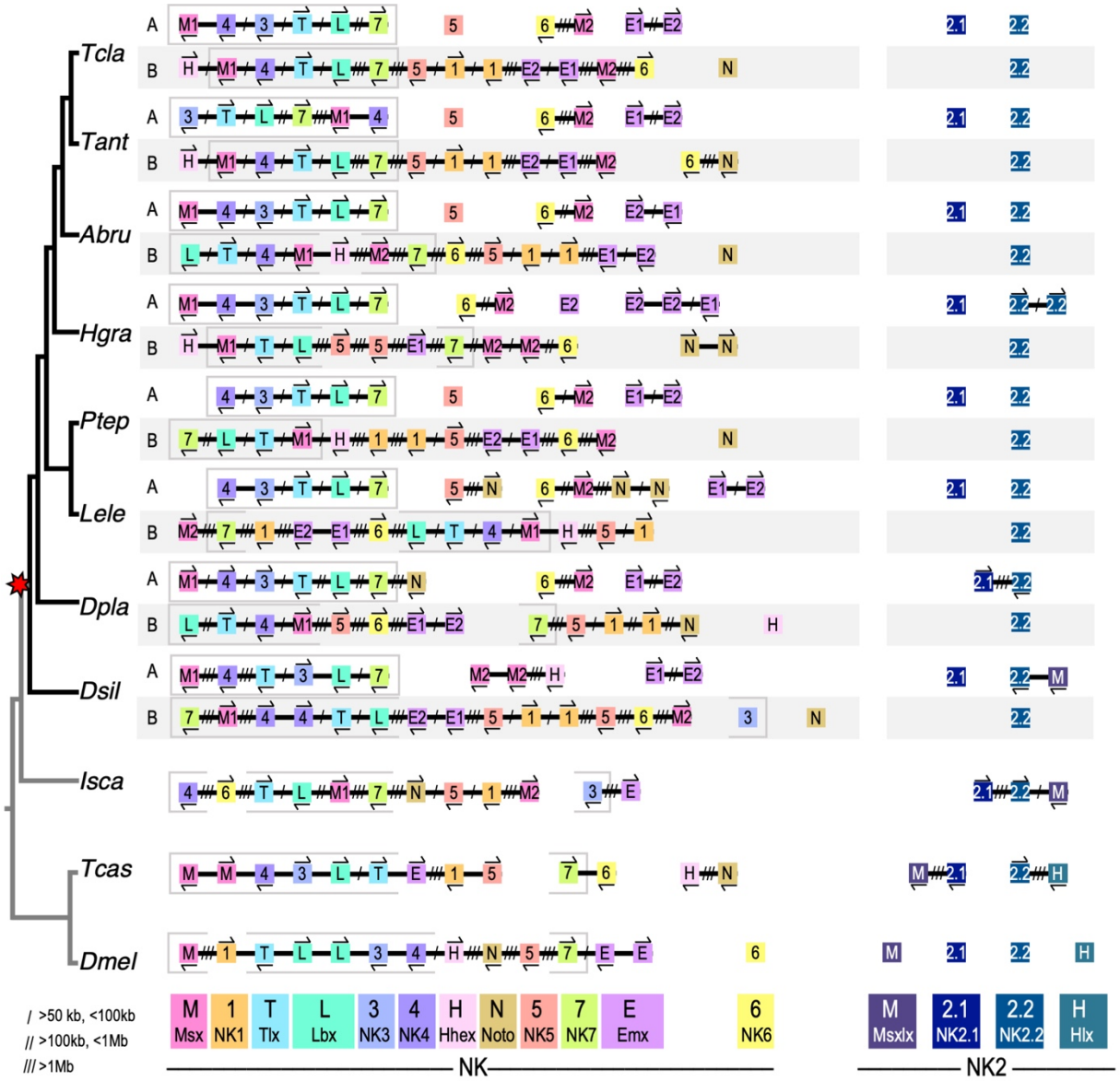


Figure 4: Two ohnologous NK clusters are retained, and a disintegrated NK2 cluster is found in spiders. Genes are represented by coloured boxes; chromosomes are black lines. Gene orientation is indicated by arrows above or below genes, and distances are indicated by slashes between genes. Gene families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

342 making it more degenerate and less conserved between species. This second cluster is,
343 however, often linked to other ancestral NK cluster genes *Hhex*, *Noto*, *Emx1-B*, *Emx2-*
344 *B*, *NK5-B*, and *NK1*, which has undergone a tandem duplication, and to *NK6-B* and
345 *Msx2-B* in all spiders (Figure 4). *NK6-A* is found linked to *Msx2-A* in all the entelegyne
346 species, and *NK6* and *Msx2* genes can be found in cluster A and/or B in all spiders,
347 suggesting this two-gene *NK6-Msx2* cluster predates the WGD and was retained in
348 duplicate some species, while others have lost either *Msx2* or *NK6* ohnologues.

349 The NK2 cluster has largely broken down in spiders, with successive gene losses: first
350 *Hlx* from chelicerates, and then *Msxlx* from entelegynes, though further lineage
351 sampling may change this pattern (Figure 4). In all spiders, *NK2.2* is retained in
352 duplicate on two separate chromosomes, indicating an ohnologous relationship, while
353 *NK2.1* has returned to the single copy (Figure 4). *NK2.1* and *NK2.2* are unlinked in all
354 spiders except for *D. plantarius*, indicating either that the cluster has separated twice, or
355 alternatively, that different ohnologues have been lost in different lineages.
356 Phylogenetics were not able to determine which ohnologue of *NK2.1* was retained in
357 different species because no examined genome contains both ohnologues.

358 The SINE cluster

359 There are five SINE-class genes in two clusters in spiders, consistent with a duplication
360 in the WGD followed by the loss of one *Six4/5* ohnologue. Spider SINE clusters likely
361 underwent different rearrangements in different lineages (Figure 5). *D. sylvatica* has
362 retained only one copy of each gene, while the entelegynes have retained B
363 ohnologues of both *Six1/2* and *Six3/6* in a conserved cluster. In araneids, in cluster A all
364 three genes are linked, though *Six4/5-A* is separated by a larger distance than that
365 separating *Six1/2-A* from *Six3/6-A*, which remain tightly clustered. There has also been
366 an inversion which may have occurred ancestrally among the theridiids that brought
367 *Six4/5-A* closer to *Six3/6-A* than *Six1/2-A*. This indicates that the separation of *Six4/5-A*
368 from the other two genes in araneids occurred after the divergence of the theridiids and
369 that the three genes making up SINE cluster A were likely still tightly clustered in their
370 shared ancestor. Among the outgroups, all three SINE genes are linked in the tick, and

371 only two are linked in the fly, while none are linked in the beetle, indicating the spider
372 SINE clusters have been retained more intact relative to other arthropod lineages.

373 The HRO cluster

374 The PRD-class HRO cluster was so named because the order of the genes in the fly
375 genome is *hbn* (*Hbn*), *Rx* (*Rax*), then *otp* (*Otp*); this order is conserved in insects and in
376 the tick genome, indicating that it is likely the ancestral arthropod arrangement (Figure
377 5). In spiders, the HRO cluster is found intact, but reordered: *Rax*, *Hbn*, then *Otp*
378 (Figure 5). There have also been several inversions of the orientations of the genes
379 within the cluster, but the R-H-O order is conserved across spiders, suggesting this
380 rearrangement occurred once ancestrally. It likely did not predate the WGD but
381 occurred in a spider ancestor because the scorpion HRO cluster A order is *Hbn*, *Rax*,
382 then *Otp*, as in insects and the tick (Table S1). As well as one conserved intact HRO
383 cluster, spiders retained only *Rax* in duplicate, while there are two *Rax* and two *Otp*
384 ohnologues retained in the scorpion (Figures 2 and 5).

385 The LIM-class gene *Isl* is consistently found linked to the HRO cluster (Figure 5). Ferrier
386 (2016) suggested this pattern is evidence for the ancestral linkage between the LIM-
387 class and PRD-class genes in the Giga-cluster of homeobox progenitor genes in an
388 animal ancestor. The linkage of *Isl* to the HRO cluster has been observed across
389 species with a most recent common ancestor in the ancestor of all animals. The
390 consistent linkage pattern between these genes in spiders, the tick, and beetle
391 genomes supports this association. This broad phylogenetic distribution makes this
392 between-class linkage unlikely to be a convergent secondary association; rather it is
393 more likely to be a conserved ancestral trait.

394 The Irx cluster

395 *Irx* genes have duplicated independently many times (see Table 1), giving rise to six
396 genes in spiders. Each of the ohnologous *Irx* clusters contains one of each gene type,
397 *Irx1*, *Irx2*, and *Irx3*, though *Irx2-B* has been lost from one of the clusters in the
398 entelegynes (Figure 5). Consistent with the findings of Kerner et al. (2009), insect *Irx*

genes fall into two clades, one containing orthologues of *D. melanogaster mirr* and the other with orthologues of *ara* and *caup* (Table 1 and Figure S1). Spider and scorpion *Irx1-A* and *Irx1-B* group with *ara* and *caup*, while spider and scorpion *Irx2-A* and *Irx2-B* group with *mirr* (Figure S1). The exact relationships between *Irx3* and other arthropod genes remains somewhat unclear. There are three *Irx* 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 *Irx1/ara/caup* or *Irx2/mirr* clade, and is resolved as sister to the *Irx3* clade in the *Irx* gene phylogeny, albeit with very low bootstrap support (Figure S1). The potential *Irx3* is clustered with *Irx1* and *Irx2* in the *S. maritima* genome in the same relative position as spider *Irx3* (Figures 5 and S3 C). There is a clear orthologue of each of the three arachnopulmonate *Irx* gene families in both the harvestman and tick genomes, indicating the three gene types originated in or before the chelicerate ancestor, and possibly *Irx3* genes were lost

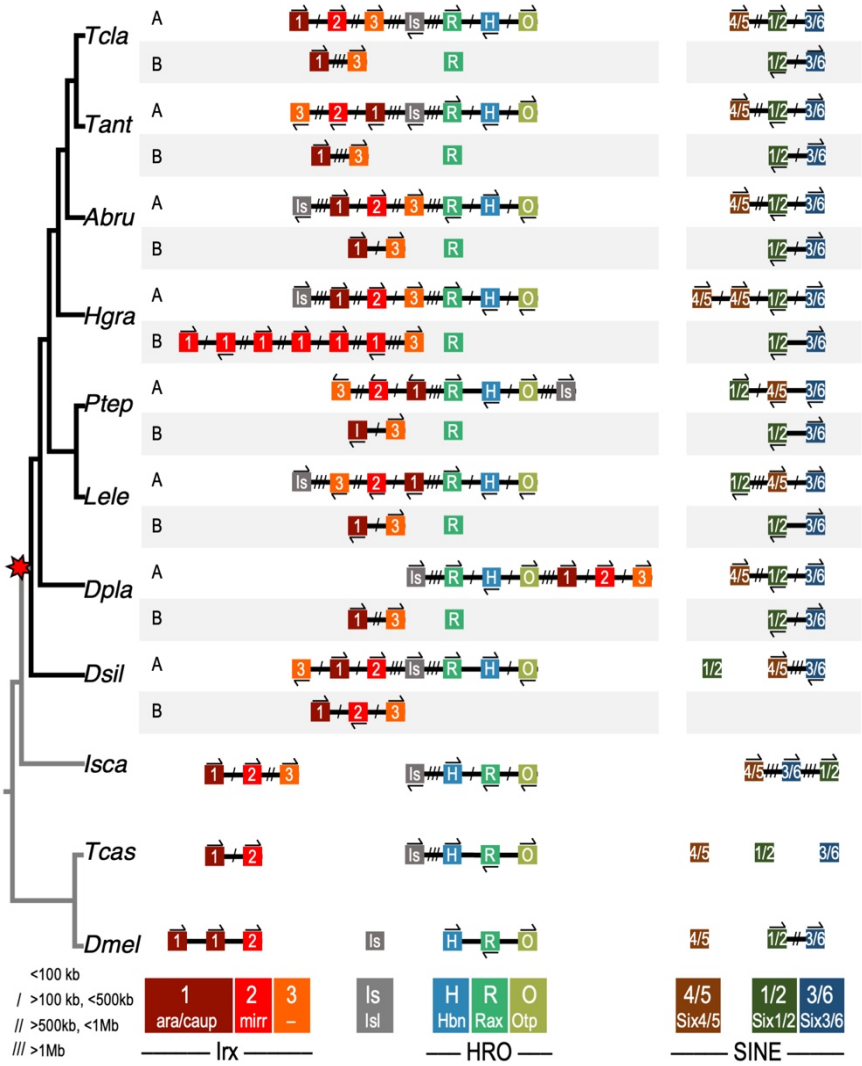


Figure 5: Two ohnologous HRO clusters, one of which is linked to one of the two ohnologous *Irx* clusters, and two ohnologous SINE clusters are retained in spiders. Genes are represented by coloured boxes; chromosomes are black lines. Gene orientation is indicated by arrows above or below genes, and distances are indicated by slashes between genes. Gene families are denoted by colour and a one- or two-letter code. Species abbreviations are as in Table 2.

430 from an insect ancestor after the divergence of Myriapoda (est. 520 MYA) (Fernández,
431 Edgecombe, et al. 2018). According to Kerner et al. (2009), *Irx* genes duplicated
432 independently in spiralians, ecdysozoans, and deuterostomes, suggesting that all
433 paralogues arose from a single ancestral bilaterian *Irx* gene. From our phylogeny
434 (Figure S1), we can tentatively infer that the duplication that gave rise to fly paralogues
435 is ancestral to the arthropods, and that there was a three-gene cluster in the ancestral
436 chelicerate that further duplicated in the arachnopulmonate WGD resulting in up to six
437 spider *Irx* paralogues in two ohnologous clusters.

438 In spiders, one of the two ohnologous *Irx* clusters is also linked to the HRO+*Isl* cluster
439 (Figure 5). Since this association occurs only to *Irx* cluster A, while *Irx* cluster B is
440 consistently not linked to the second *Rax* paralogue (Figure 5, Figure S3 E-L), a likely
441 explanation is that one *Irx* cluster translocated to be linked to HRO+*Isl* following the
442 WGD. We infer it occurred secondarily, i.e., on only one ancestral arachnopulmonate
443 chromosome, rather than deeper in animal ancestry due to the consistent lack of
444 linkage between HRO+*Isl* with *Irx* genes in any other species examined here, in contrast
445 to the linkage between HRO and *Isl* found across bilaterians.

446 *Conserved macrosynteny of spider genomes*

447 To contextualise our observations of homeobox cluster conservation and
448 rearrangements, we explored macrosynteny across spider chromosomes. We used *H.*
449 *graminicola* as the reference because the *D. silvatica* genome is highly rearranged
450 compared to all entelegynes in this study, and *H. graminicola* represents an early-
451 branching lineage realative to the majority of other entelegynes (Figure 6, Figure S4 A).
452 Across entelegynes, we detected clear patterns of orthology between chromosomes,
453 and consistency in the synteny blocks in which certain homeobox clusters are found,
454 but these patterns were not as evident when compared to the sole synspermiatan
455 representative *D. silvatica* (Figure 6, Figure S4 A). Depictions of locations of all
456 homeobox genes across genomes can be found in Figure S3.

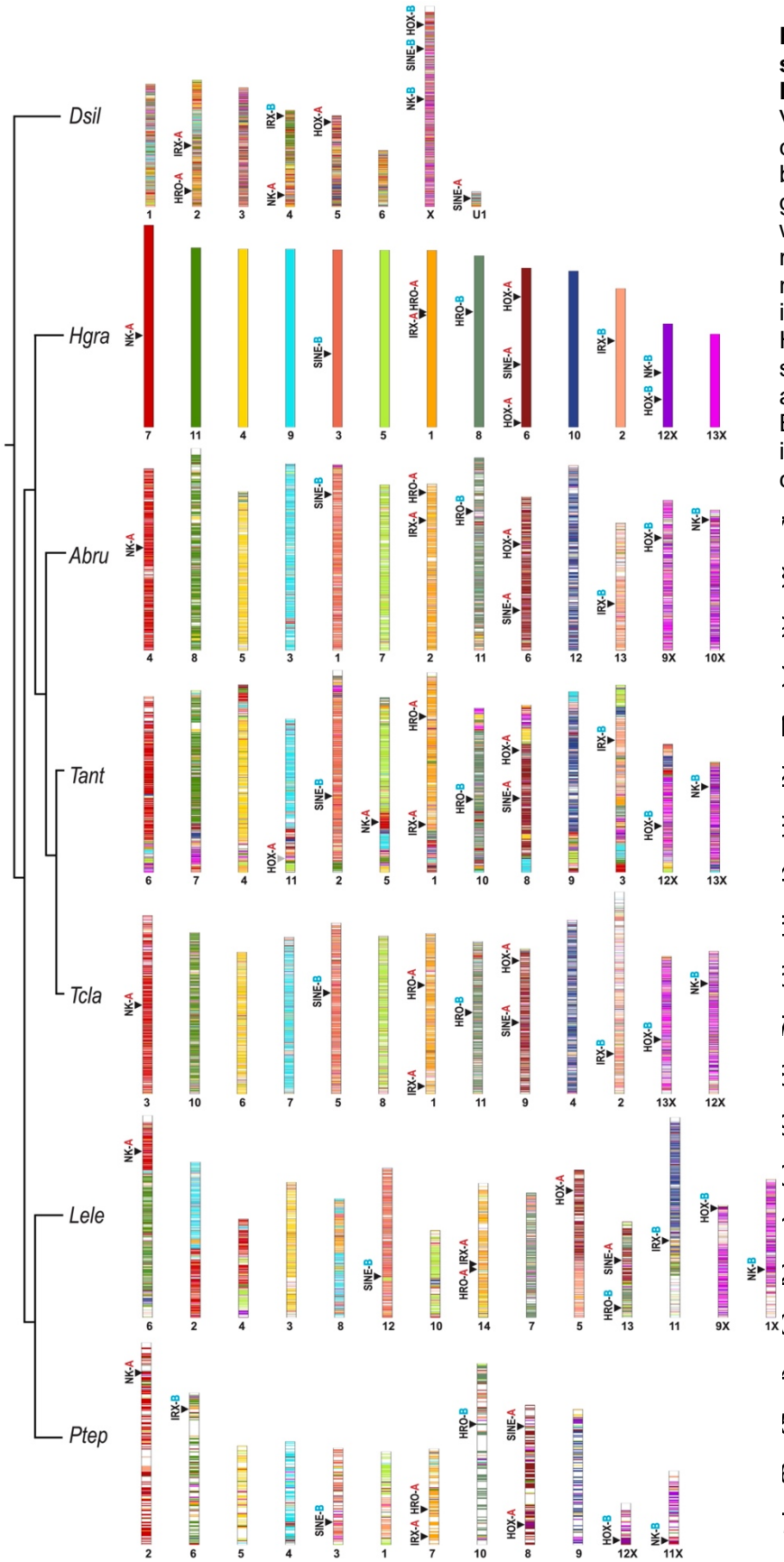


Figure 6 – Conserved macrosynteny of spider chromosomes and associated locations of homeobox gene clusters. Vertical bars show stacks of genes on each chromosome representing a macrosynteny block, with *H. graminicola* as the reference genome. Chromosomes of other species were ordered to show the orthologous relationships to *H. graminicola*. Chromosome numbers are denoted below, with 'X' indicating sex chromosomes. Locations of HOX, NK, IRX, SINE and HRO clusters are shown with black arrows, along with the assignment for cluster A (red) versus cluster B (blue). The grey triangle in *T. antipodiana* indicates the partial translocation of Hox cluster A.

In each araneid analysed here, there are 13 chromosomes; however, in the theridiids, there are 12 in *P. tepidariorum* and 14 in *L. elegans* (Figure 6, Figure S4 A). In *P. tepidariorum*, the synteny block corresponding to chromosome 2 in *H. graminicola* (coral pink in Figure 6) appears to have been mostly lost or distributed throughout the genome, and *Irx* cluster B which belongs to this block has been relocated (Figure 6, Figure S4 A). 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

478 being found on separate chromosomes, as well as the rearrangement of *Irx* cluster B
479 (Figures 6 and S3 G, Figure S4 A).

480 Sex chromosomes have been identified in several spider species (Figure 6 and Figure
481 S4 A) (Sheffer et al. 2021; Zhu et al. 2022; Miller et al. 2023). Interestingly, Hox cluster
482 B and NK cluster B are both found on the same sex chromosome in *H. graminicola* (12)
483 and *D. silvatica* (X), but on separate sex chromosomes in other entelegynes, whereas
484 Hox cluster A is located on an autosome (Figure 6, Figure S4 A). Across entelegynes,
485 NK cluster A is found in the same conserved synteny block corresponding to a full
486 chromosome in all species except for *Trichonephila antipodiana*. In this species, a
487 rearrangement brought NK cluster A onto another chromosome, and this is the only
488 entelegyne species in which the NK cluster A is rearranged (Figure 4). Overall, this
489 analysis illustrates the conservation of macrosynteny blocks and chromosomes across
490 entelegyne genomes and shows several chromosomal rearrangements that have
491 affected homeobox gene clusters.

492 *Analysis of selection on NK cluster genes*

493 We next sought to understand how homeobox genes have evolved following WGD and
494 tandem duplications. In the NK cluster, we identified families with single-copy
495 orthologues, ancient tandem paralogues, and WGD ohnologues. To determine if genes
496 with different relationships were under different selective pressures, we compared the
497 aBSREL estimated ω values calculated across coding sequence phylogenies for seven
498 NK cluster families with different inferred duplication histories: *Msx*, *NK4*, *NK3*, *Tlx*, *Lbx*,
499 *NK7*, and *Emx*. While none of the focal ohnologues or tandem paralogues were found to
500 be under positive selection, certain genes from the pre-duplicate outgroups were. *NK7*
501 was found to be under positive selection in *D. melanogaster* and *I. scapularis*, *Tlx* also
502 in *I. scapularis*, and *NK3* in *T. castaneum*. However, certain ohnologues did have longer
503 branches separating them from other ohnologues or the single-copy outgroup genes in
504 the aBSREL-generated phylogenies, indicating a higher substitution rate. These include
505 *Msx1* in arachnopulmonates and the harvestman, *NK3-A* in the entelegynes, and *Lbx-B*
506 in the scorpion. No consistent pattern was detected between A versus B ohnologues,

507 despite NK cluster A being more organised, nor was there a clear pattern between
508 ohnologues versus tandem duplicates. Supplementary material underpinning these
509 results is available on GitHub (see Materials and Methods).

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

520 **Table 3 – The ratio of non-synonymous to synonymous substitution rates between *P. tepidariorum***
521 **and *P. opilio* NK cluster genes shows no**
522 **evidence for positive selection among**
523 **paralogues**. Pairwise dN/dS values were
524 calculated with codeML for each *P. opilio*
525 and *P. tepidariorum* NK cluster gene. The
526 bars at the right show the dN/dS value
527 relative to the black bar (length of 1) at the
528 top. The relationship between each gene
529 pair is indicated in the third column, where 1
530 to 1 indicates direct orthology (the loss of an
531 ohnologue), 1 to 2 indicates the retention of
532 two ohnologues in the spider, O is the
533 comparison between the two spider
534 ohnologues, and T is the comparison
535 between tandem paralogues. For families
536 with ancient tandem duplications (*Msx* and
537 *Emx*), gene names have been updated to
538 reflect the inferred relationships between
539 paralogues in *P. opilio* and subsequent
540 ohnologues in *P. tepidariorum*, such that,
541 e.g., *P. opilio Emx1* is the pro-orthologue of
542 *P. tepidariorum Emx1-A* and *Emx1-B*, and
543 the same for *P. opilio Emx2* and *P.*
544 *tepidariorum Emx2-A* and *Emx2-B*. Raw
545 substitutions (S and N) and substitution
546 rates (dS and dN) as calculated with
547 codeML can be found in Table S3.
548

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

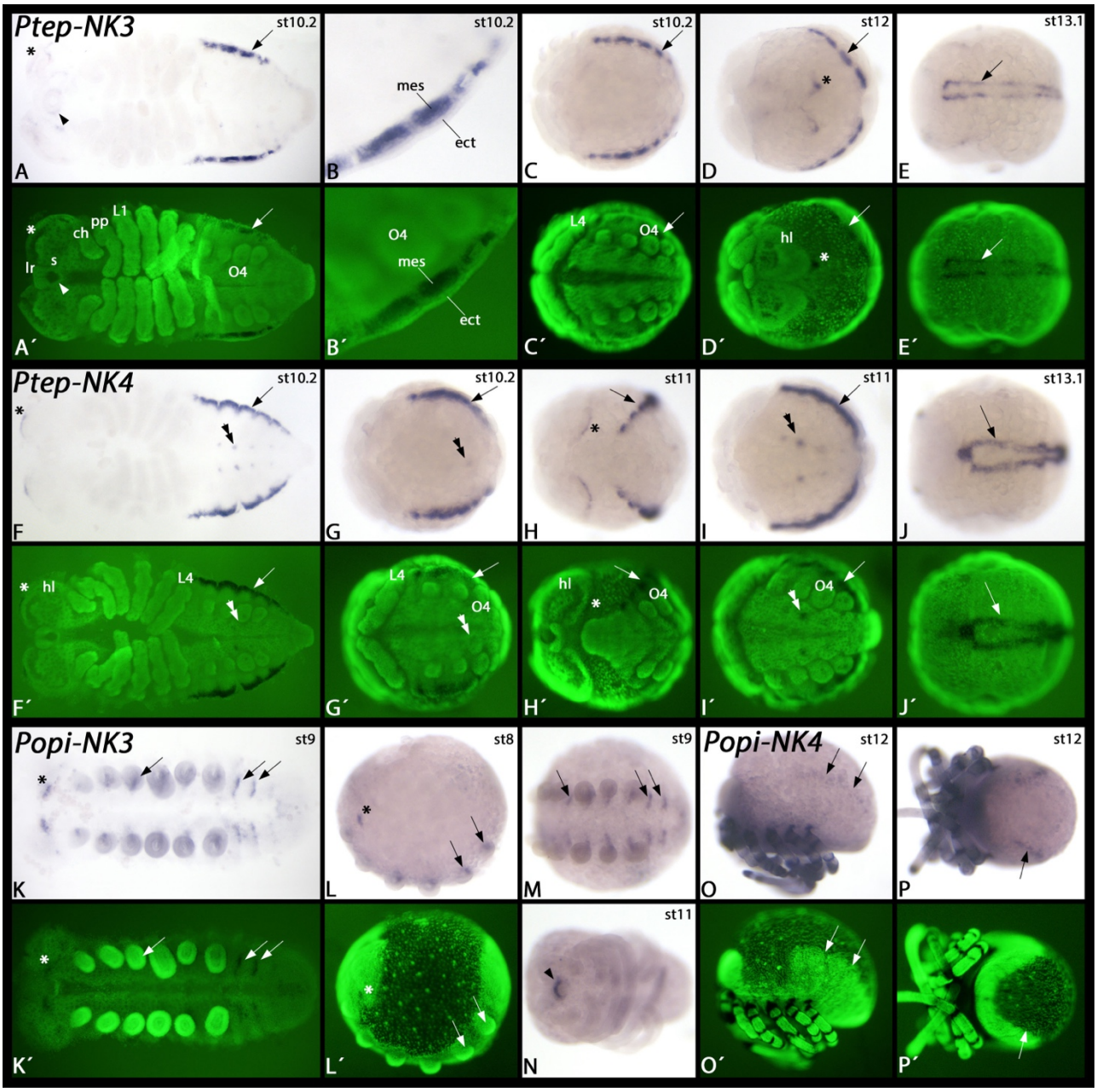


Figure 7 – Expression of *P. tepidariorum* and *P. opilio NK3* and *NK4*. In all panels, anterior is to the left. Panels A-C, F, G, I, K, M, and N show ventral views. Panels D and H show anterior-dorsal views. Panels E, and J show dorsal views. Panels L and O show lateral views and dorsal up. Panel P 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´-L´, O´ and P´ show SYBR green stained embryos as shown in panels A-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-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-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-L4, first to fourth walking leg; lr, labrum; mes, mesoderm; O, opisthosomal segment; pp, pedipalp; s, stomodaeum.

550 *Developmental expression of NK cluster genes in* P. tepidariorum *and* P. opilio

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

557 Single-copy genes

558 Only one copy each of *NK3* and *NK4* were found in both *P. opilio* and *P. tepidariorum*,
559 indicating the loss of an ohnologue of each from the spider following the WGD. The
560 single spider *NK3* is strongly expressed in the developing heart and is weakly
561 expressed around the stomodaeum and in the head (Figure 7 A-E). This is very similar
562 to the single spider *NK4*, which is also expressed strongly in the heart (Figure 7 F-J).
563 Unlike *NK3*, *NK4* is also expressed in a dot-like fashion in the ventral nervous system of
564 the opisthosoma, and at the anterior rim of the head lobes (Figure 7 asterisks and
565 double arrowheads in Figure 7 F-I). We did not detect any *NK3* expression in the
566 developing heart of the harvestman; instead *NK3* is expressed in a segmental fashion
567 ventral to the base of the appendages and in similar short transverse segmental stripes
568 in the opisthosomal segments (Figure 7 K-N). The expression of harvestman *NK3* is
569 thus more like the expression of its orthologue in *D. melanogaster* (Azpiazu and Frasch
570 1993) and the millipede *Glomeris marginata* (Figure S5) than that of the spider.
571 Expression of *NK4* in the harvestman is restricted to the developing dorsal vessel at late
572 developmental stages (Figure 7 O-P), a pattern that is conserved among panarthropods
573 (Azpiazu and Frasch 1993; Janssen and Damen 2008; Treffkorn et al. 2018).

574 Ancient tandem paralogues

575 Both *Emx* and *Msx* families have likely undergone an ancient tandem duplication,
576 resulting in two paralogues of each gene in the harvestman genome, with four *Emx* and
577 three *Msx* paralogues retained in *P. tepidariorum*. In previous transcriptome-based

578 analyses of *P. opilio*, the highly similar sequences for each *Emx* paralogue were
579 merged into a single transcript model, and only a partial sequence was found for *Msx1*,
580 split across two different transcript models (Sharma et al. 2012), but the genome
581 revealed distinct loci for each paralogue. In *P. tepidariorum*, Leite et al. (2018)
582 described *Emx1-A* and *Emx2-A* (previously designated *Emx4* and *Emx3*, respectively)
583 expression in the precheliceral region and in patches in each segment along the ventral
584 midline, and *Emx1-B* and *Emx2-B* expression (previously designated *Emx2* and *Emx1*)
585 in the anterior of each opisthosomal segment, with *Emx1-B* also expressed at the base
586 of the prosomal appendages. *P. opilio Emx2* (previously designated *Emx*) was detected
587 at the base of the appendages and in patches along the ventral midline in stage 10-11
588 embryos (Leite et al. 2018).

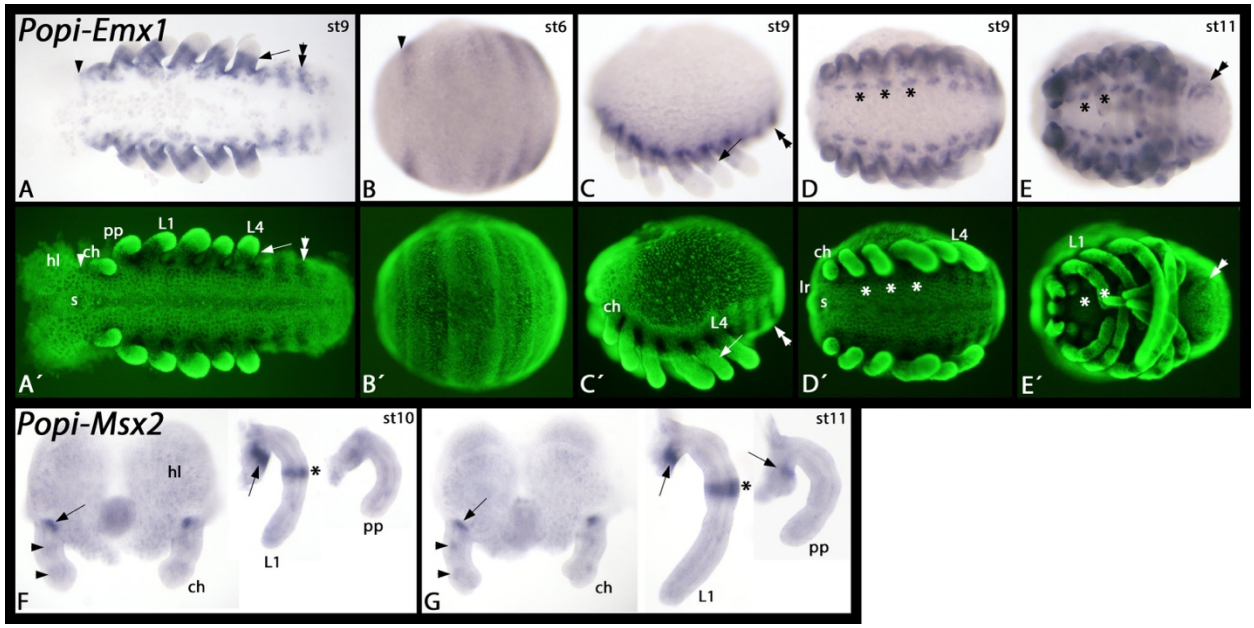


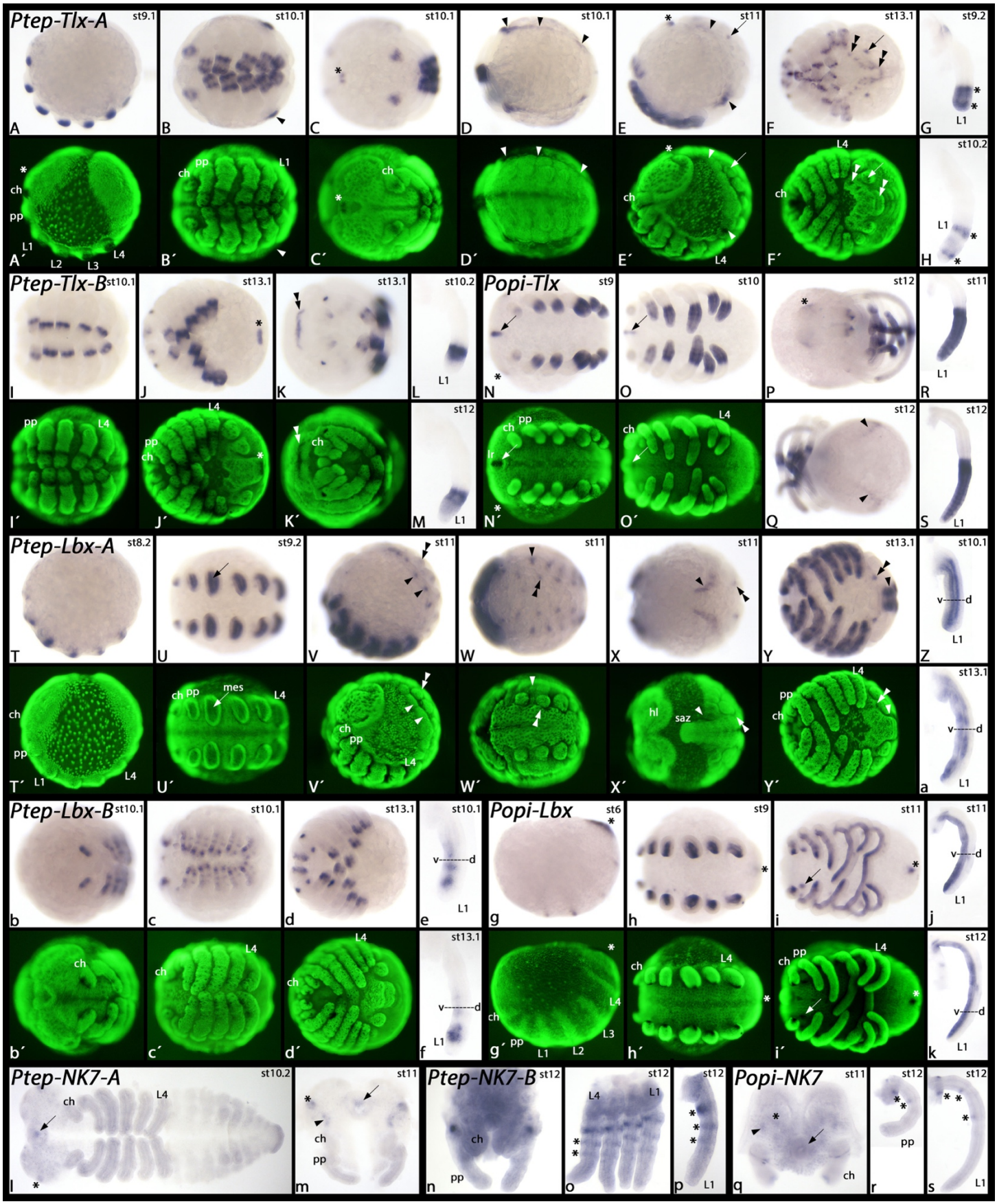
Figure 8 – Expression of *P. opilio Emx1* and *Msx2*. In panels A-E, anterior is to the left and ventral views (except panel C, lateral view and dorsal up). Panels F and G each show dissected head lobes and chelicerae, a walking leg and a pedipalp. The dissected heads are shown with anterior up, ventral 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.

589 Very similar to its paralogue *Emx2*, *P. opilio Emx1* is expressed in all posterior
590 segments including the cheliceral segment (Figure 8). The sharp anterior border of
591 expression is conserved among other arthropods, including spiders (Walldorf and
592 Gehring 1992; Simonnet et al. 2006; Schinko et al. 2008; Birkan et al. 2011; Janssen
593 2017; Leite et al. 2018) and an onychophoran (Janssen 2017). Indeed, the overall
594 expression pattern of *P. opilio Emx1* is identical with that of the previously reported
595 *Emx2*, with the exception that expression of *Emx1* is detected earlier.

596 According to Leite et al. (2018), *P. tepidariorum Msx1-B* (previously designated *Msx3*) is
597 expressed in the base of the prosomal appendages at stage 11, while *Msx2-B*
598 (previously designated *Msx2*) is expressed in the chelicerae at stage 12, and *Msx2-A*
599 (previously designated *Msx1*) is expressed in each segment along the ventral midline in
600 stages 8 and 10 (Akiyama-Oda and Oda 2020). In *P. opilio*, *Msx1* (previously
601 designated *Msx*) was detected in a stripe in each segment at stage 7 and along the
602 ventral midline in each segment at stage 13 (Leite et al. 2018). In contrast, we detected
603 *P. opilio Msx2* expression only at late developmental stages and in a very restricted dot-
604 like patch at the base of the appendages, and a ring-like domain in the legs (Figure 8 F-
605 G). Thus, *P. opilio Msx1* expression in the segments and along the ventral midline is
606 similar to *P. tepidariorum Msx2-A* expression, while *P. opilio Msx2* expression at the
607 base of appendages is similar to *P. tepidariorum Msx1-B* expression. *P. opilio Msx2*
608 expression was also detected in a small region of the chelicerae, matching the
609 expression of *Msx2-B* in stage 12 *P. tepidariorum* that was observed by Leite et al.
610 (2018) and interpreted as a novel expression domain. Our description of a matching
611 expression pattern from a previously unknown harvestman gene, *Msx2*, means that the
612 inferred neofunctionalisation is more likely a case of subfunctionalisation between the
613 tandem *Msx* paralogues.

614 Ohnologues

615 Three other NK families, *NK7*, *Tlx*, and *Lbx*, are found in the single copy in *P. opilio* and
616 have two retained ohnologues in *P. tepidariorum* (Figure 4). The two ohnologues of *P.*



617 *tepidariorum Tlx* are both expressed in the distal region of the legs, but in

Figure 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. Asterisks in panels G and H mark two stripes of stronger expression within the distal region of *Tlx-A* expression in the legs. 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-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.

complementary patterns (Figure 9 A-M). *Tlx-A* is first expressed in the complete distal region of the appendages and restricts into two rings of stronger expression around stage 10 (Figure 9 G-H). *Tlx-B* expression only begins after this point and is strongest in the region between the two rings of *Tlx-A* expression (Figure 9 M). 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 (Figure 9 B, D-E), and at later stages, in the ventral opisthosoma (Figure 9 F). In contrast, *Tlx-B* is only expressed in the opisthosoma at the very posterior at late developmental stages, after segment addition is complete (Figure 9 J). 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 Figure 9 K) 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 subfunctionalised. *P. opilio Tlx* is expressed in the distal half of the developing appendages (Figure 9 R-S), reflecting the combined expression of both spider *Tlx* ohnologues (Figure 9 G-H, L-M). Like *P. tepidariorum Tlx-A*, *P. opilio Tlx* is expressed in dorsal tissue of the opisthosoma and in the labrum (Figure 9 N-O), and like *P. tepidariorum Tlx-B*, *P. opilio Tlx* is expressed at the anterior edges of the head lobes (asterisk in Figure 9 P).

637 Expression of *P. tepidariorum Lbx-A* first appears in the mesoderm of the outgrowing
638 prosomal limb buds (Figure 9 T-U). Later, expression in the appendages is also seen in
639 the ventral ectoderm, but the dorsal ectoderm remains free of expression (Figure 9 W-
640 Y). Additional expression of *Lbx-A* is found in the dorsal ectoderm of the opisthosoma
641 and in segmental patches in the opisthosoma (Figure 9 V-Y). This latter expression is
642 possibly associated with the ventral-most tissue of the opisthosomal limb buds.
643 Compared to *Lbx-A*, expression of *P. tepidariorum Lbx-B* appears later and is restricted
644 to the prosomal appendages (Figure 9 b-d). *Lbx-B* expression within the developing
645 appendages is similar, but not identical, to that of *Lbx-A*; *Lbx-B* is expressed in fewer
646 ventral cells than *Lbx-A*. (Figure 9 e-f). Expression of the single *P. opilio Lbx* gene is
647 similar to that of *P. tepidariorum Lbx-A*, appearing early during development in the
648 mesoderm of the prosomal appendages (Figure 9 g-h), and later extending into ventral
649 ectodermal cells of the limbs (Figure 9 j-k). Unlike the two *P. tepidariorum Lbx*
650 ohnologues, *P. opilio Lbx* is also expressed at the posterior of the developing embryo
651 (asterisks in Figure 9 h-i).

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

659 **DISCUSSION**

660 *Evolution of the repertoire and organisation of spider homeobox genes before and after*
661 *the WGD*

662 We detected widespread conserved patterns of homeobox gene content and clusters
663 across spider genomes that were impacted both by tandem duplications and the WGD.
664 Leite et al. (2018) previously discovered that up to 59% of spider homeobox families

665 were duplicated in arachnopulmonates and that most of these were dispersed
666 duplicates rather than closely linked tandem duplicates. However, the few fragmented
667 spider genomes available at the time meant that whether these were ohnologues or
668 tandem duplicates could not reliably be determined. Here we showed that most
669 duplicated homeobox genes are likely ohnologues retained after WGD, and the majority
670 have been retained in all spiders surveyed (Figure 2 A, Table 2). The
671 arachnopulmonate WGD has therefore played a prominent role in shaping the
672 homeobox gene repertoires of spiders.

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

680 Contextualising the organisation of the conserved duplicated homeobox gene clusters
681 across entire spider genomes, we found that synteny is largely conserved among the
682 entelegynes (Figure 6). We also observed a few rearrangements of homeobox clusters
683 associated with chromosomal rearrangements (Figure S4 B and C). In *H. graminicola*,
684 Hox cluster B and NK cluster B are both on chromosome 12, one of the two sex
685 chromosomes, while in the other entelegynes, the synteny blocks that make up the *H.*
686 *graminicola* sex chromosomes are found on two distinct chromosomes, one with Hox
687 cluster B and the other with NK cluster B. This could have arisen by a rearrangement
688 among the sex chromosomes in *H. graminicola* or independent rearrangements in
689 theridiids and araneids. Given the X_1X_2 male and $X_1X_1X_2X_2$ female sex determination in
690 most entelegynes including the species shown here (Maddison and Leduc-Robert
691 2013), differential rates of purifying selection on sex chromosomes versus autosomes
692 may impact the extent to which the synteny of these chromosomes is conserved, thus

693 affecting the homeobox clusters on them (Kořínková and Král 2013; Charlesworth 2017;
694 Cordellier et al. 2020).

695 There seems to be a disconnect between the relative organisation of spider Hox
696 clusters and their regulation. Hox genes exhibit spatial and temporal collinearity (Ferrier
697 2019 and the references therein). The regulatory elements responsible for temporal
698 collinearity constrain Hox gene organisation, thus it has been hypothesised that Hox
699 cluster integrity is essential for the mechanism of temporal collinearity (Duboule 2007;
700 Krumlauf 2018; Ferrier 2019). While spatial collinearity is exhibited by both spider Hox
701 clusters, temporal collinearity is more evident in cluster A, though this cluster is less
702 organised (Schwager et al. 2017; Turetzek et al. 2022). Hox cluster organisation and
703 regulation vary between animals, for example, in mammals, each Hox cluster is
704 condensed and is regulated globally, while in some invertebrates the Hox cluster is
705 organised and expressed in sub-clusters (Kmita and Duboule 2003; Deschamps and
706 Duboule 2017; Wang et al. 2017). This necessitates a better understanding of the
707 relationship between the organisation and regulation of the ohnologous spider Hox
708 clusters.

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

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

722 bilaterian NK complement (Figure 4). Some NK gene organisation patterns seen in
723 spiders are similar to those in other arthropods. While the NK genes in different
724 drosophilid species have undergone a series of rearrangements, fragmenting and then
725 reuniting the NK cluster, certain gene pairs or triplets are consistently clustered and
726 move as units referred to as "contiguities" (Chan et al. 2015). In spiders, *NK3* and *NK4*
727 are adjacent, as in *D. melanogaster*, *T. castaneum*, *Bombyx terrestris*, many butterflies,
728 and *Anopheles gambiae*, and *Tlx* and *Lbx* are adjacent, as in *D. melanogaster*, *T.*
729 *castaneum*, several butterflies, *A. gambiae* and *I. scapularis* (Figure 4) (Luke et al.
730 2003; Chan et al. 2015; Mulhair et al. 2022). Though *NK1* has undergone a recent
731 tandem duplication in spiders, *NK5* and *NK1* are also adjacent in *T. castaneum*, *B.*
732 *terrestris*, *A. gambiae*, and *I. scapularis*, but not *D. melanogaster* or butterflies (Figure 4)
733 (Chan et al. 2015; Mulhair et al. 2022).

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

747 *NK cluster genes show conserved and subfunctionalised expression*

748 We observed that the genes in NK sub-clusters are expressed in similar patterns. NK
749 cluster genes are broadly involved in patterning the mesoderm (Jagla et al. 2001;
750 Saudemont et al. 2008; Treffkorn et al. 2018). The expression of *NK3* and *NK4* in both

751 arachnids is similar to the *D. melanogaster* orthologues *tin* and *bap* in the heart and
752 visceral mesoderm (Azpiazu and Frasch 1993). The fly *Tlx* orthologue, *C15*, and *Lbx*
753 paralogues *lbe* and *lbl* are also expressed in legs like their orthologues in both *P.*
754 *tepidariorum* and *P. opilio* (Jagla et al. 2001; Campbell 2005; Maqbool et al. 2006). This
755 suggests these genes share conserved cis-regulatory elements that may not only
756 preserve their clustering but also result in their co-expression. It would be interesting to
757 identify these regulatory elements and to investigate the extent of co-expression within
758 NK sub-clusters more broadly across bilaterians.

759 Though a few examples of potential neofunctionalisation have been described,
760 subfunctionalisation seems to be the more common fate for duplicated homeobox
761 genes (Schwager et al. 2017; Leite et al. 2018), and other genes (Schomburg et al.
762 2015; Janssen et al. 2021) in spiders. Using a more thorough annotation of NK genes
763 from the harvestman, we were able to determine the relative functional consequences
764 of tandem duplication versus the WGD, which was particularly important for families
765 with ancient tandem duplication events like *Emx* and *Msx*. Expression of *Emx* tandem
766 paralogues were similar while A and B ohnologues have undergone temporal
767 subfunctionalisation (Leite et al. 2018), despite the *Emx* tandem duplication now
768 estimated to predate the WGD. Furthermore, the two *Emx* paralogues in *P. opilio* are
769 expressed in near-identical patterns (Figure 8) (Leite et al. 2018), which suggests that
770 while the tandem duplication gave rise to two conserved paralogues, *Emx* genes did not
771 subfunctionalise until after the WGD. For *Msx*, we found that *P. opilio Msx2* exhibits a
772 similar expression pattern to *Msx2-B*, which now seems more likely to have undergone
773 subfunctionalisation instead of neofunctionalisation. This stresses the importance of
774 understanding the full evolutionary history of a gene family before inferring sub- and/or
775 neofunctionalisation. Furthermore, the roles of the ancestral single copies of *Msx* and
776 *Emx* are not yet known but could be determined with comparisons to earlier-branching
777 lineages of chelicerates or arachnids.

778 It is thought that duplicated genes encoding developmental transcription factors like
779 homeobox genes are more likely to undergo regulatory subfunctionalisation, driven by

780 the complementary loss of enhancers (Force et al. 1999; Jiménez-Delgado et al. 2009;
781 Espinosa-Cantú et al. 2015; Marlétaz et al. 2018). Indeed, the clear patterns of
782 subfunctionalisation observed for the *P. tepidariorum* NK ohnologues and single-copy
783 orthologues from *P. opilio* despite no NK cluster ohnologues being found to be under
784 positive selection suggests changes primarily occurred in their regulatory regions.

785 *Comparisons to the outcomes of other WGD events*

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

797 In comparison to vertebrate homeobox gene clusters resulting from the 2R WGD, the
798 ohnologous clusters of spiders are more conserved in terms of gene content. Across the
799 four human Hox clusters, complementary gene losses resulted in no cluster containing
800 all gene families, but in spiders, one Hox cluster is intact while the other has lost only
801 two genes (Figure 3). The four ohnologous human NK clusters are spread across five
802 chromosomes, and most genes are retained in two or maximum three copies (Luke et
803 al. 2003; Wotton et al. 2009). While in spiders the HRO cluster is intact, *Hbn* is not
804 found in chordates, and *Otp* is not linked to *Rax* or *Rax2* in the human genome (Mazza
805 et al. 2010). However, for SINE genes, the clusters have evolved similarly between
806 vertebrates and spiders following WGD (Kawakami et al. 2000; Ferrier 2016b). In both
807 of these lineages, linkage of all three SINE genes is conserved on one chromosome,
808 while the ohnologous cluster(s) have undergone losses and rearrangements. Therefore,

809 the pattern of one copy being retained intact while the other has degenerated, as seen
810 for most ohnologous spider homeobox gene clusters, is similar for only some human
811 homeobox clusters. Still, these few similarities illustrate how duplication can lead to
812 asymmetric degradation of otherwise conserved gene clusters.

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

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

832 **CONCLUSIONS**

833 Analyses of WGDs in arachnopulmonates, horseshoe crabs, and vertebrates shows
834 that homeobox genes, as well as other regulatory genes are commonly retained in
835 duplicate (Cañestro et al. 2013; Berthelot et al. 2014; Leite et al. 2016; Harper et al.
836 2021; Janssen et al. 2021). Across eight spiders, representing three major clades we

837 found that the majority of duplicated homeobox genes likely originated in the
838 arachnopulmonate WGD and that often, cluster or sub-cluster integrity has been
839 maintained in at least one ohnologous copy. Comparisons of expression patterns of
840 selected homeobox genes support extensive subfunctionalisation among spider
841 ohnologues. Taken together, our results help us understand the impact of WGD and
842 tandem duplication on the evolution of homeobox genes and genomes in spiders and
843 provide an important comparison to WGD in other animals.

844 MATERIALS AND METHODS

845 *Homeobox gene annotation*

846 We classified the homeobox gene complement of the publicly available chromosome-
847 level genomes from eight spider species, *D. silvatica* (Sánchez-Herrero et al. 2019;
848 Escuer et al. 2022), *D. plantarius* (Blaxter et al. 2022), *L. elegans* (Wang et al. 2022), *P.*
849 *tepidariorum* (Zhu et al. 2023), *H. graminicola* (Zhu et al. 2022), *A. bruennichi* (Sheffer
850 et al. 2021), *T. antipodiana* (Fan et al. 2021), and *T. clavata* (Hu et al. 2022). We
851 examined the previous homeobox gene classification of the scorpion *C. sculpturatus*,
852 which shared the ancestral WGD, as an outgroup to the spiders (Schwager et al. 2017;
853 Leite et al. 2018). As non-WGD outgroups, we included two arachnids, the harvestman
854 *P. opilio* (Gainett et al. 2021), and the tick *I. scapularis* (De et al. 2023), as well as three
855 mandibulates: the centipede *S. maritima* (Chipman et al. 2014; So et al. 2022), the fruit
856 fly *D. melanogaster* (BDGP6.32), in which many of these homeobox genes and families
857 were first described (Zhong et al. 2008), and the beetle *T. castaneum* (Herndon et al.
858 2020), which was used by Butts et al. (2008) to describe the Super-Hox cluster. Only
859 chromosome-level assemblies were used for synteny analyses, but homeobox genes
860 were counted and classified in all genomes. Genome accessions are shown in Table 4.

861 **Table 4 – Genome assemblies used in this study and accessions.** NCBI accession numbers or
862 sequence database DOIs are given for genomes with homeobox genes annotated in this study for the first
863 time. Citations are given for genomes with published homeobox gene annotations. Genome size, N50, and
864 BUSCO scores are shown for each genome as calculated for the annotated chromosome-level assemblies
865 with the arthropoda_odb10 database using BUSCO v5.4.7 or were taken from NCBI and the genome
866 papers (cited in the text above). Abbreviations: C: completeness; S: single copy; D: duplicated. The first-
867 listed *A. bruennichi* genome, sequenced by the Darwin Tree of Life project, had a higher N50 and larger
868 size, so was used for homeobox annotation, however it has not been annotated, so an older version, the

869 second-listed, was used for the whole genome synteny comparison as it has a corresponding annotation.
870 A homeobox annotation was also conducted on this older version and can be found at the bottom of Table
871 S1. The older *P. tepidariorum* assembly was not to chromosome-level so homeobox genes were
872 reannotated on the newer chromosome-level genome despite existing annotations on the previous version.

Species	Accession or Citation	Size	N50	BUSCO
T. clavata	https://spider.bioinfotoolkits.net/base/download/-1	2.6 Gb	202.1 Mb	C:83.8% [S:80.8%, D:3.0%]
T. antipodiana	http://dx.doi.org/10.5524/100868	2.3 Gb	172.9 Mb	C:96.3% [S:92.5%, D:3.8%]
A. bruennichi (DToL)	GCF_947563725.1	1.8 Gb	139.1 Mb	unannotated
A. bruennichi (OLD)	GCA_015342795.1	1.7 Gb	124.2 Mb	C:87.0% [S:81.5%, D:5.5%
H. graminicola	GCA_023701765.1	936.1 Mb	77.1 Mb	C:96.0% [S:92.2%, D:3.8%]
P. tepidariorum (OLD)	GCF_000365465.2; annotated by Schwager et al. 2017, Leite et al. 2018	1.4 Gb	4.1 Mb	98% completeness reported
P. tepidariorum (NEW)	https://doi.org/10.11922/sciencedb.o00019.00014	1.1 Gb	93.9 Mb	C:95.2% [S:92.4%, D:2.8%]
L. elegans	http://dx.doi.org/10.5524/102210	1.6 Gb	114.3 Mb	C:58.3% [S:55.2%, D:3.1%]
D. plantarius	GCA_907164885.2	2.8 Gb	216.7 Mb	unannotated
D. silvatica	GCA_006491805.2	1.4 Gb	174.2 Mb	C:77.2% [S:74.7%, D:2.5%]
C. sculpturatus	GCA_000671375.2; annotated by Schwager et al. 2017, Leite et al. 2018	925.5 Mb	537.5 kb	96.8% completeness reported
P. opilio	GCA_019434445.1	576.9 Mb	211 kb	95.1% reported
I. scapularis	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%]
S. maritima	GCA_000239455.1; annotated by Chipman et al. 2014, Leite et al. 2018	176.2 Mb	139.5 kb	93.8% completeness reported
T. castaneum	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%]
D. melanogaster	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%]

873

874 We used tBLASTn searches (Altschul et al. 1990) with the contents of HomeoDB
875 (accessed July 2020) (Zhong et al. 2008; Zhong and Holland 2011), as well as the
876 previous classification of *P. tepidariorum* homeobox genes (Schwager et al. 2017; Leite
877 et al. 2018) as the query set to retrieve homeodomain sequences from all eight spider
878 genomes, and re-annotated the homeobox genes for the harvestman, tick, and
879 centipede. Homeodomain sequences were then curated by tBLASTn searches against
880 annotated gene models or transcriptomes for all species except *D. plantarius*, for which
881 an annotation is not available, and classified based on sequence similarity and
882 diagnostic residues of the homeodomain for all families. For some gene families with
883 several duplication events, *e.g.*, *Irx*, *Pax4/6*, *Pax3/7*, as well as for the NK cluster genes,

884 phylogenies were constructed (see below). All homeodomain sequences can be found
885 in the supplementary information (Table S1).

886 *Phylogenetics*

887 To determine the timing of duplication of the arachnopulmonate *Irx* genes relative to
888 pre-duplicate paralogues, we made a full-length amino acid sequence alignment from a
889 from a subset of our spider species, the scorpion, and a transcriptome from another
890 synspermiatan spider, *Segestria senoculata* (Baudouin-Gonzalez unpublished). Full
891 sequences for genomes with poor annotations were curated from GenBank searches
892 and BLASTp (Altschul et al. 1990) searches against nr (Benson et al. 2018). These
893 sequences were aligned with MUSCLE set to the protein alignment defaults (Edgar
894 2004) and manually curated and trimmed to remove gaps. We inferred a Maximum-
895 Likelihood phylogeny with IQ-TREE (Nguyen et al. 2015), determined the best
896 substitution model with ModelFinder (Kalyaanamoorthy et al. 2017), and used 1,000
897 bootstrap replicates to optimise topology support. The phylogeny was visualised with
898 FigTree (FigTree 2018). The full and trimmed alignments and the phylogeny can be
899 found in the supplementary information on GitHub (Figure S1,
900 [https://github.com/madeleineaaseremedios/SpiderHomeoboxSequences]). Other family
901 phylogenies were made using only the homeodomain sequence and aligned manually,
902 but inferred as for the *Irx* family tree, except for the NK cluster and NK-linked genes
903 phylogeny, which used full protein sequences and was aligned with MUSCLE set to the
904 protein default (Edgar 2004) (Figure S2).

905 *Whole genome synteny comparison*

906 We performed whole genome synteny comparisons to explore chromosomal
907 rearrangements that might underlie trends in homeobox placement in spider genomes.
908 We used all spider species mentioned above, except for *D. plantarius* due to the lack of
909 genome annotation, and compared these to non-spider species *I. scapularis*, *T.*
910 *castaneum*, and *D. melanogaster*. Hierarchical orthogroups (HOGs) of all non-
911 homeobox genes were identified using Orthofinder (Emms and Kelly 2019) with default

912 settings, specifying the species tree:
913 ((Dmel,Tcas)N1,(Isca,(Dsil,((Ptep,Lele)N5,(Hgra,(Abru,(Tcla,Tant)N8)N7)
914 N6)N4)N3)N2)N0);. These were combined with HOGs of our homeobox orthology and
915 used to reconstruct ancestral gene orders with Agora (Muffato et al. 2023), using the
916 basic pipeline. However, we used Agora to illustrate chromosomes with macrosynteny
917 and orthology relationships, rather than the ancestral synteny blocks, termed
918 Contiguous Ancestral Regions (CARs) detected by Agora, because it resolved a better
919 signal of chromosomal rearrangements across the whole genome compared to the
920 smaller coverage of CARs. We also used MCScanX (Wang et al. 2012) to capture
921 microsynteny relationships between spider genomes and used Circos (Krzywinski et al.
922 2009) to plot synteny relationships and homeobox positions in the genome.

923 *Coding sequence evolutionary analyses*

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

936 We used aBSREL v2.3 (Smith et al. 2015) to detect selection on each branch of the
937 phylogeny for each of the spider NK core cluster gene families (*Msx*, *NK3*, *NK4*, *Tlx*,
938 *Lbx*, and *NK7*) and *Emx*. We used sequences from three spiders, *P. tepidariorum*, *D.*
939 *silvatica*, and *A. bruennichi*, and the scorpion, harvestman, tick, fly, and beetle, as
940 above, and rooted the trees with sequences from the onychophoran (Treffkorn et al.

941 2018). We created and trimmed codon alignments with PAL2NAL (Suyama et al. 2006)
942 from coding sequences and protein sequence alignments inferred with MUSCLE set to
943 the protein alignment default (Edgar 2004). We ran aBSREL using hyphy v2.5.46 and
944 the outputs were visualised with HyPhy Vision at [http://vision.hyphy.org/] (Pond et al.
945 2005; Kosakovsky Pond et al. 2020). We also conducted pairwise comparisons of
946 dN/dS ratios between *P. opilio* and *P. tepidariorum* for NK cluster genes calculated with
947 codeML (Goldman and Yang 1994) from codon alignments inferred as above for each
948 gene pair. Sequence data and aBSREL outputs can be found in the supplementary
949 information on GitHub at [https://github.com/madeleineaaseremedios/NKseqevoldata]
950 and visualised with HyPhy Vision as above.

951 *Animal husbandry, embryo collection, PCR, riboprobe synthesis, and* in situ
952 *hybridization*

953 Embryos of the spider *P. tepidariorum* were collected from the colony established in
954 Uppsala, Sweden, and maintained as described in Prpic et al. (2008). Embryos of the
955 harvestman *P. opilio* were collected from wild caught specimens in Uppsala, Sweden.
956 Adult specimens and embryos of *P. opilio* were treated as described in Janssen et al.
957 (2021). We applied the staging system of Mittmann and Wolff (2012) for *P.*
958 *tepidariorum*, and Gainett et al. (2022) for *P. opilio*. We investigated all developmental
959 stages of *P. tepidariorum* and *P. opilio* from the formation of the early germ band to
960 dorsal closure. Total RNA from *P. tepidariorum* and *P. opilio* was isolated from a mix of
961 embryonic stages using TRIzol (Invitrogen). mRNA was isolated from total RNA using
962 the Dynabeads mRNA Purification System (Invitrogen) and reverse transcribed into
963 cDNA using the SuperScriptII system (Invitrogen). PCRs were performed applying sets
964 of gene specific primers. For all investigated genes, reverse primers were ordered with
965 additional 5-prime T7-promotor sequences (gggTAATACGACTCACTATAG) for
966 subsequent antisense RNA probe synthesis (David and Wedlich 2001). *In situ*
967 hybridizations were performed as per Janssen et al. (2018). Primer sequences can be
968 found in Table S2.

969 **DATA AVAILABILITY STATEMENT**

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

AUTHOR CONTRIBUTIONS

APM and LSR conceived of the project. MEAR annotated the homeobox genes, conducted the phylogenetics and protein sequence analyses, and drafted the manuscript. RJ characterised the developmental expression of NK genes. DJL conducted the whole genome macrosynteny analysis. All authors contributed to writing and approved the final version of the manuscript.

ACKNOWLEDGEMENTS

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.

REFERENCES

Aase-Remedios ME, Ferrier DEK. 2021. Improved Understanding of the Role of Gene and Genome Duplications in Chordate Evolution With New Genome and Transcriptome Sequences. *Front. Ecol. Evol.* 9:429. https://doi.org/10.3389/fevo.2021.703163.

Abzhanov A, Popadic A, Kaufman TC. 1999. Chelicerate Hox genes and the homology of arthropod segments. *Evol. Dev.* 1:77–89. https://doi.org/10.1046/j.1525-142x.1999.99014.x.

Akiyama-Oda Y, Oda H. 2020. Hedgehog signaling controls segmentation dynamics and diversity via msx1 in a spider embryo. *Sci. Adv.* 6:eaba7261. https://doi.org/10.1126/sciadv.aba7261.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J. Mol. Biol.* 215:403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.

Averof M, Patel NH. 1997. Crustacean appendage evolution associated with changes in Hox gene expression. *Nature* 388:682–686. https://doi.org/10.1038/41786.

998 Azpiazu N, Frasch M. 1993. tinman and bagpipe: two homeo box genes that determine
999 cell fates in the dorsal mesoderm of Drosophila. *Genes Dev.* 7:1325–1340.
1000 https://doi.org/10.1101/gad.7.7b.1325.

1001 Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Ostell J, Pruitt KD, Sayers EW.
1002 2018. GenBank. *Nucleic Acids Res.* 46:D41–D47.
1003 https://doi.org/10.1093/nar/gkx1094.

1004 Berthelot C, Brunet F, Chalopin D, Juanchich A, Bernard M, Noël B, Bento P, Da Silva
1005 C, Labadie K, Alberti A, et al. 2014. The rainbow trout genome provides novel
1006 insights into evolution after whole-genome duplication in vertebrates. *Nat.*
1007 *Commun.* 5:3657. https://doi.org/10.1038/ncomms4657.

1008 Birkan M, Schaeper ND, Chipman AD. 2011. Early patterning and blastodermal fate
1009 map of the head in the milkweed bug Oncopeltus fasciatus. *Evol. Dev.* 13:436–
1010 447. https://doi.org/10.1111/j.1525-142X.2011.00497.x.

1011 Blaxter M, Mieszkowska N, Di Palma F, Holland P, Durbin R, Richards T, Berriman M,
1012 Kersey P, Hollingsworth P, Wilson W, et al. 2022. Sequence locally, think
1013 globally: The Darwin Tree of Life Project. *Proc. Natl. Acad. Sci.* [Internet] 119.
1014 https://doi.org/10.1073/pnas.2115642118. Available from:
1015 http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2115642118/-
1016 /DCSupplemental.

1017 Brooke NM, Garcia-Fernàndez J, Holland PWH. 1998. The ParaHox gene cluster is an
1018 evolutionary sister of the Hox gene cluster. *Nature* 392:920–922.
1019 https://doi.org/10.1038/31933.

1020 Brunet F, Lorin T, Bernard L, Haftek-Terreau Z, Galiana D, Schartl M, Volff JN. 2017.
1021 Case studies of seven gene families with unusual high retention rate since the
1022 vertebrate and teleost whole-genome duplications. In: Evolutionary Biology:
1023 Self/Nonself Evolution, Species and Complex Traits Evolution, Methods and
1024 Concepts. Springer International Publishing. p. 369–396.

1025 Bürglin TR. 2011. Homeodomain Subtypes and Functional Diversity. In: Hughes TR,
1026 editor. A Handbook of Transcription Factors. Subcellular Biochemistry.
1027 Dordrecht: Springer Netherlands. p. 95–122. Available from:
1028 https://doi.org/10.1007/978-90-481-9069-0_5

1029 Butts T, Holland PWH, Ferrier DEK. 2008. The Urbilaterian Super-Hox cluster. *Trends*
1030 *Genet.* 24:259–262. https://doi.org/10.1016/j.tig.2007.09.006.

1031 Butts T, Holland PWH, Ferrier DEK. 2010. Ancient homeobox gene loss and the
1032 evolution of chordate brain and pharynx development: deductions from
1033 amphioxus gene expression. *Proc. R. Soc. B Biol. Sci.* 277:3381–3389.
1034 https://doi.org/10.1098/rspb.2010.0647.

1035 Campbell G. 2005. Regulation of gene expression in the distal region of the Drosophila
1036 leg by the Hox11 homolog, C15. *Dev. Biol.* 278:607–618.
1037 https://doi.org/10.1016/j.ydbio.2004.12.009.

1038 Cañestro C, Albalat R, Irimia M, Garcia-Fernàndez J. 2013. Impact of gene gains,
1039 losses and duplication modes on the origin and diversification of vertebrates.
1040 *Semin. Cell Dev. Biol.* 24:83–94. https://doi.org/10.1016/j.semcdb.2012.12.008.

1041 Chan C, Jayasekera S, Kao B, Páramo M, von Grotthuss M, Ranz JM. 2015.
1042 Remodelling of a homeobox gene cluster by multiple independent gene reunions
1043 in Drosophila. *Nat. Commun.* 6:6509. https://doi.org/10.1038/ncomms7509.

1044 Charlesworth D. 2017. Evolution of recombination rates between sex chromosomes.
1045 *Philos. Trans. R. Soc. B Biol. Sci.* 372:20160456.
1046 https://doi.org/10.1098/rstb.2016.0456.

1047 Chipman AD, Ferrier DEK, Brena C, Qu J, Hughes DST, Schröder R, Torres-Oliva M,
1048 Znassi N, Jiang H, Almeida FC, et al. 2014. The First Myriapod Genome
1049 Sequence Reveals Conservative Arthropod Gene Content and Genome
1050 Organisation in the Centipede Strigamia maritima.Tyler-Smith C, editor. *PLoS*
1051 *Biol.* 12:e1002005. https://doi.org/10.1371/journal.pbio.1002005.

1052 Cordellier M, Schneider JM, Uhl G, Posnien N. 2020. Sex differences in spiders: from
1053 phenotype to genomics. *Dev. Genes Evol.* 230:155–172.
1054 https://doi.org/10.1007/s00427-020-00657-6.

1055 David R, Wedlich D. 2001. PCR-Based RNA Probes: A Quick and Sensitive Method to
1056 Improve Whole Mount Embryo In Situ Hybridizations. *BioTechniques* 30:768–
1057 774. https://doi.org/10.2144/01304st02.

1058 De S, Kingan SB, Kitsou C, Portik DM, Foor SD, Frederick JC, Rana VS, Paulat NS,
1059 Ray DA, Wang Y, et al. 2023. A high-quality Ixodes scapularis genome advances
1060 tick science. *Nat. Genet.* 55:301–311. https://doi.org/10.1038/s41588-022-01275-
1061 w.

1062 Deschamps J, Duboule D. 2017. Embryonic timing, axial stem cells, chromatin
1063 dynamics, and the Hox clock. *Genes Dev.* 31:1406–1416.
1064 https://doi.org/10.1101/gad.303123.117.

1065 Duboule D. 2007. The rise and fall of Hox gene clusters. *Development* 134:2549–2560.
1066 https://doi.org/10.1242/dev.001065.

1067 Edgar RC. 2004. MUSCLE: Multiple sequence alignment with high accuracy and high
1068 throughput. *Nucleic Acids Res.* 32:1792–1797.
1069 https://doi.org/10.1093/nar/gkh340.

1070 Emms DM, Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for
1071 comparative genomics. *Genome Biol.* 20:238. https://doi.org/10.1186/s13059-
1072 019-1832-y.

1073 Escuer P, Pisarenco VA, Fernández-Ruiz AA, Vizueta J, Sánchez-Herrero JF, Arnedo
1074 MA, Sánchez-Gracia A, Rozas J. 2022. The chromosome-scale assembly of the
1075 Canary Islands endemic spider Dysdera silvatica (Arachnida, Araneae) sheds
1076 light on the origin and genome structure of chemoreceptor gene families in
1077 chelicerates. *Mol. Ecol. Resour.* 22:375–390. https://doi.org/10.1111/1755-
1078 0998.13471.

1079 Espinosa-Cantú A, Ascencio D, Barona-Gómez F, De Luna A. 2015. Gene duplication
1080 and the evolution of moonlighting proteins. *Front. Genet.* 6:227.
1081 https://doi.org/10.3389/fgene.2015.00227.

1082 Fan Z, Yuan T, Liu P, Wang LY, Jin JF, Zhang F, Zhang ZS. 2021. A chromosome-level
1083 genome of the spider Trichonephila antipodiana reveals the genetic basis of its
1084 polyphagy and evidence of an ancient whole-genome duplication event.
1085 *GigaScience* 10:1–15. https://doi.org/10.1093/GIGASCIENCE/GIAB016.

1086 Fernández R, Edgecombe GD, Giribet G. 2018. Phylogenomics illuminates the
1087 backbone of the Myriapoda Tree of Life and reconciles morphological and
1088 molecular phylogenies. *Sci. Rep.* 8:83. https://doi.org/10.1038/s41598-017-
1089 18562-w.

1090 Fernández R, Kallal RJ, Dimitrov D, Ballesteros JA, Arnedo MA, Giribet G, Hormiga G.
1091 2018. Phylogenomics, Diversification Dynamics, and Comparative
1092 Transcriptomics across the Spider Tree of Life. *Curr. Biol.* 28:1489-1497.e5.
1093 https://doi.org/10.1016/j.cub.2018.03.064.

1094 Ferrier DEK. 2016a. The origin of the Hox/ParaHox genes, the Ghost Locus hypothesis
1095 and the complexity of the first animal. *Brief. Funct. Genomics* 15:333–341.
1096 https://doi.org/10.1093/bfgp/elv056.

1097 Ferrier DEK. 2016b. Evolution of Homeobox Gene Clusters in Animals: The Giga-
1098 Cluster and Primary vs. Secondary Clustering. *Front. Ecol. Evol.* 4:1–13.
1099 https://doi.org/10.3389/fevo.2016.00036.

1100 Ferrier DEK. 2019. Space and time in Hox/ParaHox gene cluster evolution. In: Fusco G,
1101 editor. Perspectives on Evolutionary and Developmental Biology: Essays for
1102 Alessandro Minelli. Padova: Padova University Press. p. 245–274.

1103 FigTree. 2018. FigTree v1.4.4. Available from: http://tree.bio.ed.ac.uk/software/figtree/

1104 Flot J-F, Hespeels B, Li X, Noel B, Arkhipova I, Danchin EGJ, Hejnol A, Henrissat B,
1105 Koszul R, Aury J-M, et al. 2013. Genomic evidence for ameiotic evolution in the

bdelloid rotifer Adineta vaga. *Nature* 500:453–457. https://doi.org/10.1038/nature12326.

Force A, Lynch M, Pickett FB, Amores A, Yan Y, Postlethwait J. 1999. Preservation of Duplicate Genes by Complementary, Degenerative Mutations. *Genetics* 151:1531–1545.

Gainett G, Crawford AR, Klementz BC, So C, Baker CM, Setton EVW, Sharma PP. 2022. Eggs to long-legs: embryonic staging of the harvestman Phalangium opilio (Opiliones), an emerging model arachnid. *Front. Zool.* 19:11. 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. 2021. The genome of a daddy-long-legs (Opiliones) illuminates the evolution of arachnid appendages. *Proc. R. Soc. B Biol. Sci.* 288:20211168. 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. 1999. 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* 61:82–91. https://doi.org/10.1006/geno.1999.5916.

Garcia-Fernàndez J. 2005. The genesis and evolution of homeobox gene clusters. *Nat. Rev. Genet.* 6:881–892. https://doi.org/10.1038/nrg1723.

Goldman N, Yang Z. 1994. A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol. Biol. Evol.* 11:725–736. https://doi.org/10.1093/oxfordjournals.molbev.a040153.

Harper A, Baudouin Gonzalez L, Schönauer A, Janssen R, Seiter M, Holzem M, Arif S, McGregor AP, Sumner-Rooney L. 2021. Widespread retention of ohnologs in key developmental gene families following whole-genome duplication in arachnopulmonates.Lott S, editor. *G3 GenesGenomesGenetics* 11:2020.07.10.177725. 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, Siemanowski J, et al. 2020. Enhanced genome assembly and a new official gene set for Tribolium castaneum. *BMC Genomics* 21:47. https://doi.org/10.1186/s12864-019-6394-6.

Hoegg S, Meyer A. 2005. Hox clusters as models for vertebrate genome evolution. *Trends Genet.* 21:421–424. https://doi.org/10.1016/j.tig.2005.06.004.

1141 Holland LZ, Ocampo Daza D. 2018. A new look at an old question: when did the second
1142 whole genome duplication occur in vertebrate evolution? *Genome Biol.* 19:209.
1143 https://doi.org/10.1186/s13059-018-1592-0.

1144 Holland PWH. 2013. Evolution of homeobox genes. *Wiley Interdiscip. Rev. Dev. Biol.*
1145 2:31–45. https://doi.org/10.1002/wdev.78.

1146 Holland PWH. 2015. Did homeobox gene duplications contribute to the Cambrian
1147 explosion? *Zool. Lett.* 1:1. https://doi.org/10.1186/s40851-014-0004-x.

1148 Holland PWH, Booth HAF, Bruford EA. 2007. Classification and nomenclature of all
1149 human homeobox genes. *BMC Biol.* 5:47. https://doi.org/10.1186/1741-7007-5-
1150 47.

1151 Holland PWH, Garcia-Fernàndez J. 1996. HoxGenes and Chordate Evolution. *Dev.*
1152 *Biol.* 173:382–395. https://doi.org/10.1006/dbio.1996.0034.

1153 Holland PWH, Marlétaz F, Maeso I, Dunwell TL, Paps J. 2017. New genes from old:
1154 asymmetric divergence of gene duplicates and the evolution of development.
1155 *Philos. Trans. R. Soc. B Biol. Sci.* 372:20150480.
1156 https://doi.org/10.1098/rstb.2015.0480.

1157 Hu W, Jia A, Ma S, Zhang G, Wei Z, Lu F, Luo Y, Zhang Z, Sun J, Yang T, et al. 2022.
1158 Molecular atlas reveals the tri-sectional spinning mechanism of spider dragline
1159 silk. :2022.06.21.496984. https://doi.org/10.1101/2022.06.21.496984. Available
1160 from: https://www.biorxiv.org/content/10.1101/2022.06.21.496984v1

1161 Hui JHL, McDougall C, Monteiro AS, Holland PWH, Arendt D, Balavoine G, Ferrier
1162 DEK. 2012. Extensive Chordate and Annelid Macrosynteny Reveals Ancestral
1163 Homeobox Gene Organization. *Mol. Biol. Evol.* 29:157–165.
1164 https://doi.org/10.1093/molbev/msr175.

1165 Irimia M, Maeso I, Garcia-Fernàndez J. 2008. Convergent Evolution of Clustering of
1166 Iroquois Homeobox Genes across Metazoans. *Mol. Biol. Evol.* 25:1521–1525.
1167 https://doi.org/10.1093/molbev/msn109.

1168 Irimia M, Tena JJ, Alexis MS, Fernandez-Miñan A, Maeso I, Bogdanovic O, de la Calle-
1169 Mustienes E, Roy SW, Gómez-Skarmeta JL, Fraser HB. 2012. Extensive
1170 conservation of ancient microsynteny across metazoans due to cis-regulatory
1171 constraints. *Genome Res.* 22:2356–2367. https://doi.org/10.1101/gr.139725.112.

1172 Jagla K, Bellard M, Frasch M. 2001. A cluster of Drosophila homeobox genes involved
1173 in mesoderm differentiation programs. *BioEssays* 23:125–133.
1174 https://doi.org/10.1002/1521-1878.

1175 Janssen R. 2017. Comparative analysis of gene expression patterns in the arthropod
1176 labrum and the onychophoran frontal appendages, and its implications for the

1177 arthropod head problem. *EvoDevo* 8:1. https://doi.org/10.1186/s13227-016-0064-
1178 4.

1179 Janssen R, Andersson E, Betnér E, Bijl S, Fowler W, Höök L, Leyhr J, Mannelqvist A,
1180 Panara V, Smith K, et al. 2018. Embryonic expression patterns and phylogenetic
1181 analysis of panarthropod sox genes: insight into nervous system development,
1182 segmentation and gonadogenesis. *BMC Evol. Biol.* 18:88.
1183 https://doi.org/10.1186/s12862-018-1196-z.

1184 Janssen R, Damen WGM. 2008. Diverged and conserved aspects of heart formation in
1185 a spider. *Evol. Dev.* 10:155–165. https://doi.org/10.1111/j.1525-
1186 142X.2008.00223.x.

1187 Janssen R, Eriksson BJ, Tait NN, Budd GE. 2014. Onychophoran Hox genes and the
1188 evolution of arthropod Hox gene expression. *Front. Zool.* 11:22.
1189 https://doi.org/10.1186/1742-9994-11-22.

1190 Janssen R, Pechmann M. 2023. Expression of posterior Hox genes and opisthosomal
1191 appendage development in a mygalomorph spider. *Dev. Genes Evol.* [Internet].
1192 https://doi.org/10.1007/s00427-023-00707-9. Available from:
1193 https://doi.org/10.1007/s00427-023-00707-9

1194 Janssen R, Pechmann M, Turetzek N. 2021. A chelicerate Wnt gene expression atlas:
1195 novel insights into the complexity of arthropod Wnt-patterning. *EvoDevo* 12:12.
1196 https://doi.org/10.1186/s13227-021-00182-1.

1197 Jiménez-Delgado S, Pascual-Anaya J, Garcia-Fernàndez J. 2009. Implications of
1198 duplicated cis-regulatory elements in the evolution of metazoans: the DDI model
1199 or how simplicity begets novelty. *Brief. Funct. Genomic. Proteomic.* 8:266–275.
1200 https://doi.org/10.1093/bfgp/elp029.

1201 Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. 2017.
1202 ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat.*
1203 *Methods* 14:587–589. https://doi.org/10.1038/nmeth.4285.

1204 Kawakami K, Sato S, Ozaki H, Ikeda K. 2000. Six family genes-structure and function
1205 as transcription factors and their roles in development. *BioEssays* 22:616–626.
1206 https://doi.org/10.1002/1521-1878(200007)22:7<616::AID-BIES4>3.0.CO;2-R.

1207 Kenny NJ, Chan KW, Nong W, Qu Z, Maeso I, Yip HY, Chan TF, Kwan HS, Holland
1208 PWH, Chu KH, et al. 2016. Ancestral whole-genome duplication in the marine
1209 chelicerate horseshoe crabs. *Heredity* 116:190–199.
1210 https://doi.org/10.1038/hdy.2015.89.

1211 Kerner P, Ikmi A, Coen D, Vervoort M. 2009. Evolutionary history of the iroquois/Irx
1212 genes in metazoans. *BMC Evol. Biol.* 9:74. https://doi.org/10.1186/1471-2148-9-
1213 74.

1214 Kmita M, Duboule D. 2003. Organizing Axes in Time and Space; 25 Years of Colinear
1215 Tinkering. *Science* 301:331–333. https://doi.org/10.1126/science.1085753.

1216 Kořínková T, Král J. 2013. Karyotypes, Sex Chromosomes, and Meiotic Division in
1217 Spiders. In: Nentwig W, editor. Spider Ecophysiology. Berlin, Heidelberg:
1218 Springer. p. 159–171. Available from: https://doi.org/10.1007/978-3-642-33989-
1219 9_12

1220 Kosakovsky Pond SL, Poon AFY, Velazquez R, Weaver S, Hepler NL, Murrell B, Shank
1221 SD, Magalis BR, Bouvier D, Nekrutenko A, et al. 2020. HyPhy 2.5—A
1222 Customizable Platform for Evolutionary Hypothesis Testing Using Phylogenies.
1223 *Mol. Biol. Evol.* 37:295–299. https://doi.org/10.1093/molbev/msz197.

1224 Krumlauf R. 2018. Hox genes, clusters and collinearity. *Int. J. Dev. Biol.* 62:659–663.
1225 https://doi.org/10.1387/ijdb.180330rr.

1226 Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra
1227 MA. 2009. Circos: an information aesthetic for comparative genomics. *Genome*
1228 *Res.* 19:1639–1645. https://doi.org/10.1101/gr.092759.109.

1229 Kumar JP. 2009. The sine oculis homeobox (SIX) family of transcription factors as
1230 regulators of development and disease. *Cell. Mol. Life Sci.* 66:565–583.
1231 https://doi.org/10.1007/s00018-008-8335-4.

1232 Larroux C, Fahey B, Degnan SM, Adamski M, Rokhsar DS, Degnan BM. 2007. The NK
1233 Homeobox Gene Cluster Predates the Origin of Hox Genes. *Curr. Biol.* 17:706–
1234 710. https://doi.org/10.1016/j.cub.2007.03.008.

1235 Larroux C, Luke GN, Koopman P, Rokhsar DS, Shimeld SM, Degnan BM. 2008.
1236 Genesis and Expansion of Metazoan Transcription Factor Gene Classes. *Mol.*
1237 *Biol. Evol.* 25:980–996. https://doi.org/10.1093/molbev/msn047.

1238 Leite DJ, Baudouin-Gonzalez L, Iwasaki-Yokozawa S, Lozano-Fernandez J, Turetzek
1239 N, Akiyama-Oda Y, Prpic N-M, Pisani D, Oda H, Sharma PP, et al. 2018.
1240 Homeobox Gene Duplication and Divergence in Arachnids. *Mol. Biol. Evol.*
1241 35:2240–2253. https://doi.org/10.1093/molbev/msy125.

1242 Leite DJ, Ninova M, Hilbrant M, Arif S, Griffiths-Jones S, Ronshaugen M, McGregor AP.
1243 2016. Pervasive microRNA Duplication in Chelicerates: Insights from the
1244 Embryonic microRNA Repertoire of the Spider Parasteatoda tepidariorum.
1245 *Genome Biol. Evol.* 8:2133–2144. https://doi.org/10.1093/gbe/evw143.

1246 Luke GN, Castro LFC, McLay K, Bird C, Coulson A, Holland PWH. 2003. Dispersal of
1247 NK homeobox gene clusters in amphioxus and humans. *Proc. Natl. Acad. Sci.*
1248 100:5292–5295. https://doi.org/10.1073/pnas.0836141100.

1249 Maddison WP, Leduc-Robert G. 2013. MULTIPLE ORIGINS OF SEX CHROMOSOME
1250 FUSIONS CORRELATED WITH CHIASMA LOCALIZATION IN HABRONATTUS
1251 JUMPING SPIDERS (ARANEAE: SALTICIDAE). *Evolution* 67:2258–2272.
1252 https://doi.org/10.1111/evo.12109.

1253 Maqbool T, Soler C, Jagla T, Daczewska M, Lodha N, Palliyil S, VijayRaghavan K,
1254 Jagla K. 2006. Shaping Leg Muscles in Drosophila: Role of ladybird, a
1255 Conserved Regulator of Appendicular Myogenesis. *PLOS ONE* 1:e122.
1256 https://doi.org/10.1371/journal.pone.0000122.

1257 Marlétaz F, Firbas PN, Maeso I, Tena JJ, Bogdanovic O, Perry M, Wyatt CDR, de la
1258 Calle-Mustienes E, Bertrand S, Burguera D, et al. 2018. Amphioxus functional
1259 genomics and the origins of vertebrate gene regulation. *Nature* 564:64–70.
1260 https://doi.org/10.1038/s41586-018-0734-6.

1261 Martin A, Serano JM, Jarvis E, Bruce HS, Wang J, Ray S, Barker CA, O'Connell LC,
1262 Patel NH. 2016. CRISPR/Cas9 Mutagenesis Reveals Versatile Roles of Hox
1263 Genes in Crustacean Limb Specification and Evolution. *Curr. Biol.* 26:14–26.
1264 https://doi.org/10.1016/j.cub.2015.11.021.

1265 Mazza ME, Pang K, Reitzel AM, Martindale MQ, Finnerty JR. 2010. A conserved cluster
1266 of three PRD-class homeobox genes (homeobrain, rx and orthopedia) in the
1267 Cnidaria and Protostomia. *EvoDevo* 1:3. https://doi.org/10.1186/2041-9139-1-3.

1268 McLysaght A, Hokamp K, Wolfe KH. 2002. Extensive genomic duplication during early
1269 chordate evolution. *Nat. Genet.* 31:200–204. https://doi.org/10.1038/ng884.

1270 Meyer A, Schartl M. 1999. Gene and genome duplications in vertebrates: the one-to-
1271 four (-to-eight in fish) rule and the evolution of novel gene functions. *Curr. Opin.*
1272 *Cell Biol.* 11:699–704. https://doi.org/10.1016/S0955-0674(99)00039-3.

1273 Miller J, Zimin AV, Gordus A. 2023. Chromosome-level genome and the identification of
1274 sex chromosomes in Uloborus diversus. *GigaScience* 12:giad002.
1275 https://doi.org/10.1093/gigascience/giad002.

1276 Mittmann B, Wolff C. 2012. Embryonic development and staging of the cobweb spider
1277 Parasteatoda tepidariorum C. L. Koch, 1841 (syn.: Achaearanea tepidariorum;
1278 Araneomorphae; Theridiidae). *Dev. Genes Evol.* 222:189–216.
1279 https://doi.org/10.1007/s00427-012-0401-0.

1280 Muffato M, Louis A, Nguyen NTT, Lucas J, Berthelot C, Roest Crollius H. 2023.
1281 Reconstruction of hundreds of reference ancestral genomes across the
1282 eukaryotic kingdom. *Nat. Ecol. Evol.* 7:355–366. https://doi.org/10.1038/s41559-
1283 022-01956-z.

1284 Mulhair P, Crowley L, Boyes DH, Harper A, Lewis OT, Consortium DT of L, Holland
1285 PWH. 2022. Diversity, duplication, and genomic organization of homeobox genes

in Lepidoptera. *Genome Res.*:gr.277118.122. https://doi.org/10.1101/gr.277118.122.

Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Mol. Biol. Evol.* 32:268–274. 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. 2021. Horseshoe crab genomes reveal the evolution of genes and microRNAs after three rounds of whole genome duplication. *Commun. Biol.* 4:83. https://doi.org/10.1038/s42003-020-01637-2.

Nossa CW, Havlak P, Yue J-X, Lv J, Vincent KY, Brockmann HJ, Putnam NH. 2014. Joint assembly and genetic mapping of the Atlantic horseshoe crab genome reveals ancient whole genome duplication. *GigaScience* 3:2047-217X-3–9. https://doi.org/10.1186/2047-217X-3-9.

Ohno S. 1970. Evolution by Gene Duplication. Berlin, Heidelberg: Springer Berlin Heidelberg Available from: http://link.springer.com/10.1007/978-3-642-86659-3

Pollard SL, Holland PWH. 2000. Evidence for 14 homeobox gene clusters in human genome ancestry. *Curr. Biol.* 10:1059–1062. https://doi.org/10.1016/S0960-9822(00)00676-X.

Pond SLK, Frost SDW, Muse SV. 2005. HyPhy: hypothesis testing using phylogenies. *Bioinformatics* 21:676–679. https://doi.org/10.1093/bioinformatics/bti079.

Prpic N-M, Schoppmeier M, Damen WGM. 2008. Collection and Fixation of Spider Embryos. *Cold Spring Harb. Protoc.* 2008: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. 2008. The amphioxus genome and the evolution of the chordate karyotype. *Nature* 453:1064–1071. https://doi.org/10.1038/nature06967.

Sánchez-Herrero JF, Frías-López C, Escuer P, Hinojosa-Alvarez S, Arnedo MA, Sánchez-Gracia A, Rozas J. 2019. The draft genome sequence of the spider Dysdera silvatica (Araneae, Dysderidae): A valuable resource for functional and evolutionary genomic studies in chelicerates. *GigaScience* 8:giz099. https://doi.org/10.1093/gigascience/giz099.

Saudemont A, Dray N, Hudry B, Le Gouar M, Vervoort M, Balavoine G. 2008. Complementary striped expression patterns of NK homeobox genes during segment formation in the annelid Platynereis. *Dev. Biol.* 317:430–443. https://doi.org/10.1016/j.ydbio.2008.02.013.

1322 Schinko JB, Kreuzer N, Offen N, Posnien N, Wimmer EA, Bucher G. 2008. Divergent
1323 functions of orthodenticle, empty spiracles and buttonhead in early head
1324 patterning of the beetle Tribolium castaneum (Coleoptera). *Dev. Biol.* 317:600–
1325 613. https://doi.org/10.1016/j.ydbio.2008.03.005.

1326 Schomburg C, Turetzek N, Schacht MI, Schneider J, Kirfel P, Prpic N-M, Posnien N.
1327 2015. Molecular characterization and embryonic origin of the eyes in the
1328 common house spider Parasteatoda tepidariorum. *EvoDevo* 6:15.
1329 https://doi.org/10.1186/s13227-015-0011-9.

1330 Schwager EE, Schoppmeier M, Pechmann M, Damen WGM. 2007. Duplicated Hox
1331 genes in the spider Cupiennius salei. *Front. Zool.* 4:10.
1332 https://doi.org/10.1186/1742-9994-4-10.

1333 Schwager EE, Sharma PP, Clarke T, Leite DJ, Wierschin T, Pechmann M, Akiyama-
1334 Oda Y, Esposito L, Bechsgaard J, Bilde T, et al. 2017. The house spider genome
1335 reveals an ancient whole-genome duplication during arachnid evolution. *BMC*
1336 *Biol.* 15:62. https://doi.org/10.1186/s12915-017-0399-x.

1337 Serano JM, Martin A, Liubicich DM, Jarvis E, Bruce HS, La K, Browne WE, Grimwood J,
1338 Patel NH. 2016. Comprehensive analysis of Hox gene expression in the
1339 amphipod crustacean Parhyale hawaiensis. *Dev. Biol.* 409:297–309.
1340 https://doi.org/10.1016/j.ydbio.2015.10.029.

1341 Shao L, Li S. 2018. Early Cretaceous greenhouse pumped higher taxa diversification in
1342 spiders. *Mol. Phylogenet. Evol.* 127:146–155.
1343 https://doi.org/10.1016/j.ympev.2018.05.026.

1344 Sharma PP, Schwager EE, Extavour CG, Giribet G. 2012. Hox gene expression in the
1345 harvestman Phalangium opilio reveals divergent patterning of the chelicerate
1346 opisthosoma. *Evol. Dev.* 14:450–463. https://doi.org/10.1111/j.1525-
1347 142X.2012.00565.x.

1348 Sheffer MM, Hoppe A, Krehenwinkel H, Uhl G, Kuss AW, Jensen L, Jensen C, Gillespie
1349 RG, Hoff KJ, Prost S. 2021. Chromosome-level reference genome of the
1350 European wasp spider Argiope bruennichi: a resource for studies on range
1351 expansion and evolutionary adaptation. *GigaScience* 10:1–12.
1352 https://doi.org/10.1093/GIGASCIENCE/GIAA148.

1353 Shimeld SM, Holland PWH. 2000. Vertebrate innovations. *Proc. Natl. Acad. Sci. U. S.*
1354 *A.* 97:4449–4452. https://doi.org/10.1073/pnas.97.9.4449.

1355 Shingate P, Ravi V, Prasad A, Tay BH, Garg KM, Chattopadhyay B, Yap LM, Rheindt
1356 FE, Venkatesh B. 2020. Chromosome-level assembly of the horseshoe crab
1357 genome provides insights into its genome evolution. *Nat. Commun.* 11:1–13.
1358 https://doi.org/10.1038/s41467-020-16180-1.

1359 Simakov O, Kawashima T. 2017. Independent evolution of genomic characters during
1360 major metazoan transitions. *Dev. Biol.* 427:179–192.
1361 https://doi.org/10.1016/j.ydbio.2016.11.012.

1362 Simakov O, Kawashima T, Marlétaz F, Jenkins J, Koyanagi R, Mitros T, Hisata K,
1363 Bredeson J, Shoguchi E, Gyoja F, et al. 2015. Hemichordate genomes and
1364 deuterostome origins. *Nature* 527:459–465. https://doi.org/10.1038/nature16150.

1365 Simakov O, Marlétaz F, Yue J, O'Connell B, Jenkins J, Brandt A, Calef R, Tung C,
1366 Huang T, Schmutz J, et al. 2020. Deeply conserved synteny resolves early
1367 events in vertebrate evolution. *Nat. Ecol. Evol.* 4:820–830.
1368 https://doi.org/10.1038/s41559-020-1156-z.

1369 Simonnet F, Célérier M-L, Quéinnec E. 2006. Orthodenticle and empty spiracles genes
1370 are expressed in a segmental pattern in chelicerates. *Dev. Genes Evol.* 216:467–
1371 480. https://doi.org/10.1007/s00427-006-0093-4.

1372 Smith MD, Wertheim JO, Weaver S, Murrell B, Scheffler K, Kosakovsky Pond SL. 2015.
1373 Less is more: an adaptive branch-site random effects model for efficient
1374 detection of episodic diversifying selection. *Mol. Biol. Evol.* 32:1342–1353.
1375 https://doi.org/10.1093/molbev/msv022.

1376 So WL, Nong W, Xie Y, Baril T, Ma H, Qu Z, Haimovitz J, Swale T, Gaitan-Espitia JD,
1377 Lau KF, et al. 2022. Myriapod genomes reveal ancestral horizontal gene transfer
1378 and hormonal gene loss in millipedes. *Nat. Commun.* 13:3010.
1379 https://doi.org/10.1038/s41467-022-30690-0.

1380 Suyama M, Torrents D, Bork P. 2006. PAL2NAL: robust conversion of protein sequence
1381 alignments into the corresponding codon alignments. *Nucleic Acids Res.*
1382 34:W609–W612. https://doi.org/10.1093/nar/gkl315.

1383 Tinti M, Johnson C, Toth R, Ferrier DEK, Mackintosh C. 2012. Evolution of signal
1384 multiplexing by 14-3-3-binding 2R-ohnologue protein families in the vertebrates.
1385 *Open Biol.* 2:120103. https://doi.org/10.1098/rsob.120103.

1386 Treffkorn S, Kahnke L, Hering L, Mayer G. 2018. Expression of NK cluster genes in the
1387 onychophoran Euperipatoides rowelli: implications for the evolution of NK family
1388 genes in nephrozoans. *EvoDevo* 9:17. https://doi.org/10.1186/s13227-018-0105-
1389 2.

1390 Turetzek N, Pechmann M, Janssen R, Prpic N-M. 2022. Hox genes in spiders: Their
1391 significance for development and evolution. *Semin. Cell Dev. Biol.*:in press.
1392 https://doi.org/10.1016/j.semcdb.2022.11.014.

1393 Walldorf U, Gehring W j. 1992. Empty spiracles, a gap gene containing a homeobox
1394 involved in Drosophila head development. *EMBO J.* 11:2247–2259.
1395 https://doi.org/10.1002/j.1460-2075.1992.tb05284.x.

1396 Wang S, Zhang J, Jiao W, Li J, Xun X, Sun Y, Guo X, Huan P, Dong B, Zhang L, et al.
1397 2017. Scallop genome provides insights into evolution of bilaterian karyotype and
1398 development. *Nat. Ecol. Evol.* 1:1–12. https://doi.org/10.1038/s41559-017-0120.

1399 Wang Y, Tang H, DeBarry JD, Tan X, Li J, Wang X, Lee T, Jin H, Marler B, Guo H, et
1400 al. 2012. MCScanX: a toolkit for detection and evolutionary analysis of gene
1401 synteny and collinearity. *Nucleic Acids Res.* 40:e49.
1402 https://doi.org/10.1093/nar/gkr1293.

1403 Wang Z, Zhu K, Li H, Gao L, Huang H, Ren Y, Xiang H. 2022. Chromosome-level
1404 genome assembly of the black widow spider Latrodectus elegans illuminates
1405 composition and evolution of venom and silk proteins. *GigaScience* 11:1–11.
1406 https://doi.org/10.1093/GIGASCIENCE/GIAC049.

1407 Wotton KR, Weierud FK, Juárez-Morales JL, Alvares LE, Dietrich S, Lewis KE. 2009.
1408 Conservation of gene linkage in dispersed vertebrate NK homeobox clusters.
1409 *Dev. Genes Evol.* 219:481–496. https://doi.org/10.1007/s00427-009-0311-y.

1410 Zhong YF, Butts T, Holland PWH. 2008. HomeoDB: A database of homeobox gene
1411 diversity. *Evol. Dev.* 10:516–518. https://doi.org/10.1111/j.1525-
1412 142X.2008.00266.x.

1413 Zhong YF, Holland PWH. 2011. HomeoDB2: Functional expansion of a comparative
1414 homeobox gene database for evolutionary developmental biology. *Evol. Dev.*
1415 13:567–568. https://doi.org/10.1111/j.1525-142X.2011.00513.x.

1416 Zhu B, Jin P, Hou Z, Li J, Wei S, Li S. 2022. Chromosomal-level genome of a sheet-
1417 web spider provides insight into the composition and evolution of venom. *Mol.*
1418 *Ecol. Resour.* 22:2333–2348. https://doi.org/10.1111/1755-0998.13601.

1419 Zhu B, Jin P, Zhang Y, Shen Y, Wang W, Li S. 2023. Genomic and transcriptomic
1420 analyses support a silk gland origin of spider venom glands. *BMC Biol.* 21:82.
1421 https://doi.org/10.1186/s12915-023-01581-7.

1422