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

Whole genome duplication (WGD) generates new genetic material that can contribute to the evolution of gene regulatory networks and phenotypes. After WGD, retained ohnologues can undergo subfunctionalisation to partition ancestral functions and/or neofunctionalisation, where one copy assumes a new function. We previously found that there had been a 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 were produced by WGD versus tandem duplications. Understanding this is key to determining the relative contributions of tandem duplications and WGD to arachnopulmonate genome evolution. Here we characterised the distribution of duplicated homeobox genes across eight chromosome-level spider genomes. We found that the majority of retained duplicate homeobox genes in spiders likely originated from WGD. We also found two copies of conserved homeobox gene clusters, including the Hox, NK, HRO, Irx, and SINE, in all eight species. Consistently, we noticed one degenerated copy of each cluster 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* to their single-copy orthologues in the harvestman *Phalangium opilio*. Our study provides new insights into the contributions of multiple modes of duplication to the homeobox gene repertoire during the evolution of spiders and the function of NK genes.

INTRODUCTION

Homeobox genes are a superclass of transcription factor, characterised by encoding the homeodomain, a highly conserved DNA-binding domain. There are eleven recognised classes of homeobox genes, with diverse and wide-reaching roles in development (Holland et al. 2007; Zhong et al. 2008; Kumar 2009; Butts et al. 2010; Bürglin 2011; Zhong and Holland 2011; Holland 2013; Holland et al. 2017). Many of these classes are thought to have expanded by tandem duplication early in animal evolution to give rise to more than 100 homeobox genes in the ancestral bilaterian (Larroux et al. 2008; Holland et al. 2017). It has been argued that certain homeobox gene expansions coincided with important innovations that occurred at the evolutionary origins of animals and subsequently bilaterians (Garcia-Fernàndez 2005; Holland 2013; Holland 2015). These tandem expansions resulted in clusters of related genes that have been conserved across the animal phylogeny (Pollard and Holland 2000; Larroux et al. 2007; Larroux et al. 2008; Ferrier 2016a). Homeobox clusters are thought to be maintained via 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).

The best known homeobox cluster is the Hox cluster. In arthropods, the Hox cluster contains ten ANTP-class genes: *labial* (*lab*, *Hox1*), *proboscipedia* (*pb*, *Hox2*), *Hox3*, *Deformed* (*Dfd*, *Hox4*), *Sex combs reduced* (*Scr*, *Hox5*), *fushi tarazu* (*ftz*, *Hox6-8*), *Antennapedia* (*Antp*, *Hox6-8*), *Ultrabithorax* (*Ubx*, *Hox6-8*), *abdominal-A* (*abd-A*, *Hox6-8*), and *Abdominal-B* (*Abd-B*, *Hox9-15*) (Table 1), organised in the genomes of most arthropods in this specific order. In some genomes, the cluster has been expanded or rearranged, for instance, the *Drosophila melanogaster* Hox cluster contains three *Hox3*-family genes, *zerknüllt* (*zen)*, *zen2*, and *bicoid* (*bcd*), and the complex has been separated between the Antennapedia complex from the Bithorax complex (Duboule 2007; Negre and Ruiz 2007). Nonetheless, Hox clusters are generally highly conserved in organisation and function, patterning the antero-posterior (A-P) axis in the same order as their relative genomic positions (Krumlauf 2018 and the references therein). Similarly, the ParaHox cluster, considered sister to the Hox cluster (Brooke et al. 1998),

also exhibits some collinearity in expression in a few extant animals and is thought to have regulated the A-P patterning of the through-gut of the ancestral bilaterian (Hui et al. 2009; Holland 2013). In deuterostomes and spiralians, the ParaHox cluster consists of three families, *Gsx*, *Xlox/Pdx*, and *Cdx*; however, in ecdysozoans, *Xlox/Pdx* is presumed lost (Ferrier and Minguillón 2003). The linkage between the two remaining ParaHox genes has broken down in flies and nematodes, indicating the overall disintegration of the ecdysozoan ParaHox cluster (Ferrier and Minguillón 2003).

Table 1 – Homeobox gene family names and synonyms in *D. melanogaster* for gene families belonging to conserved clusters. Table adapted from Ferrier (2016). The dash for *Irx3* indicates it is not found in *D. melanogaster*. Some of the 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).

	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*	
	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*

	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	*Six1/2*	*Six1/2*	*so*
	Six3/6	*Six3/6*	*optix*
	Six4/5	*Six4/5*	*Six4*

The NK cluster also belongs to the ANTP class. It was first described in *D. melanogaster* as a cluster of genes involved in mesodermal patterning (Jagla et al. 2001). This cluster contains the orthologues in the bilaterian gene families *Msx*, *NK1*, *Tlx*, *Lbx*, *NK3*, and *NK4* in *D. melanogaster*, which along with *Hhex*, *Noto*, *NK5*, *NK7*, *Emx*, and *NK6* make up the ancestral NK complement (Table 1) (Larroux et al. 2007; Butts et al. 2008; Li et al. 2020). Related to the NK cluster, the NK2 cluster contains *NK2.1*, *NK2.2*, and *Msxlx*, as well as *Hlx* (Table 2), which seem to exhibit conserved roles patterning the pharynx in deuterostomes (Butts et al. 2010; Simakov et al. 2015; Simakov and Kawashima 2017).

The PRD-class HRO cluster consists of orthologues of *D. melanogaster hbn*, *Rx*, and *otp* (Table 1), involved in brain and eye development across bilaterians, and neural development in a cnidarian (Mazza et al. 2010). SINE-class genes have conserved roles in the *Pax-Six-Eya-Dach* network in myogenesis, brain, and eye development across animals (Kawakami et al. 2000; Stierwald et al. 2004; Kumar 2009). The SINE class contains three bilaterian gene families, *Six1/2/sine oculis*, *Six3/6/optix*, and *Six4/5* (Table 1), which sit in a conserved three-gene cluster (Gallardo et al. 1999; Ferrier 2016b). The TALE-class genes in the *Iroquois* (*Irx*) family have undergone multiple independent tandem duplications resulting in clusters of various sizes in different bilaterian lineages, indicating that these genes may be uniquely prone to tandem duplication and retention (Irimia et al. 2008; Kerner et al. 2009). Their complex roles in development are less well characterised from an evolutionary perspective, but *Irx* genes have been implicated in eye development in *D. melanogaster* and neural crest in vertebrates, but within both flies and vertebrate models, *Irx* paralogues seem to act redundantly to one another (Cavodeassi et al. 2001).

Conserved linkage between homeobox clusters of the same and different classes in multiple animal lineages gave rise to the Giga-cluster hypothesis, which posits that the progenitor genes of each homeobox class were themselves situated in a tandem array in a pre-metazoan ancestor (Ferrier 2016b). Linkage between classes may be descendant from an ancestral association or reflect a recent association due to chromosomal rearrangements.

110 Homeobox genes and clusters have also been pervasively retained in duplicate
111 following whole genome duplications (WGD). There were two rounds (2R) of WGD in
112 the gnathostome ancestor, and many homeobox ohnologues were retained, including
113 four ohnologous Hox clusters (Holland and Garcia-Fernàndez 1996; McLysaght et al.
114 2002; Holland et al. 2007; Putnam et al. 2008; Cañestro et al. 2013; Holland 2013;
115 Holland and Ocampo Daza 2018). The 2R WGD appears to have relaxed constraint on
116 Hox cluster organisation so that while the complement of genes from *Hox1* to *Hox14* are
117 present in at least one copy, no cluster contains all 14 genes (Hoegg and Meyer 2005).
118 Positive selection has also been detected between paralogues among the four
119 vertebrate Hox clusters (Lynch et al. 2006).

120 It is thought that retained WGD paralogues (ohnologues) have contributed to the
121 evolution of vertebrate novelties (Meyer and Schartl 1999; Shimeld and Holland 2000;
122 Cañestro et al. 2013; Aase-Remedios and Ferrier 2021). Duplicated genes are
123 hypothesised to be released from functional constraints, allowing paralogues to
124 subfunctionalise, specialise, or neofunctionalise while collectively maintaining full
125 functionality, either by changes to their cis-regulatory controls or protein sequence
126 (Force et al. 1999; Jiménez-Delgado et al. 2009; Tinti et al. 2012; Espinosa-Cantú et al.
127 2015; Marlétaz et al. 2018). While the 2R WGD has been credited with facilitating the
128 evolution of vertebrate novelties, the extent to which WGDs in other animal lineages
129 have contributed to diversification is not well understood.

130 An independent WGD took place in the arachnopulmonate ancestor (Figure 1), resulting
131 in many homeobox genes retained in duplicate in spiders and scorpions and their
132 relatives, including two ohnologous Hox clusters (Schwager et al. 2017; Leite et al.
133 2018; Harper et al. 2021; Ontano et al. 2021). Between the Hox genes, each ohnologue
134 exhibits different spatial and temporal expression, though overall, spatial and temporal
135 collinearity is largely intact (Schwager et al. 2007; Schwager et al. 2017; Turetzek et al.
136 2022). Beyond the two Hox clusters, nearly 60% of all homeobox gene families were
137 found to have retained duplicates in spiders and scorpions, with previous descriptions
138 totalling 145 homeobox genes belonging to 80 families in the spider *Parasteatoda*
139 *tepidariorum* and 156 genes in 82 families in the scorpion *Centruroides sculpturatus*

140 (Schwager et al. 2017; Leite et al. 2018). Of these retained paralogues, many exhibit
141 divergent expression consistent with both sub- and neofunctionalisation (Leite et al.
142 2018).

143 In previous surveys, certain paralogues were identified as likely tandem duplicates
144 based on their relative positions in the scaffold-level genome assembly for *P.*
145 *tepidariorum*. For instance, both two-gene clusters of *Emx* paralogues were complete
146 on their own scaffold (Leite et al. 2018). For other genes, however, their relationships
147 could not be reliably determined, particularly for families with more than two paralogues
148 or dispersed tandem duplicates like the *Msx* family (Leite et al. 2018). In the previous *P.*
149 *tepidariorum* assembly, only one of the two Hox clusters was found on a single scaffold,
150 so it could not be determined if the other was also intact (Schwager et al. 2017). Now
151 with several chromosome-level assemblies available for spiders including *P.*
152 *tepidariorum* (Figure 1), relative genomic positions of paralogous genes across species
153 can provide important information in determining their evolutionary origins. This includes
154 using chromosomal location and linkage to distinguish tandem duplication from WGD
155 and comparing species from different spider and outgroup lineages to estimate the
156 timing of duplications.

157 We surveyed the homeobox gene complement and synteny across eight recently
158 released chromosome-level spider genome assemblies (Figure 1). For the NK cluster,
159 we complemented the characterisation of the expression of *Emx* and *Msx* family genes
160 (Leite et al. 2018) by describing the temporal and spatial expression of the rest of the
161 NK cluster genes during embryogenesis in the harvestman *P. opilio* and their
162 orthologues and ohnologues from *P. tepidariorum*.

Our analysis of chromosome-level spider genomes allowed us to distinguish ohnologues resulting from WGD versus paralogues from tandem duplication for the first time. We found that many spider homeobox genes and clusters were retained in duplicate following the WGD. 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. For NK cluster genes, we detected no evidence of selection on coding sequence in our comparisons between paralogues, but for each family with retained ohnologues in *P. tepidariorum*, we found evidence for subfunctionalisation during embryogenesis. Combined, these results provide new insight into both the patterns of ohnologue retention and their consequences in spider evolution.

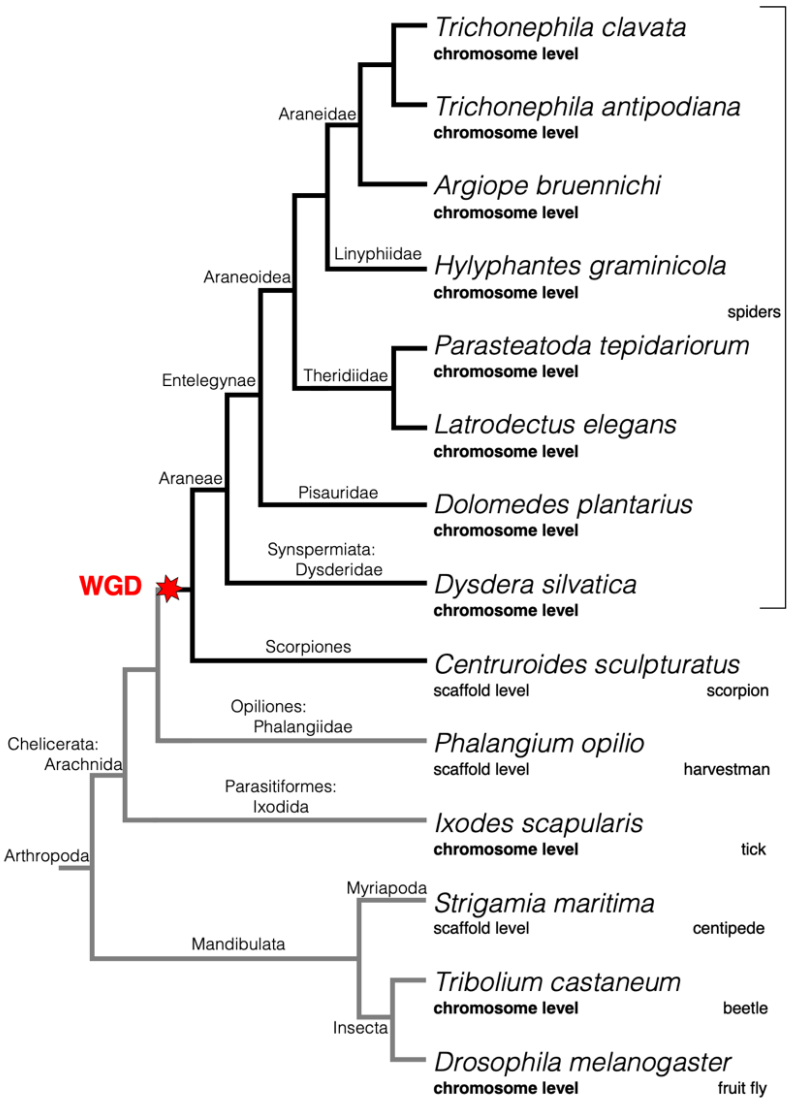

Figure 1 – Phylogeny of arthropods used in this study. The timing of the whole genome duplication (WGD) is indicated by the red star; pre-duplicate lineages have grey lines, duplicate genomes have black lines. Genome assembly level is indicated below the species binomial, in bold for the eight chromosome-level spider and three non-WGD outgroup genome assemblies used for the synteny analyses.

185 **MATERIALS AND METHODS**

186 *Homeobox gene annotation*

187 We classified the homeobox gene complement of the publicly available chromosome-
188 level genomes from eight spider species, *Dysdera silvatica* (Sánchez-Herrero et al.
189 2019; Escuer et al. 2022), *Dolomedes plantarius* (Blaxter et al. 2022), *Latrodectus*
190 *elegans* (Wang et al. 2022), *P. tepidariorum* (Zhu et al. 2023), *Hylyphantes graminicola*
191 (Zhu et al. 2022), *Argiope bruennichi* (Sheffer et al. 2021), *Trichonephila antipodiana*
192 (Fan et al. 2021), and *Trichonephila clavata* (Hu et al. 2022). We examined the previous
193 homeobox gene classification of the scorpion *C. sculpturatus*, which shared the
194 ancestral WGD, as an outgroup to the spiders (Schwager et al. 2017; Leite et al. 2018).
195 As non-WGD outgroups, we included two arachnids, the harvestman *P. opilio* (Gainett
196 et al. 2021), and the tick *Ixodes scapularis* (De et al. 2023), as well as three
197 mandibulates: the centipede *Strigamia maritima* (Chipman et al. 2014; So et al. 2022),
198 the fruit fly *D. melanogaster* (BDGP6.32), in which many of these homeobox genes and
199 families were first described (Zhong et al. 2008), and the beetle *Tribolium castaneum*
200 (Herndon et al. 2020), which was used by Butts et al. (2008) to describe the Super-Hox
201 cluster. Only chromosome-level assemblies were used for synteny analyses, but
202 homeobox genes were counted and classified in all genomes. We used tBLASTn
203 searches (Altschul et al. 1990) with the contents of HomeoDB (accessed July 2020)
204 (Zhong et al. 2008; Zhong and Holland 2011), as well as the previous classification of *P.*
205 *tepidariorum* homeobox genes (Schwager et al. 2017; Leite et al. 2018) as the query set
206 to retrieve homeodomain sequences from all eight spider genomes, and re-annotated
207 the homeobox genes for the harvestman, tick, and centipede. Homeodomain
208 sequences were then curated by tBLASTn searches against annotated gene models or
209 transcriptomes for all species except *D. plantarius*, for which an annotation is not
210 available, and classified based on sequence similarity and diagnostic residues of the
211 homeodomain for all families. For some gene families with several duplication events,
212 *e.g.*, *Irx*, *Pax4/6*, *Pax3/7*, as well as the NK cluster genes, phylogenies were constructed
213 (see below). All homeodomain sequences can be found in the supplementary
214 information (Table S1).

215 *Phylogenetics*

216 To determine the timing of duplication of the arachnopulmonate *Irx* genes relative to
217 pre-duplicate paralogues, we made a full-length amino acid sequence alignment from a
218 from a subset of our spider species, the scorpion, and a transcriptome from another
219 synspermiatan spider, *Segestria senoculata* (Baudouin-Gonzalez unpublished). Full
220 sequences for genomes with poor annotations were curated from GenBank searches
221 and BLASTp (Altschul et al. 1990) searches against nr (Benson et al. 2018). These
222 sequences were aligned with MUSCLE set to the protein alignment defaults (Edgar
223 2004) and manually curated and trimmed to remove gaps. We inferred a Maximum-
224 Likelihood phylogeny with IQ-TREE (Nguyen et al. 2015), determined the best
225 substitution model with ModelFinder (Kalyaanamoorthy et al. 2017), and used 1,000
226 bootstrap replicates to optimise topology support. The phylogeny was visualised with
227 FigTree (FigTree 2018). The full and trimmed alignments and the phylogeny can be
228 found in the supplementary information (Figure S1,
229 [https://github.com/madeleineaaseremedios/SpiderHomeoboxSequences]). Other family
230 phylogenies were made using only the homeodomain sequence and aligned manually,
231 but inferred as for the *Irx* family tree, except for the NK cluster and NK-linked genes
232 phylogeny, which used full protein sequences and was aligned with MUSCLE set to the
233 protein default (Edgar 2004) (Figure S2).

234 *Whole genome synteny comparison*

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

244 basic pipeline. However, we used Agora to illustrate chromosomes with macrosynteny
245 and orthology relationships, rather than the Contiguous Ancestral Regions detected by
246 Agora, because we achieved better signal of chromosomal macrosynteny.

247 *Coding sequence evolutionary analyses*

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

260 We used aBSREL v2.3 (Smith et al. 2015) to detect selection on each branch of the
261 phylogeny for each of the spider NK core cluster gene families (*Msx*, *NK3*, *NK4*, *Tlx*,
262 *Lbx*, and *NK7*) as well as *Emx*. We used sequences from three spiders, *P. tepidariorum*,
263 *D. silvatica*, and *A. bruennichi*, and the scorpion, harvestman, tick, fly, and beetle, as
264 above, and rooted the trees with sequences from the onychophoran (Treffkorn et al.
265 2018). We created and trimmed codon alignments with PAL2NAL (Suyama et al. 2006)
266 from coding sequences and protein sequence alignments inferred with MUSCLE set to
267 the protein alignment default (Edgar 2004). We ran aBSREL using hyphy v2.5.46 and
268 the outputs were visualised with HyPhy Vision at [http://vision.hyphy.org/]. We also
269 conducted pairwise comparisons of dN/dS ratios between *P. opilio* and *P. tepidariorum*
270 for NK cluster genes calculated with codeML (Goldman and Yang 1994) from codon
271 alignments inferred as above for each gene pair. Sequence data and aBSREL outputs

272 can be found at [https://github.com/madeleineaaseremedios/NKseqevoldata] and
273 visualised with HyPhy Vision as above.

274 *Animal husbandry, embryo collection, PCR, riboprobe synthesis, and* in situ
275 *hybridization*

276 Embryos of the spider *P. tepidariorum* were collected from the colony established in
277 Uppsala, Sweden, maintained as described in Prpic et al. (2008). Embryos of the
278 harvestman *P. opilio* were collected from wild caught specimens in Uppsala, Sweden.
279 Adult specimens and embryos of *P. opilio* were treated as described in Janssen et al.
280 (2021). We applied the staging system of Mittmann and Wolff (2012) for *P.*
281 *tepidariorum*, and Gainett et al. (2022) for *P. opilio*. We investigated all developmental
282 stages of *P. tepidariorum* and *P. opilio* from the formation of the early germ band to
283 dorsal closure. Total RNA from *P. tepidariorum* and *P. opilio* was isolated from a mix of
284 embryonic stages using TRIzol (Invitrogen). mRNA was isolated from total RNA using
285 the Dynabeads mRNA Purification System (Invitrogen). mRNA was reverse transcribed
286 into cDNA using the SuperScriptII system (Invitrogen). PCRs were performed applying
287 sets of gene specific primers. For all investigated genes, reverse primers were ordered
288 with additional 5-prime T7-promotor sequences (gggTAATACGACTCACTATAG) for
289 subsequent antisense RNA probe synthesis (David and Wedlich 2001). *In situ*
290 hybridizations were performed as per Janssen et al. (2018). Primer sequences can be
291 found in Table S2.

292 *Data availability statement*

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

295 **RESULTS**

296 *Homeobox gene repertoires in arthropods*

297 We surveyed the homeobox gene repertoires of eight spider species with chromosome-
298 level genome assemblies, representing seven entelegyne and one synspermiatan

species and compared these to the scorpion *C. sculpturatus*, the harvestman *P. opilio* and the tick *I. scapularis*, as well as three mandibulate species (Figures 1, 2 and Table 2). We found that spiders and scorpions have similar numbers of homeobox genes, and on average, have 1.4 times more homeobox genes than outgroup species that did not have an ancestral WGD (Table 2). The spiders and scorpion had an average of 150 homeobox genes; the most are found in *H. graminicola* with 172 and the fewest in *D. silvatica* with 140. Mandibulates and non-arachnopulmonate arachnids have on average 107 homeobox genes, with the fewest in the insects (103) and the most in *S. maritima* (113). Previous surveys found 96 and 69 homeobox genes in *I. scapularis* and *P. opilio*, respectively, 145 in *P. tepidariorum* and 156 in *C. sculpturatus* (Schwager et al. 2017; Leite et al. 2018), but we identified between 1-55% more in each of these species (Table 2). For *P. opilio*, searching both the recent genome assembly (Gainett et al. 2021) as well as the developmental transcriptome (Sharma et al. 2012) provided a more complete dataset and enabled our identification of more homeobox genes than previously described.

A total of 90 described homeobox families was represented in our arthropod species set (Table 2), though some were absent in different species such that no species had a gene from every family (Figure 2 A, Table S1). In addition, we also found several new homeobox genes that could not be classified to family. A few of these may represent new lineage-specific unclassified (un) families, temporarily called *ANTP-un1*, *PRD-un1*, *SINE-un1*, and *TALE-un1*; *ANTP-un1* is found only in the chelicerates, while *PRD-un1* and *SINE-un1* was found only in the Araneids, and *TALE-un1* only in the *Trichonephila* spp. (Figure 2 A, Table S1). The ‘1’ is simply a placeholder to indicate that these genes have an orthologue in at least two species and to distinguish them from other unclassified homeodomain sequences that have no similarity to described families and were found only in one species, sometimes in several copies.

Comparing this taxonomic breadth of species in concert with genomic synteny enabled us to distinguish ancient and recent tandem duplicates from likely ohnologues and determine the timing of duplications relative to major radiations in arthropod evolution. We have defined ancient tandem duplicates as those that are often linked or clustered

329 in most spider genomes and date to a duplication node in an ancestor before the
330 arachnopulmonate WGD, while recent tandem duplicates are also linked or clustered,
331 but have occurred either after the WGD or in the outgroups. Six homeobox gene
332 families, *Arx*, *Barhl*, *Irx*, *Pax3/7*, *Pax4/6*, and *Lhx2/9*, underwent ancient tandem
333 duplication events that are shared by arthropods and may have ancestry even deeper in
334 the bilaterian phylogeny (Figure 2 B). Only three of these ancient tandem paralogues
335 have been subsequently lost; just one *Lhx2/9* gene is found in *D. melanogaster*, and
336 only one *Barhl* gene is found in *T. castaneum* and *D. silvatica*. Within the chelicerates
337 but predating the arachnopulmonate WGD, three further tandem duplications occurred.
338 *Bsx* and *Msx* likely duplicated in the common ancestor of ticks, harvestmen and
339 arachnopulmonates, and *Emx* likely duplicated in the ancestor of harvestmen and
340 arachnopulmonates. *NK1* duplicated in tandem, likely just after the WGD, while the
341 other ohnologue was probably lost. Besides these duplications, the majority of tandem
342 duplications in spiders were more recent and lineage-specific. In particular, *H.*
343 *graminicola* has the highest number of families with recent tandem duplications (23
344 families) (Figure 2 B), resulting in this species having the highest number of homeobox
345 genes identified (172). Although this could be an assembly artefact, since overall, the *H.*
346 *graminicola* genome did not exhibit an exceptionally large number of duplicated genes
347 (Zhu et al. 2022).

348 **Table 2 – Homeobox genes across arthropods are largely stable, except for an increase in the**
349 **arachnopulmonates.** Species abbreviations: *Dmel Drosophila melanogaster*, *Tcas Tribolium castaneum*,
350 *Smar Strigamia maritima*, *Isca Ixodes scapularis*, *Popi Phalangium opilio*, *Cscu Centruroides sculpturatus*,
351 *Dsil Dysdera silvatica*, *Dpla Dolomedes plantarius*, *Lele Latrodectus elegans*, *Ptep Parasteatoda*
352 *tepidariorum*, *Hgra Hylyphantes graminicola*, *Abru Argiope bruennichi*, *Tant Trichonephila antipodiana*, *Tcla*
353 *Trichonephila clavata*.

	Dmel	*Tcas*	*Smar*	*Isca*	*Popi*	*Cscu*	*Dsil*	*Dpla*	*Lele*	*Ptep*	*Hgra*	*Abru*	*Tant*	*Tcla*
Homeobox genes	103	103	113	109	107	157	140	146	148	147	172	145	143	148
Homeobox families	83	85	90	88	90	90	79	86	88	91	86	87	86	88
Families with paralogues from ancient TD(s)	5	5	6	8	9	8	8	9	9	9	9	9	9	9
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	75	75	79	42	34	40	44	43	41	43	43	44
Families with retained ohnologues	n/a	n/a	n/a	n/a	n/a	44	41	43	41	43	37	40	40	40

354

In contrast to tandem duplicates, dispersed paralogues consistently found on different chromosomes in surveyed spider genomes were inferred to be ohnologues. There were between 37 and 44 families with retained ohnologues in spider species, and the families that were retained were largely consistent across species. All the Hox genes besides *Hox3* and *ftz* were retained as ohnologues, as previously reported for two spider species (Schwager et al. 2017), as well as a few Super-Hox genes, namely *Dbx*, *En*, and *Gbx* (Figure 2 A). Many of the NK cluster genes were also retained as ohnologues, including *NK4*, *NK5*, *NK6*, *NK7*, *Lbx*, *Tlx*, and tandem paralogues of *Emx* and *Msx* (Figure 2 A).

Ohnologues were likely lost from several families following the WGD (Figure 2 A). 32 families are found as single-copy in scorpions and spiders (discounting subsequent lineage-specific tandem duplications of single-copy genes), indicating an ohnologue was lost between the WGD and the divergence of scorpions and spiders. These include many of the Super-Hox families (*Ro*, *Nedx*, *Mnx*, *Evx*, and *Abox*), the ParaHox family *Gsx*, most of the NK2 cluster families (*Hlx*, *Msxlx*, and *NK2.1*) and the HRO

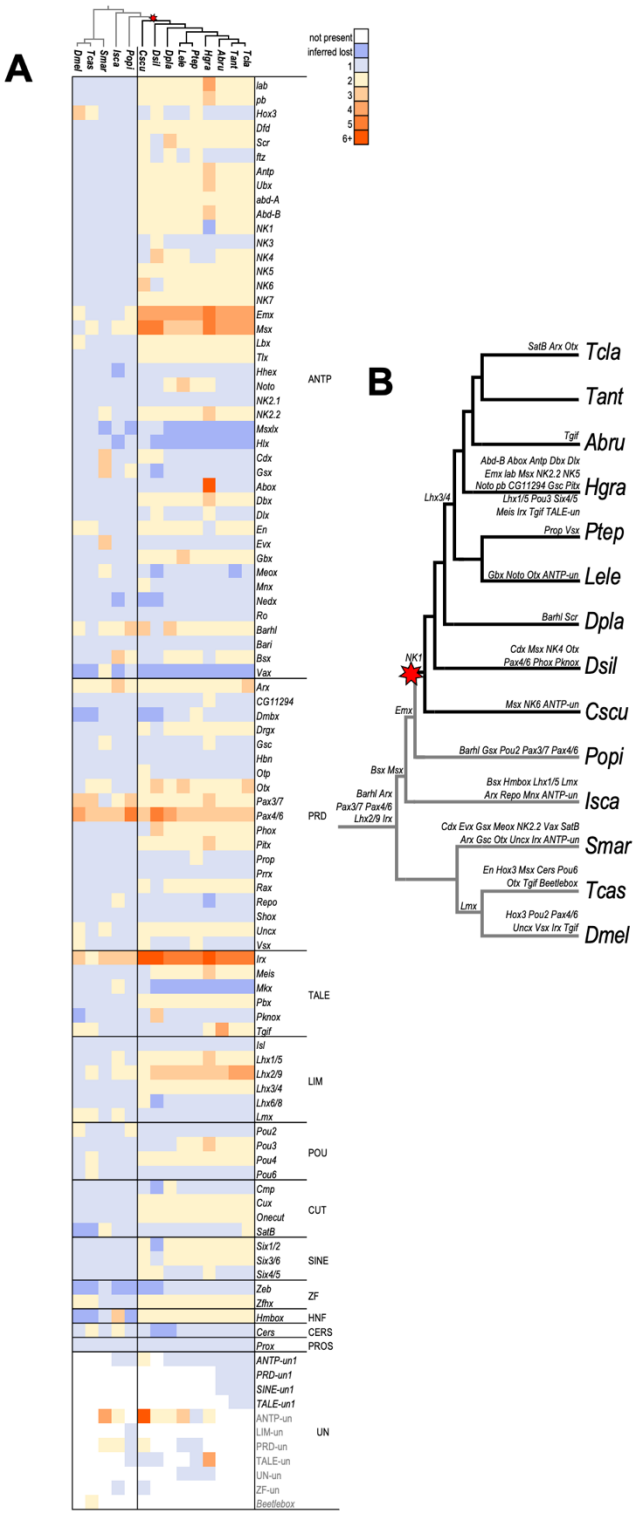

Figure 2: Many homeobox genes were retained in duplicate in scorpion and spider genomes. A) Heatmap of homeobox genes by family. B) Homeobox gene families having undergone tandem duplication events specific lineages of arthropods. *ftz* ohnologues in *D. plantarius* and *P. tepidariorum* are pseudogenes. Species abbreviations are as in Table 2.

386 cluster family *Hbn*. Six families including the ParaHox family *Cdx* and the HRO cluster
387 family *Otp* were found in the single copy in spiders while both ohnologues were retained
388 in the scorpion, indicating the loss of an ohnologue in the spider ancestor (Figure 2 A).
389 Within spiders, an ohnologue was lost from five families in the entelegyne ancestor: *ftz*,
390 *Dlx*, *Six4/5*, *Msxlx*, and *NK3*. *D. silvatica*, the sole synspermiatan representative, has
391 lost the largest number of ohnologues from eighteen families. Though the extent to
392 which this applies across synspermiatans cannot be determined, it is greater than the
393 twelve losses that have occurred across the entelegynes, or the eleven losses in *C.*
394 *sculpturatus*. Only a few families underwent losses multiple times, namely *Hox3* in *C.*
395 *sculpturatus*, *D. plantarius*, *L. elegans*, and the araneids; *Tgif* in the theridiids, *D.*
396 *plantarius*, *D. silvatica* and *C. sculpturatus*; and *Bsx* in both *H. graminicola* and *D.*
397 *silvatica*. Only one family, *Vax*, was potentially lost before the WGD, or both ohnologues
398 were lost in the arachnopulmonate ancestor (Figure 2 A). Though it was not reported in
399 the harvestman previously (Leite et al. 2018), we identified a single *Vax* gene in the
400 harvestman genome (Figure 2 A, Table S1).

401 *Homeobox gene clusters in spider genomes*

402 The Hox cluster

403 Each of the spider genomes here contains two Hox clusters (Figure 3), consistent with
404 previous descriptions for *P. tepidariorum* and other spiders (Schwager et al. 2007;
405 Schwager et al. 2017; Fan et al. 2021; Sheffer et al. 2021; Wang et al. 2022). With the
406 new *P. tepidariorum* chromosome-level assembly, it could be confirmed that each
407 cluster is on a single chromosome. Cluster B (Figure 3), is typically intact, organised,
408 and contains orthologues of all ten arthropod Hox genes, while cluster A often lacks
409 *Hox3* and *ftz*, and is sometimes rearranged. There are some exceptions to this,
410 however. For one, there have been several tandem duplications of *H. graminicola* Hox
411 genes resulting in four *lab* paralogues, and three *pb*, *Antp*, *Ubx*, and *Abd-B* paralogues
412 (Figure 3). While these duplications have occurred in both Hox clusters, cluster B is
413 ordered, and contains the single orthologues of *Hox3* and *ftz*, while cluster A lacks
414 those genes and has been rearranged so that *lab-A-1*, *lab-A-2*, *pb-A-1*, *pb-A-2*, and

415 *Dfd-A* lie to the right of *Scr-A* through *abd-A-2* (Figure 3). Furthermore, the Super-Hox
416 gene *Dlx-A* is between *pb-A-2* and *Dfd-A* separated by 124 kb from *pb-A-2* and 380 kb
417 from *Dlx-A* and *Dfd-A*. In *D. plantarius* there is a new gene, identified with highest
418 similarity in the homeodomain to *Scr*, but not linked to either Hox cluster, possibly
419 representing a divergent *Scr* paralogue (blue *'S'* in Figure 3). A gene with similarity to
420 *Dbx* (grey *'Db'* in Figure 3) is also found between *pb-A* and *Dfd-A*, in the same relative
421 position as *Dlx* in the *H. graminicola* Hox cluster A. In *D. silvatica*, both Hox clusters are
422 rearranged, though one does contain both *Hox3* and *ftz*, while the other has lost only *ftz*
423 (Figure 3). Pseudogenes of *ftz* were found in *D. plantarius* and *P. tepidariorum*,
424 identified by the homeodomain disrupted by a stop codon. In *P. tepidariorum*, the *ftz*-
425 pseudogene is not linked to either Hox cluster, while in *D. plantarius,* it is adjacent to *ftz*-
426 B (Table S1, Figure S3 F).

427 Generally, the more degenerated Hox cluster, cluster A, contains more intervening
428 genes, but this greatly depends on the quality of annotation. In several genomes, Hox
429 genes encompass multiple predicted gene models, even for genomes with annotations
430 generated using transcriptomic data, resulting in an inflation of the estimation of the
431 number of intervening genes. Some estimates count as many as 98 protein coding
432 genes interspersed within the Hox clusters in *P. tepidariorum* (Turetzek et al. 2022),
433 which may be inaccurate both because of the fragmentation of gene models, as well as
434 the scaffold break disrupting Hox cluster B in the previous assembly. The triangle icons
435 in Figure 2 reflect the presence of any intervening genes, rather than the number, to
436 avoid overestimation due to fragmented gene models.

437 The Super-Hox genes consist of the Hox genes along with several ANTP-class genes
438 consistently found linked to each other and/or the Hox cluster across bilaterians (Butts
439 et al. 2008; Ferrier 2016b). While the Super-Hox complement is fairly intact in the
440 single-copy genomes studied here, with many Super-Hox genes still linked to one
441 another and/or to the Hox cluster in the tick and insects, Super-Hox genes have
442 dispersed in spider genomes (Figure 3), and many Super-Hox families have also
443 returned to the single copy (Figure 2 A, Figure 3). Amid the disintegration of the Super-
444 Hox cluster, there are still some associations between certain genes that indicate

conserved clustering and possibly functional constraint. *Abox* and *Meox* are commonly found linked or clustered in all the entelegyne species (Figure 3). Other apparent linkage patterns likely rather reflect the passive maintenance of linkage over time, like the linkage of *Dbx* and *Gbx* to Hox cluster B in *H. graminicola*, *P. tepidariorum*, and *D. plantarius*. Due to the >1 Mb distances between these genes and the Hox cluster, it is unlikely that there is a functional aspect to their association. These passive remnants of Super-Hox linkage also include the linkage of *En* with *Mnx*, *Gbx* with *Nedx*, and *Dbx*

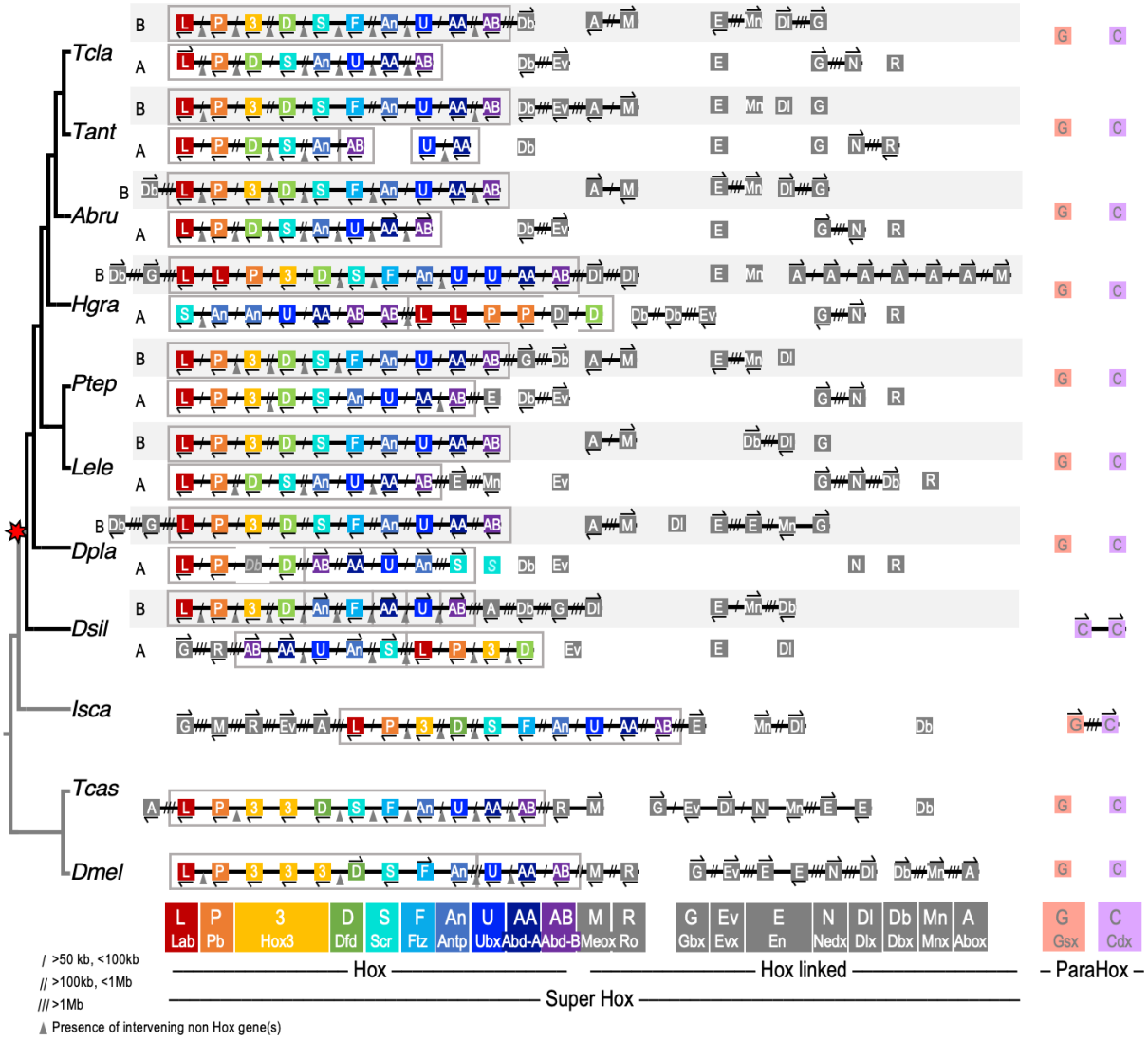

Figure 3: Two ohnologous Hox clusters and a disintegrated arthropod ParaHox cluster are found in spiders. Genes are represented by coloured boxes; chromosomes are black lines. Gene orientation is indicated by arrows above or below genes, and distances are indicated by slashes between genes. Gene families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

452 with *Evx*, though these have more frequently been disrupted and dispersed, unlike *Abox*
453 and *Meox*. Thus, while the Hox cluster is maintained intact in both copies, most of the
454 Super-Hox genes have largely dispersed and their linkage arrangements have broken
455 down following the WGD.

456 The ParaHox cluster has largely disintegrated in spiders. Neither ParaHox gene is
457 linked to the other, and both have returned to the single copy (Figure 3). The two *Cdx*
458 paralogues in *D. silvatica* likely arose in a recent lineage-specific tandem duplication, as
459 they are found in a two-gene cluster only in this species. In both insects, *Gsx* and *Cdx*
460 family genes are unlinked, while in the tick they are on the same chromosome but
461 separated by a large distance (28 Mb) (Figure 3). This is consistent with the ancestral
462 breakdown of the ParaHox clustering in an arthropod ancestor, and the further
463 separation of the genes onto separate chromosomes in insects and spiders (Figure 3).
464 It is not clear whether these genes belong to the same ohnology group or not (*i.e.*, both
465 A or both B versus one A and one B), as there is only one ohnologue retained in any of
466 the species surveyed.

467 The NK cluster

468 We found two ohnologous NK clusters in all spider genomes (Figure 4). One consists of
469 *Msx1-A*, *NK4-A*, *NK3-A*, *Tlx-A*, *Lbx-A*, and *NK7-A* clustered on one chromosome, while
470 *Msx1-B*, *Tlx-B*, *Lbx-B*, and *NK7-B* are found on a separate chromosome, with *NK3-B*
471 and *NK4-B* in the species that have retained these genes (core NK clusters are
472 delineated with grey boxes in Figure 4). Cluster B is less organised; *NK3-B* has likely
473 been lost in entelegynes and *NK4-B* has likely been lost in *H. graminicola* and *P.*
474 *tepidariorum*. NK core cluster A is conserved across spiders, except that *Msx1-A* is not
475 found in *L. elegans* or *P. tepidariorum*, which could indicate the loss of *Msx1-A* from NK
476 cluster A in a theridiid ancestor. While NK cluster A is found in a conserved complement
477 and order in almost all spider species analysed, it is rarely linked to any other genes
478 belonging to the larger ancestral NK complement. In contrast, NK cluster B has a
479 varying gene complement, having undergone lineage-specific gene losses and more
480 rearrangements, making it more degenerate and less conserved between species. This

481 second cluster is, however, often linked to other ancestral NK cluster genes *Hhex*, *Noto*,
482 *Emx1-B*, *Emx2-B*, *NK5-B*, and *NK1*, which has undergone a tandem duplication, and to
483 *NK6-B* and *Msx2-B* in all spiders (Figure 4). *NK6-A* is found linked to *Msx2-A* in all the
484 entelegyne species, suggesting this two-gene *NK6-Msx2* cluster predates the WGD and
485 was retained in duplicate some species, while others have lost either *Msx2-A* or *NK6-A*.

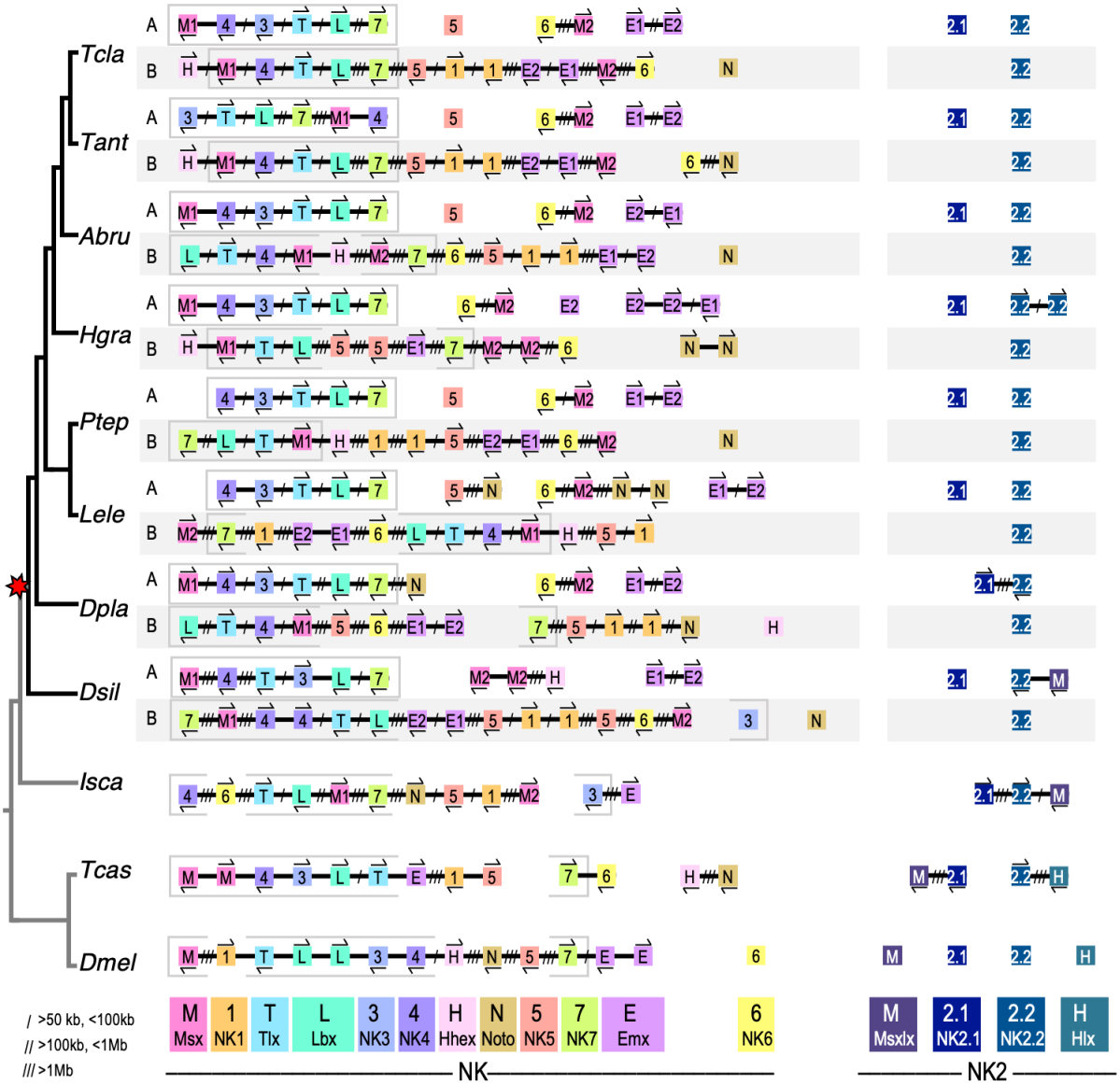

Figure 4: Two ohnologous NK clusters are retained, and a disintegrated NK2 cluster is found in spiders. Genes are represented by coloured boxes; chromosomes are black lines. Gene orientation is indicated by arrows above or below genes, and distances are indicated by slashes between genes. Gene families 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.

486 The NK2 cluster has largely broken down in spiders, with successive gene losses, first
487 *Hlx* from chelicerates, and then *MsxIx* from entelegynes, though further lineage
488 sampling may change this pattern (Figure 4). In all spiders, *NK2.2* is retained in
489 duplicate on two separate chromosomes, indicating an ohnologous relationship, while
490 *NK2.1* has returned to the single copy (Figure 4). *NK2.1* and *NK2.2* are unlinked in all
491 spiders, except for *D. plantarius*, indicating the two genes have separated twice.
492 Alternatively, different ohnologues could have been lost in different lineages. If the
493 ancestral linkage between *NK2.1* and *NK2.2* was intact following the WGD, either one
494 ohnologue was lost from each of the NK2 chromosomes, resulting in unlinked genes, or
495 both were lost from one of the ohnologous chromosomes and a rearrangement
496 separated the *NK2.1* and *NK2.2*. Either scenario would have to occur multiple times in
497 the lineages leading to *D. silvatica* and the Araneoidea, but not in *D. plantarius*, unless
498 the linkage between *NK2.1* and *NK2.2* in *D. plantarius* occurred secondarily.
499 Phylogenetics were not able to determine which ohnologue of *NK2.1* was retained in
500 different species because no genome here contains both ohnologues.

501 The HRO cluster

502 The PRD-class HRO cluster was so named because the order of the genes in the fly
503 genome is *hbn* (*Hbn*), *Rx* (*Rax*), then *otp* (*Otp*); this gene order is conserved in insects
504 and the tick genome, indicating it is likely the ancestral arthropod arrangement (Figure
505 5). In spiders, the HRO cluster is found intact, but reordered: *Rax*, *Hbn*, then *Otp*
506 (Figure 5). There have also been several inversions of the orientations of the genes
507 within the cluster, but the R-H-O order is conserved across spiders, suggesting this
508 rearrangement occurred once ancestrally. It likely did not predate the WGD but
509 occurred in a spider ancestor because the scorpion HRO cluster A order is *Hbn*, *Rax*,
510 then *Otp*, as in insects and the tick (Table S1). Also consistently found linked to the
511 HRO cluster is the LIM-class gene *Isl* (Figure 5). Ferrier (2016) suggested this pattern is
512 evidence for the ancestral linkage between the LIM-class and PRD-class progenitor
513 genes in the ancestral Giga-cluster as it has been observed across species with a
514 common ancestor in the ancestor of all animals (see Introduction). The consistent
515 linkage pattern between these genes in spiders, the tick, and beetle genomes supports

516 this association. This broad phylogenetic distribution makes this between-class linkage
517 unlikely to be a convergent secondary association; rather it is more likely to be a
518 conserved ancestral trait.

519 The Irx cluster

520 *Irx* genes have duplicated independently many times, giving rise to six genes in spiders.
521 Each of the ohnologous Irx
522 clusters contains one of each
523 gene type, *Irx1*, *Irx2*, and
524 *Irx3*, though *Irx2-B* has been
525 lost from one of the clusters
526 in the entelegynes (Figure 5).
527 Consistent with the findings
528 of Kerner et al. (2009), insect
529 *Irx* genes fall into two clades,
530 one containing orthologues of
531 *D. melanogaster mirr* and the
532 other with orthologues of *ara*
533 and *caup* (Figure S1). Spider
534 and scorpion *Irx2-A* and *Irx2-*
535 *B* group with *mirr*, while
536 spider and scorpion *Irx1-A*
537 and *Irx1-B* group with *ara* and
538 *caup* (Figure S1). The exact
539 relationships between *Irx3*
540 and other arthropod genes
541 remains somewhat unclear.
542 There are three *Irx* genes in
543 the centipede *S. maritima* and
544 the millipede *Trigoniulus*
545 *corallinus* genomes (Chipman

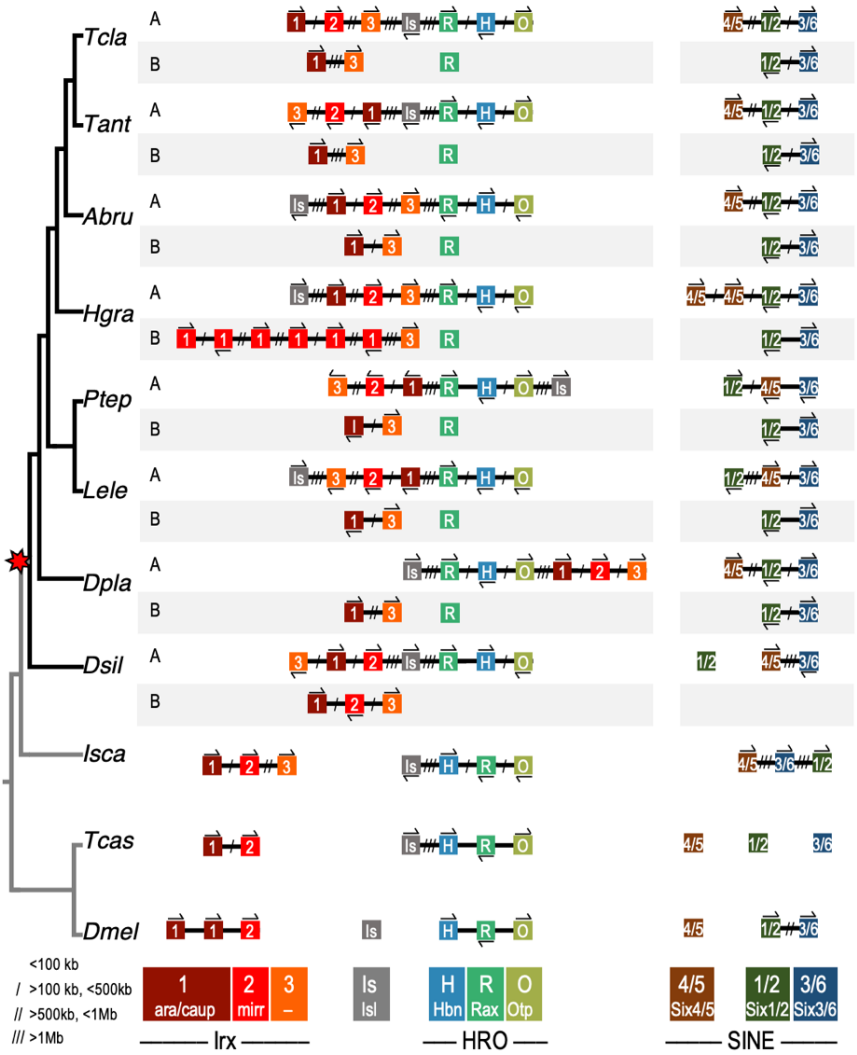

Figure 5: Two ohnologous HRO clusters, one of which is linked to one of the two ohnologous *Irx* clusters, and two ohnologous SINE clusters are retained in spiders. Genes are represented by coloured boxes; chromosomes are black lines. Gene orientation is indicated by arrows above or below genes, and 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.

546 et al. 2014; So et al. 2022), one of which is not contained within either the *Irx1/ara/caup*
547 or *Irx2/mirr* clade, and groups sister to the *Irx3* clade, albeit with very low bootstrap
548 support (Figure S1). The potential *Irx3* is clustered with *Irx1* and *Irx2* in the *S. maritima*
549 genome in the same relative position as spider *Irx3* (Figures S3 C and 5). There is a
550 clear orthologue of each of the three arachnopulmonate *Irx* gene families in both the
551 harvestman and tick genomes, indicating the three gene types originated in or before
552 the chelicerate ancestor, and possibly *Irx3* genes were lost from an insect ancestor after
553 the divergence of Myriapoda. According to Kerner et al. (2009), *Irx* genes duplicated
554 independently in spiralians, ecdysozoans, and deuterostomes, suggesting that all
555 paralogues arose from a single ancestral bilaterian *Irx* gene. From our phylogeny
556 (Figure S1), we can tentatively infer that the duplication that gave rise to fly paralogues
557 is ancestral to the arthropods, and that there was a three-gene cluster in the ancestral
558 chelicerate that further duplicated in the arachnopulmonate WGD resulting in spider *Irx*
559 paralogues in two ohnologous clusters.

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

568 *Conserved macrosynteny of spider genomes*

569 To contextualise our observations of homeobox cluster conservation and
570 rearrangements, we explored macrosynteny across spider chromosomes, using *H.*
571 *graminicola* as the reference (Figure 6). Across entelegynes, we detected clear patterns
572 of orthology between chromosomes, and consistency in the synteny blocks in which
573 certain homeobox clusters are found, but these patterns were not as evident when

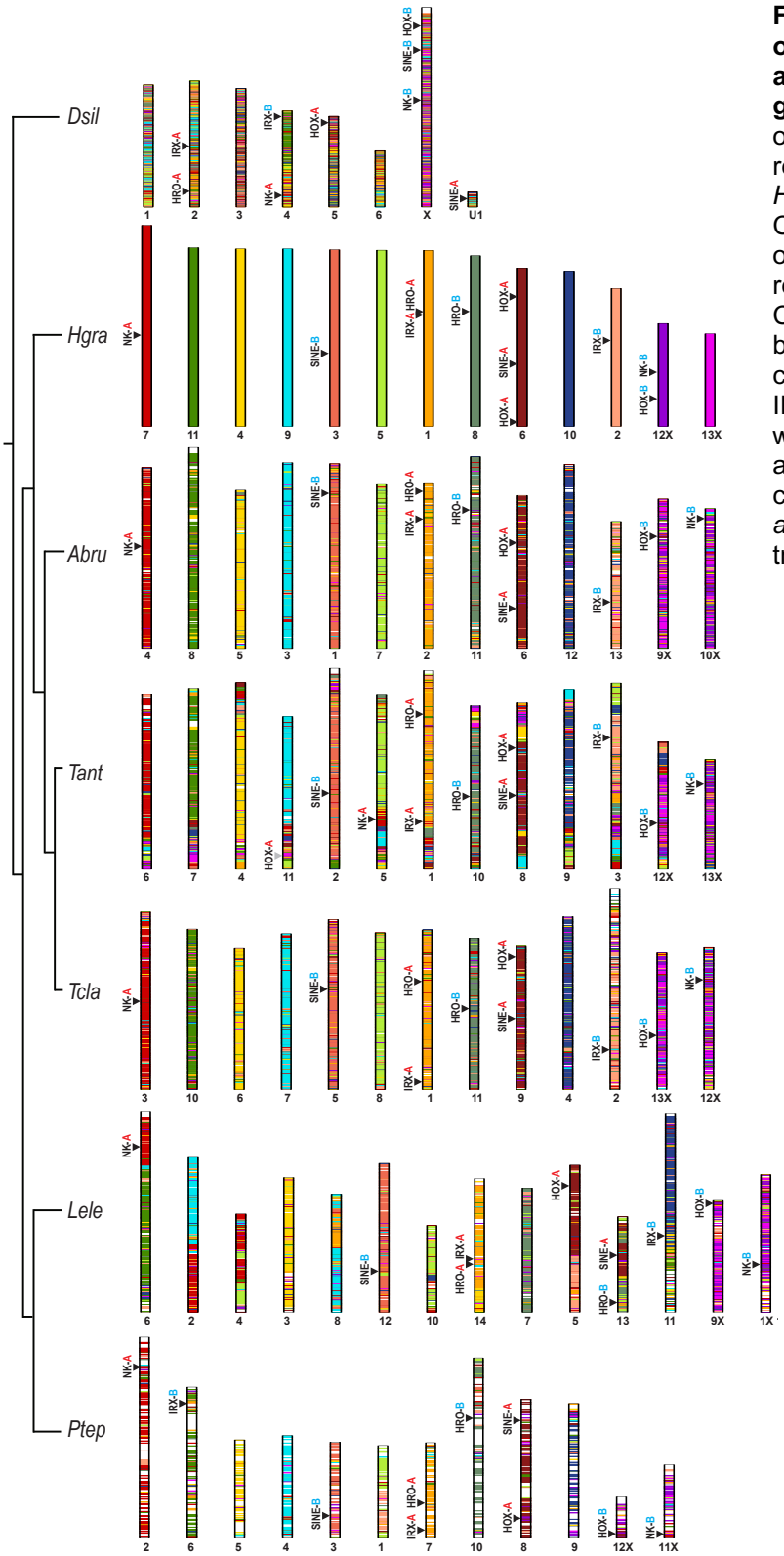

Figure 6 – Conserved macrosynteny of spider chromosomes and associated location of homeobox gene clusters. Vertical bars show stacks of genes on each chromosome representing a macrosynteny block, with *H. graminicola* as the reference genome. Chromosomes of other species were ordered to show the orthologous 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.

compared to the sole synspermiatan representative *D. silvatica* (Figure 6). Depictions of locations of all homeobox genes across genomes can be found in Figure S3.

In each araneid analysed here, there are 13 chromosomes, however there are only 12 in *P. tepidariorum* and 14 in *L. elegans* (Figure 6). In *P. tepidariorum*, the synteny block corresponding to

591 chromosome 2 in *H. graminicola* appears to have been mostly lost or distributed
592 throughout the genome, although Irx cluster B which belongs to this block has been
593 relocated (Figure 6). In *L. elegans* there appear to have been large fission and fusion
594 events on chromosomes 2, 4, 5, 6,8, 11, and 13. Such events in *L. elegans* are
595 consistent with Hox cluster A and SINE cluster A being found on separate
596 chromosomes, as well as the rearrangement of Irx cluster B (Figures 6 and S3 G).

597 Sex chromosomes have been identified in multiple spider species (Sheffer et al. 2021;
598 Zhu et al. 2022; Miller et al. 2023), which are denoted with an 'X' in Figure 6.
599 Interestingly, Hox cluster B and NK cluster B are both found on the same sex
600 chromosome in *H. graminicola* (12) and *D. silvatica* (X), but on separate sex
601 chromosomes in other entelegynes, whereas Hox cluster A is located on an autosome
602 (Figure 6). Across entelegynes, NK cluster A is found in the same conserved synteny
603 block corresponding to a full chromosome in all species except for *T. antipodiana*. In
604 this species, a small rearrangement brought NK cluster A onto another chromosome,
605 and this is the only entelegyne species in which the NK cluster A is rearranged (Figure
606 4). Overall, this analysis illustrates the conservation of macrosynteny blocks and
607 chromosomes across entelegyne genomes and shows several chromosomal
608 rearrangements affecting homeobox gene clusters.

609 *Analysis of selection on NK cluster genes*

610 We next sought to determine whether single copy orthologues, ancient tandem
611 paralogues, or WGD ohnologues of NK cluster genes are under different selective
612 pressures. To do this we compared the aBSREL estimated ω values across
613 phylogenies for seven families with different inferred duplication histories: *Msx*, *NK4*,
614 *NK3*, *Tlx*, *Lbx*, *NK7*, and *Emx* (aBSREL output JSON files available on Github).

615 While none of the focal ohnologues were found to be under positive selection, certain
616 genes from the pre-duplicate outgroups were (Supplementary material on GitHub
617 repository, see Materials and Methods). *NK7* was found to be under positive selection in
618 *D. melanogaster* and *I. scapularis*, *Tlx* also in *I. scapularis*, and *NK3* in *T. castaneum*.

619 **Table 3 – Synonymous and nonsynonymous substitution rates between *P. tepidariorum* and *P.***
620 ***opilio* NK cluster genes calculated with codeML**. The relationship between each gene pair is indicated
621 in the third column, where 1 to 1 indicates direct orthology (the loss of an ohnologue), 1 to 2 indicates the
622 retention of two ohnologues in the spider, O is comparison between the two spider ohnologues, and T is
623 the comparison between tandem paralogues. Gene names have been updated to reflect orthology between
624 paralogues in *P. opilio* and subsequent ohnologues in *P. tepidariorum*, such that, e.g., *P. opilio Emx1* is the
625 pro-orthologue of *P. tepidariorum Emx1-A* and *Emx1-B*, and the same goes for *P. opilio Emx2* and *P.*
626 *tepidariorum Emx2-A* and *Emx2-B*.

Popi	*Ptep*		S	N	dS	dN	dN/dS
NK4	*NK4*	1 to 1	184.1	574.9	7.8104	0.5245	0.0673
NK3	*NK3*	1 to 1	137.5	387.5	5.2370	0.3058	0.0584
Tlx	*Tlx-B*	1 to 2	209.7	591.3	48.8193	0.2948	0.0060
	Tlx-A	1 to 2	291.1	761.9	42.6118	0.3399	0.0094
	Tlx-A x *Tlx-B*	O	203.9	603.1	50.8092	0.2147	0.0042
Lbx	*Lbx-B*	1 to 2	136.5	433.5	4.8815	0.2053	0.0421
	Lbx-A	1 to 2	201.8	635.2	7.6580	0.2289	0.0299
	Lbx-A x *Lbx-B*	O	137.4	489.6	33.8338	0.2574	0.0076
NK7	*NK7-B*	1 to 2	223.8	739.2	12.2032	0.7833	0.0642
	NK7-A	1 to 2	307.5	835.5	25.5859	1.1008	0.0430
	NK7-A x *NK7-B*	O	286.8	1012	12.8423	0.7380	0.0575
Msx1	*Msx1-B*	1 to 1	211.1	595.9	48.2788	0.5018	0.0104
Msx2	*Msx2-A*	1 to 2	187.8	487.2	45.3891	0.5191	0.0011
	Msx2-B	1 to 2	156.9	398.1	44.7737	0.4745	0.0106
	Msx2-A x *Msx2-B*	O	145.2	457.8	53.1546	0.2699	0.0051
	Msx1-B x *Msx2-B*	T	135.8	461.2	4.4066	0.3723	0.0845
Emx1	*Emx1-A*	1 to 2	204.7	653.3	41.5860	0.5123	0.0123
	Emx1-B	1 to 2	184.5	613.5	53.9284	0.6923	0.0128
Emx2	*Emx2-A*	1 to 2	207.8	650.2	51.7318	0.6215	0.0120
	Emx2-B	1 to 2	249.6	686.4	47.5150	0.4485	0.0094
	Emx1-A x *Emx1-B*	O	176.6	636.4	5.3808	0.7658	0.1423
	Emx2-A x *Emx2-B*	O	190.4	670.6	56.4343	0.6652	0.0118
	Emx1-A x *Emx2-A*	T	196.9	673.1	28.1173	0.5279	0.0188
	Emx1-B x *Emx2-B*	T	149.2	621.8	63.8842	0.7941	0.0124

627 However, certain ohnologues did have longer branches separating them from other
628 ohnologues or the single copy outgroup genes in the aBSREL-generated phylogenies,
629 indicating a higher substitution rate. These include *Msx1* in arachnopulmonates and the
630 harvestman, *NK3-A* in the entelegynes, and *Lbx-B* in the scorpion. No consistent
631 pattern was detected between A versus B ohnologues, despite the A cluster being more
632 highly conserved in organisation, nor was there a clear pattern between ohnologues
633 versus tandem duplicates. In the one-to-one comparisons between *P. tepidariorum* and
634 *P. opilio* genes, and between *P. tepidariorum* ohnologues, the dN/dS ratio varied from
635 0.0011 (between *P. opilio Msx2* and *P. tepidariorum Msx2-A*) to 0.1423 (between *P.*
636 *tepidariorum Emx1-A* and *Emx1-B*) (Table 3). There are some patterns among gene
637 families, with low dN/dS ratios for *Tlx* ohnologues compared to a consistently higher
638 rate for *NK7* ohnologues. The *Emx1* and *Emx2* and *Msx1* and *Msx2* ohnologues had

639 consistently low dN/dS values when compared to their *P. opilio* orthologues but varied
640 when compared between ohnologues in *P. tepidariorum*. Overall, these low dN/dS
641 ratios likely reflect the highly conserved nature of homeobox genes.

642 *Developmental expression of NK cluster genes in P. tepidariorum*

643 It was previously shown that several ohnologues of spider homeobox genes had been
644 subject to sub- and/or neofunctionalisation with respect to their single copy orthologues
645 in the harvestman (Leite et al. 2018). To better understand the roles of homeobox
646 genes better following WGD in the arachnopulmonate ancestor we assayed the
647 expression of NK cluster genes in *P. tepidariorum* in comparison to *P. opilio*.

648 Single-copy genes

649 Both *NK3* and *NK4* were found in single copies in both *P. opilio* and *P. tepidariorum*,
650 indicating the loss of an ohnologue of each from the spider following the WGD. The
651 single identified orthologue of *NK3* in the spider is strongly expressed in the developing
652 heart and is weakly expressed around the stomodaeum and in the head (Figure 7 A-E).
653 This is very similar to the single spider orthologue of *NK4*, which is also expressed
654 strongly in the heart (Figure 7 F-J). Unlike *NK3*, *NK4* is also expressed in a dot-like
655 fashion in the ventral nervous system of the opisthosoma, and at the anterior rim of the
656 head lobes (Figure 7 asterisks and double arrowheads in Figure 7 F-I). There is no *NK3*
657 expression in the developing heart of the harvestman; instead *NK3* is expressed in a
658 segmental fashion ventral to the base of the appendages and in similar short transverse
659 segmental stripes in the opisthosomal segments (Figure 7 K-N). The expression of
660 harvestman *NK3* is thus more like the expression of its orthologue in *D. melanogaster*
661 (Azpiazu and Frasch 1993) and the millipede *Glomeris marginata* (Figure S4) than that
662 of the spider. Detectable expression of *NK4* in the harvestman is restricted to the
663 developing dorsal vessel at late developmental stages (Figure 7 O-P), a pattern that is
664 conserved among panarthropods (Azpiazu and Frasch 1993; Janssen and Damen
665 2008; Treffkorn et al. 2018).

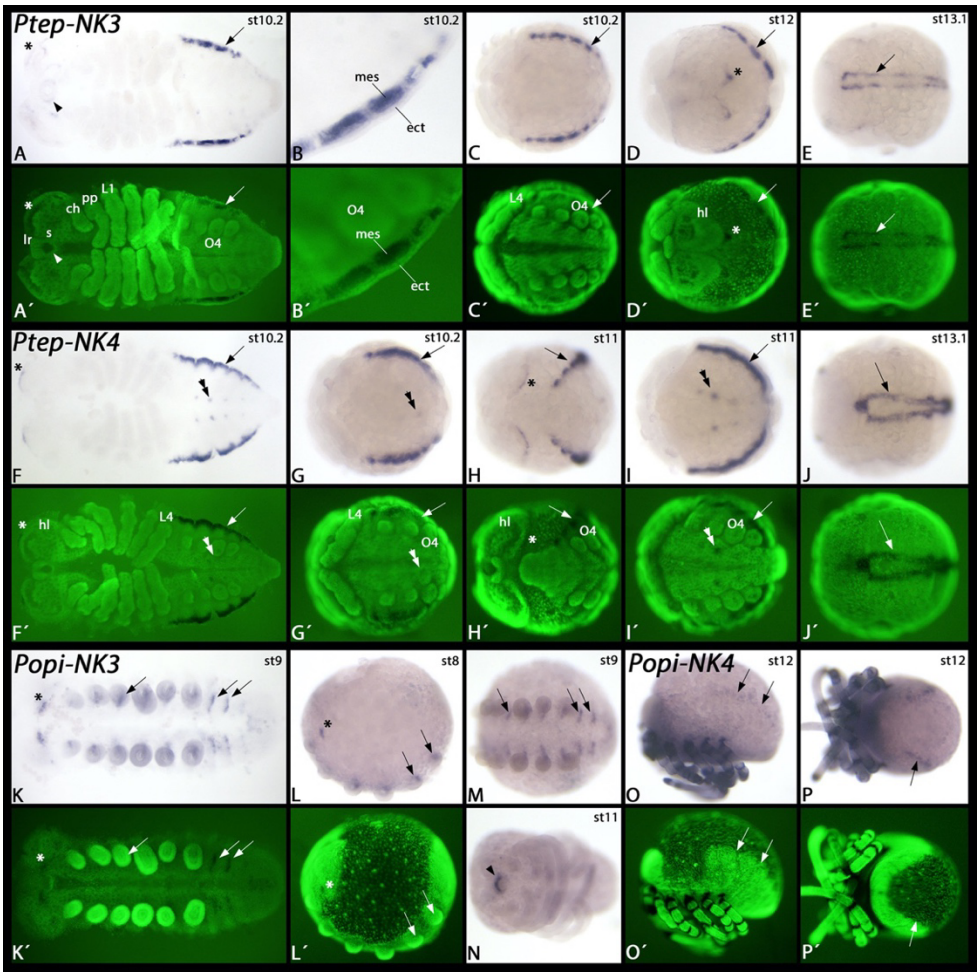

Figure 7 – Expression of *P. tepidariorum* and *P. opilio NK3* and *NK4*. In all panels, anterior is to the left. Panels A-C, F, G, I, K, M, and N show ventral views. Panels D and H show anterior-dorsal views. Panels E, and J show dorsal views. Panels L and O show lateral views and dorsal up. Panel P 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 photograph. 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.

Ancient tandem paralogues

Both *Emx* and *Msx* families appear to have undergone an ancient tandem duplication, resulting in two paralogues of each gene in the harvestman genome, with four *Emx* and three *Msx* paralogues retained in *P. tepidariorum*. Leite et al. (2018) described *Emx1-A*

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

677 Very similar to its paralogue *Emx2*, *P. opilio Emx1* is expressed in all posterior
678 segments including the cheliceral segment (Figure 8). The sharp anterior border of
679 expression is conserved among other arthropods, including spiders (Walldorf and
680 Gehring 1992; Simonnet et al. 2006; Schinko et al. 2008; Birkan et al. 2011; Janssen
681 2017; Leite et al. 2018) and an onychophoran (Janssen 2017). Indeed, the overall
682 expression pattern of *P. opilio Emx1* is identical with that of the previously reported
683 *Emx2*, with the exception that expression of *Emx1* starts earlier than that of *Emx2*.

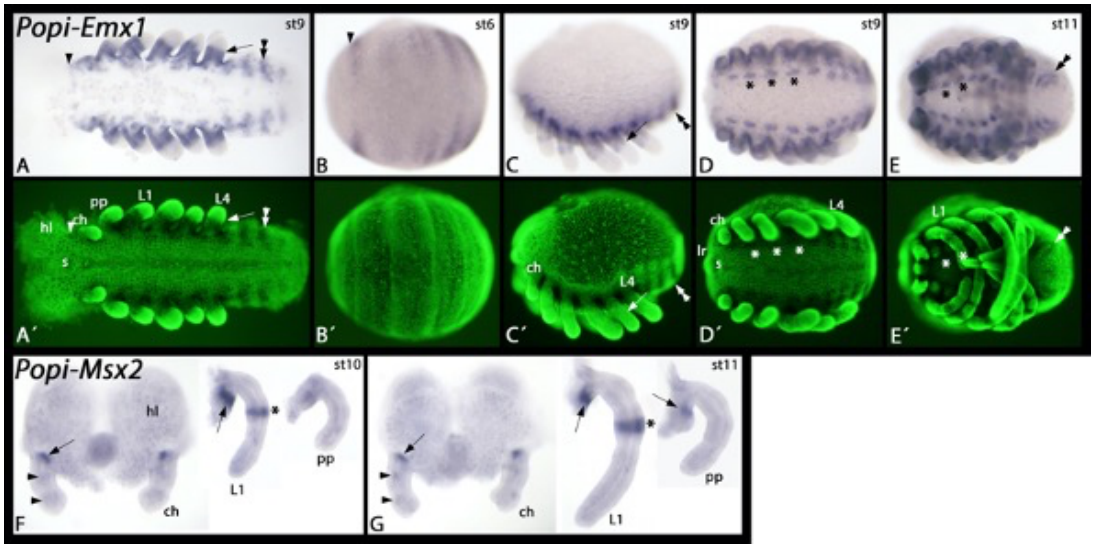

Figure 8 – Expression of *P. opilio Emx1* and *Msx2*. In panels A-E, anterior is to the left and ventral views (except panel C, lateral view and dorsal up). Panels F and G each show dissected head lobes + chelicerae, a walking leg and a pedipalp. The dissected heads are shown with anterior up, ventral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

684 According to Leite et al. (2018), *P. tepidariorum Msx1-B* (previously designated *Msx3*) is
685 expressed in the base of the prosomal appendages at stage 11, while *Msx2-B*
686 (previously designated *Msx2*) is expressed in the chelicerae at stage 12, and *Msx2-A*
687 (previously designated *Msx1*) is expressed in each segment along the ventral midline in
688 stages 8 and 10. In *P. opilio*, *Msx1* (previously designated *Msx*) was detected in a stripe
689 in each segment at stage 7 and along the ventral midline in each segment at stage 13
690 (Leite et al. 2018). However, we detected *P. opilio Msx2* expression only at late
691 developmental stages and in a very restricted dot-like patch at the base of the
692 appendages, and a ring-like domain in the legs (Figure 8 F-G). Thus, *P. tepidariorum*
693 *Msx2-A* expression is similar to *P. opilio Msx1* in the segments and along the ventral
694 midline, while *Msx1-B* is similar to *P. opilio Msx2*, expressed in the base of appendages.
695 *P. opilio Msx2* expression was also detected in a small region of the chelicerae,
696 matching the expression of *Msx2-B* observed by Leite et al. (2018) in stage 12 *P.*
697 *tepidariorum*. This means that the inferred neofunctionalisation is more likely a case of
698 subfunctionalisation between the tandem *Msx* paralogues, which was not detected
699 previously due to the absence of *Msx2* in the transcriptome-based survey.

700 Ohnologues

701 Three other NK families, *NK7*, *Tlx*, and *Lbx*, are found in the single copy in *P. opilio* and
702 have two retained ohnologues in *P. tepidariorum* (Figure 4). The two ohnologues of *P.*
703 *tepidariorum Tlx* are both expressed in the distal region of the legs, but in
704 complementary patterns (Figure 9 A-M). *Tlx-A* is first expressed in the complete distal
705 region of the appendages and restricts into two rings of stronger expression around
706 stage 10 (Figure 9 G-H). *Tlx-B* expression only begins after this point and is strongest in
707 the region between the two rings of *Tlx-A* expression (Figure 9 M). Apart from its
708 expression in the developing appendages, *Tlx-A* is also expressed in the dorsal tissue
709 of the opisthosoma in a pattern that suggests a role in heart development (Figure 9 B,
710 D-E), and at later stages, in the ventral opisthosoma (Figure 9 F). In contrast, *Tlx-B* is
711 only expressed in the opisthosoma at the very posterior at late developmental stages,
712 after segment addition is complete (Figure 9 J). Unlike *Tlx-A*, *Tlx-B* is expressed in the
713 form of a transverse stripe in the developing head lobe at late developmental stages

714 (double arrowheads in Figure 9 K). This is at the anterior edge of non-neurogenic

715 ectoderm that overgrows the neurogenic ectoderm of the head lobe from the anterior.

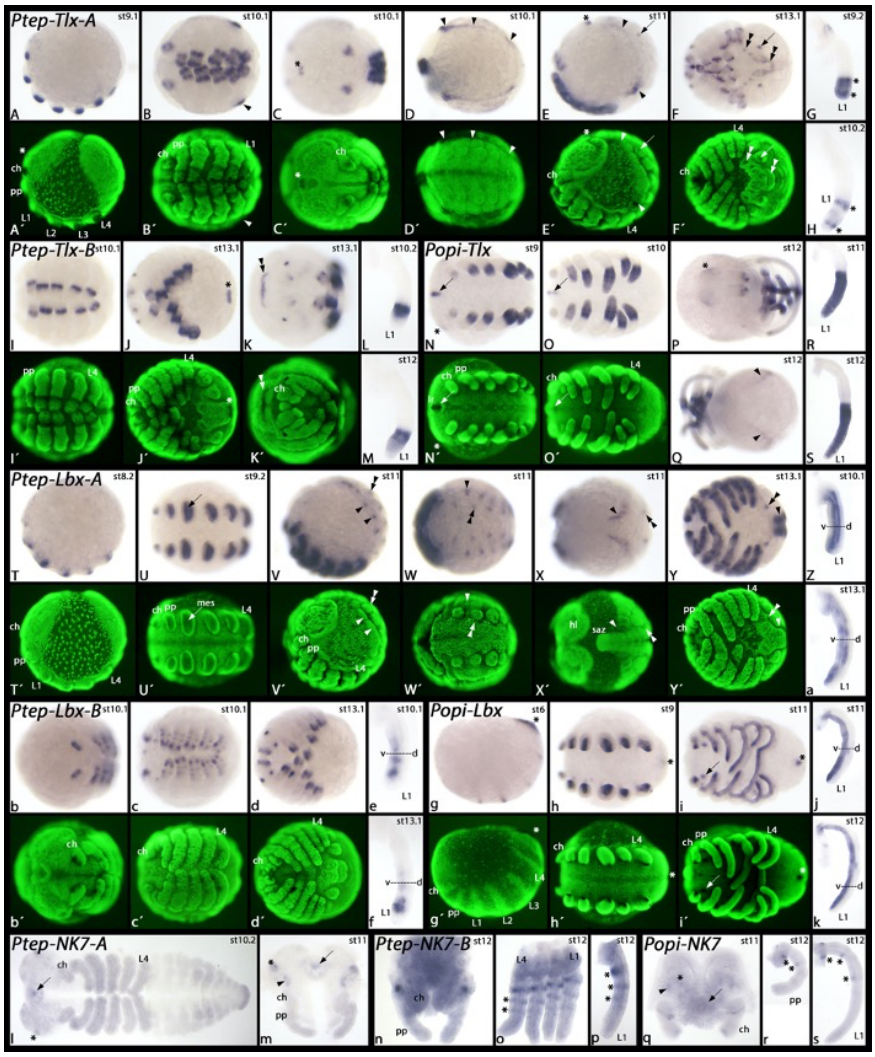

Figure 9 – Expression of *P. tepidariorum* and *P. opilio Tlx*, *Lbx* and *NK7* ohnologues. In all panels showing whole embryos, anterior is to the left. Dissected heads in panels m, n, and q, are anterior up. Dissected appendages are shown by lateral views, dorsal to the right. Panels A, E, T, V, and 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 photograph. 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 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.

716 The single *Tlx* gene of the harvestman shares many aspects with one or both spider
717 ohnologues. *P. opilio Tlx* is expressed in the distal half of the developing appendages
718 (Figure 9 R-S), reflecting the combined expression of both spider *Tlx* ohnologues
719 (Figure 9 G-H, L-M). Like *P. tepidariorum Tlx-A*, *P. opilio Tlx* is expressed in dorsal
720 tissue of the opisthosoma and in the labrum (Figure 9 N-O), and like *P. tepidariorum*
721 *Tlx-B*, *P. opilio Tlx* is expressed at the anterior edges of the head lobes (Figure 9 Q).
722 The single *P. opilio Tlx* gene thus combines aspects of expression that are either found
723 in *P. tepidariorum Tlx-A* or *Tlx-B*, suggesting a case of subfunctionalisation in the
724 spider.

725 The first expression of *P. tepidariorum Lbx-A* appears in the mesoderm of the
726 outgrowing prosomal limb buds (Figure 9 T-U). Later, expression in the appendages is
727 also in the ventral ectoderm, but the dorsal ectoderm remains free of expression (Figure
728 9 W-Y). Additional expression of *Lbx-A* is found in the dorsal ectoderm of the
729 opisthosoma and in the form of segmental patches in the opisthosoma (Figure 9 V-Y).
730 This latter expression is possibly associated with the ventral-most tissue of the
731 opisthosomal limb buds. Compared to *Lbx-A*, expression of *P. tepidariorum Lbx-B*
732 appears later and is restricted to the prosomal appendages (Figure 9 b-d). The pattern
733 of expression in these appendages is similar, but not identical, to that of *Lbx-A* (Figure 9
734 e-f). *Lbx-B* is expressed in fewer ventral cells of the developing appendages than *Lbx-A*.
735 Expression of the single *P. opilio Lbx* gene is similar to that of *P. tepidariorum Lbx-A*. It
736 appears early during development in the mesoderm of the prosomal appendages
737 (Figure 9 g-h), and later extends into ventral ectodermal cells of the limbs (Figure 9 j-k).
738 Unlike the two *P. tepidariorum Lbx* ohnologues, *P. opilio Lbx* is also expressed at the
739 posterior of the developing embryo (asterisks in Figure 9 h-i).

740 Expression of both ohnologues of *P. tepidariorum NK7* are restricted to relatively late
741 stages of embryogenesis. *NK7-A* is exclusively expressed in the stomodaeum and in a
742 few patches in the head lobes (Figure 9 l-m). *NK7-B* is expressed in the form of three
743 rings in the developing walking legs (Figure 9 n-p). The expression of *P. opilio NK7*
744 combines the expression of both spider *NK7* genes as it is expressed in the head lobes
745 and in three ventral domains in the legs (Figure 9 q-s).

746 **DISCUSSION**

747 *Evolution of the repertoire and organisation of spider homeobox genes before and after*
748 *the WGD*

749 We characterised the homeobox gene complement of eight recently released
750 chromosomal-level assembly spider genomes and compared their genomic positions to
751 detect conserved clustering and evidence for ancestral linkage. Leite et al. (2018)
752 previously discovered that up to 59% spider homeobox families were duplicated in
753 arachnopulmonates compared to non-arachnopulmonates and that most of these were
754 dispersed duplicates rather than closely linked tandem duplicates. However, the
755 fragmented nature of the genomes available at the time meant that whether these were
756 ohnologues or tandem duplicates could not be determined. Here with synteny analysis
757 across eight chromosome-level spider genome assemblies we showed that most
758 duplicated homeobox genes are likely ohnologues retained after WGD, e.g., 44 out of
759 91 families in *P. tepidariorum*, and the vast majority have been retained in all spiders
760 surveyed (Table 2). We detected far fewer gene families (up to 9) with homeobox
761 paralogues from ancient tandem duplications and more recent lineage specific tandem
762 duplicates (up to 23 families in *H. graminicola* but fewer in other spiders) (Table 2).
763 WGD has therefore had a massive impact on the homeobox gene repertoire in spiders.

764 We also examined the impact of the arachnopulmonate WGD on clusters of homeobox
765 genes. In addition to the two Hox clusters in spiders and scorpions that have been
766 described before, as well as some evidence for the duplication of other homeobox
767 clusters in *P. tepidariorum* (Leite et al. 2018), we found two ohnologous NK, Irx, and
768 SINE clusters, but only one intact HRO cluster and degenerate ParaHox and NK2
769 clusters (Figure 3-6 and S3). For those that were largely retained in duplicate, we
770 noticed an asymmetry in the level of organisation between the clusters. While one
771 ohnologous cluster was more conserved across spiders in gene content, orientation,
772 and order, the other was more disorganised, having undergone rearrangements, gene
773 losses, and duplications. We observed this pattern across the two Hox NK, Irx, and
774 SINE clusters (Figures 3-6). This pattern was associated with chromosomal

775 organisation. Hox-B and NK-B clusters on the sex chromosomes were less degenerated
776 than those on the autosomes, though in different ways. Hox-B was retained intact, while
777 NK-B was more rearranged but exhibited larger-scale patterns of conserved linkage.
778 Given the X_1X_2 male and $X_1X_1X_2X_2$ female sex determination in entelegynes, differential
779 rates of purifying selection on sex chromosomes versus autosomes may have a role in
780 maintaining the synteny of these chromosomes, thus affecting the homeobox clusters
781 on them (Kořínková and Král 2013; Charlesworth 2017; Cordellier et al. 2020).

782 Subfunctionalisation has already been described for the Hox genes in *P. tepidariorum*,
783 with differences observed between paralogues in the timing of their activation and their
784 spatial expression domains (Schwager et al. 2007; Schwager et al. 2017; Turetzek et al.
785 2022). While spatial collinearity is exhibited by both Hox clusters, temporal collinearity is
786 maintained more for cluster A (Schwager et al. 2017; Turetzek et al. 2022), though this
787 cluster has lost *ftz* and *Hox3* in many species (Figure 3). Previously, it has been
788 hypothesised that cluster integrity is essential for the mechanism of temporal
789 collinearity, whereas spatial collinearity is controlled at the gene level (Duboule 2007;
790 Krumlauf 2018; Ferrier 2019). Here, we find the opposite; the cluster that exhibits a
791 stricter pattern of temporal collinearity is apparently more organised.

792 In different lineages within Drosophilidae, NK cluster genes have undergone an
793 interesting series of rearrangements, fragmenting and then reuniting (Chan et al. 2015).
794 While the NK genes in different drosophilid species are found in different orders, certain
795 gene pairs or triplets are consistently clustered and move as units: *slou* (NK1) with *Hmx*
796 (NK5), *lbe* (Lbx) and *lbl* (Lbx) with *C15* (Tlx), and *tin* (NK4) with *bap* (NK3). In spiders,
797 we found a core NK cluster made up of *Msx1*, *NK4*, *NK3*, *Tlx*, *Lbx* and *NK7*, largely
798 conserved in this gene order and complement and retained in duplicate (Figure 4).
799 While the NK cluster B has undergone more rearrangements and lost certain genes,
800 *Lbx-B* and *Tlx-B* are still adjacent in all eight spider genomes. NK cluster B is also linked
801 to a conserved three-gene cluster of *NK5* and two recent tandem *NK1* paralogues, and
802 sometimes to *NK6-B* and *Msx2-B*, which are ohnologous to the cluster of *NK6-A* and
803 *Msx2-A* seen across the entelegynes (Figure 4).

804 Some sub-cluster gene organisation patterns seen in spiders are similar to those seen
805 across other arthropods. In spiders, *NK3* and *NK4* are adjacent, as in *D. melanogaster*,
806 *T. castaneum*, *Bombyx terrestris*, many butterflies, and *Anopheles gambiae*, and *Tlx*
807 and *Lbx* are adjacent, as in *D. melanogaster*, *T. castaneum*, several butterflies, *A.*
808 *gambiae* and *I. scapularis* (Figure 4) (Luke et al. 2003; Chan et al. 2015; Mulhair et al.
809 2022). Though *NK1* has undergone a recent tandem duplication in spiders, *NK5* and
810 *NK1* are also adjacent in *T. castaneum*, *B. terrestris*, *A. gambiae*, and *I. scapularis*, but
811 not *D. melanogaster* or butterflies (Figure 4) (Chan et al. 2015; Mulhair et al. 2022).

812 These NK sub-clusters are even conserved across to other bilaterians (Luke et al. 2003;
813 Larroux et al. 2007; Wotton et al. 2009; Hui et al. 2012; Mulhair et al. 2022). *NK3* and
814 *NK4* are clustered, as are *Lbx* and *Tlx* in the annelid *Platynereis dumerilii* and the
815 invertebrate chordate amphioxus (*Branchiostoma floridae*) (Hui et al. 2012), while in
816 human, one *NK4* ohnologue clusters with *Tlx* and *Lbx* ohnologues on chromosome 10
817 (Luke et al. 2003; Wotton et al. 2009). In contrast, the NK sub-clusters of tardigrades
818 consist of different subgroups: *Msx* with *NK3*, *NK1* with *Lbx*, and *NK4* with *NK6* and *Tlx*,
819 while *NK5* is alone (*NK7* was not shown) (Treffkorn et al. 2018), which most likely is a
820 result of rearrangements in this lineage, but could also arise from convergent
821 rearrangements in arthropods, annelids, and chordates. Overall, these patterns suggest
822 that while the entire NK cluster has undergone many rearrangements in different animal
823 lineages, these events have repeatedly occurred between these sub-clusters, and far
824 less often within them, indicating a conserved functional constraint on these gene pairs
825 and triplets.

826 *Conserved and subfunctionalised expression of NK cluster genes*

827 NK cluster genes are broadly involved in similar processes patterning the mesoderm
828 (Jagla et al. 2001; Saudemont et al. 2008; Treffkorn et al. 2018). Here, NK cluster
829 genes were expressed during development in the mesoderm of segments, in the
830 developing heart, along the limbs, and in the head (Figure 7-9). Interestingly, we
831 observed similar expression patterns between NK cluster genes that make up
832 conserved NK sub-clusters. In *P. tepidariorum*, *NK3* and *NK4* were both expressed in

833 the developing heart and the head (Figure 7), while *Tlx* and *Lbx* were both primarily
834 expressed in the legs in both *P. tepidariorum* and *P. opilio* (Figure 9). Perhaps these
835 similar functions are a result of shared cis-regulatory elements and the selective
836 constraint for their preservation has resulted in the conservation of these sub-clusters.
837 Future study is required to identify those regulatory elements and to determine whether
838 they are indeed responsible for similar developmental roles. It would also be of interest
839 to investigate whether these are shared with other bilaterians that exhibit similar NK
840 sub-clusters.

841 We previously found a few examples of neofunctionalisation among spider homeobox
842 gene paralogues, but subfunctionalisation appeared to be the more common fate
843 (Schwager et al. 2017; Leite et al. 2018), as is the case for other genes (Schomburg et
844 al. 2015; Janssen et al. 2021). This is consistent with our analysis of the spatial
845 expression patterns of NK cluster genes, with clear complementary expression between
846 the ohnologues of *Tlx*, *Lbx*, and *NK7* in *P. tepidariorum* that largely summed to the
847 expression of the single genes in *P. opilio*. For *Emx* and *Msx*, the newly found
848 paralogues in *P. opilio* contradict previous conclusions about the timing of duplication
849 and subfunctionalisation. Searching in the recently released *P. opilio* genome (Gainett
850 et al. 2021) allowed us to identify paralogues of *Emx* and *Msx* in this harvestman,
851 suggesting these families underwent ancient tandem duplications. In previous
852 transcriptome-based analyses the highly similar sequences for each *Emx* paralogue
853 were merged into a single transcript model, and only a partial sequence was found for
854 *Msx1*, split across two different transcript models (Sharma et al. 2012).

855 Each *P. tepidariorum Emx* ohnologue is expressed in overlapping subsets of the
856 expression observed for *P. opilio Emx* paralogues, though the pattern between single-
857 copy genes and their orthologues is not so clear: the expression of *Emx1-A* and *Emx1-*
858 *B* of the spider does not equal that of *Emx1* in the harvestman, nor does *Emx2-A* and
859 *Emx2-B* equal that of *Emx2.* Rather, expression of *Emx1-A* was more similar to that of
860 *Emx2-A* than to *Emx1-B* or *Emx2-B*, and likewise, expression patterns were similar
861 between the B ohnologues (Leite et al. 2018), despite the duplication giving rise to
862 *Emx1* and *Emx2* now estimated to predate the WGD. The main observed difference in

863 *Emx* ohnologue expression patterns was temporal, which may be controlled by an
864 element shared by both tandem paralogues within a cluster, that only diverged following
865 the WGD. Furthermore, the two *Emx* paralogues in *P. opilio* were expressed in near-
866 identical patterns (Figure 8) (Leite et al. 2018), which suggests that while the tandem
867 duplication gave rise to two conserved paralogues, they seem to have redundantly
868 retained much of the ancestral expression pattern, and the spider *Emx* ohnologues only
869 diverged in function after the WGD.

870 For *Msx*, the newly described *P. opilio* gene exhibits a similar expression pattern to
871 *Msx2-B*, which now seems more likely to have subfunctionalised. This result stresses
872 the importance of a correct ancestral comparison to identify neofunctionalisation. The
873 other NK cluster genes that are missing from *P. tepidariorum*, *Msx1-A* and *NK4-B*,
874 would also be of interest in the context of subfunctionalisation. Furthermore, because of
875 shared ancient tandem duplications of *Emx* and *Msx* in spiders and the harvestman, the
876 ancestral function of the single-copy gene is not yet known but could be determined
877 from earlier-branching lineages of chelicerates or arachnids. Future investigations of NK
878 cluster ohnologues in other spiders, as well as comparisons to other arthropods
879 outgroups, may reveal further examples of sub- and/or neofunctionalisation.

880 No NK cluster ohnologues were found to be under positive selection in our analysis, yet
881 several of these genes exhibited clear patterns of subfunctionalisation. It has been
882 proposed that duplicated genes encoding developmental transcription factors like
883 homeobox genes are more likely to undergo regulatory subfunctionalisation, driven by
884 the complementary loss of cis-regulatory elements (Force et al. 1999; Jiménez-Delgado
885 et al. 2009; Espinosa-Cantú et al. 2015; Marlétaz et al. 2018). While other types of
886 paralogues may evolve by changes to coding sequence, the changes we observed
887 between *P. tepidariorum* NK ohnologues and single-copy orthologues from *P. opilio*
888 may primarily have occurred in the regulatory controls of these genes.

889 Our results suggest that there have been similar outcomes between the
890 arachnopulmonate WGD and the likely independent WGDs in horseshoe crabs (Nossa
891 et al. 2014; Kenny et al. 2016; Shingate et al. 2020; Nong et al. 2021). Not only were

892 many homeobox gene families expanded to four or more copies as a result of up to
893 three WGDs in horseshoe crabs, but there are also multiple homeobox clusters
894 conserved in duplicate, including up to six Hox clusters, six NK clusters, and seven
895 SINE clusters in *Limulus polyphemus* (Nong et al. 2021). Similar to the pattern seen in
896 spiders, some ohnologous clusters are more intact than others; for instance, two of the
897 six Hox clusters are intact and ordered, while the others have lost two, seven, and eight
898 genes in *Carcinoscorpius rotundicauda* (Shingate et al. 2020). However, in the absence
899 of additional chromosome-level assemblies, the extent to which this applies to other
900 horseshoe crabs is not fully clear yet.

901 In comparison to vertebrate homeobox gene clusters, which were duplicated in the 2R
902 WGD, the ohnologous clusters of spiders are more conserved in terms of gene content.
903 Across the four human Hox clusters, complementary gene losses have occurred such
904 that no cluster contains all gene families, with between two and four gene losses per
905 cluster, but in spiders, one Hox cluster is intact while the other has lost two of the ten
906 genes (Figure 3). The four ohnologous human NK clusters are spread across five
907 chromosomes, and most genes are retained in two or three copies, and none in four
908 (Luke et al. 2003; Wotton et al. 2009). For the human HRO cluster, *Hbn* is not found in
909 chordates, and *Otp* is not linked to *Rax* or *Rax2* in the human genome (Mazza et al.
910 2010), but in spiders an HRO cluster is conserved and intact (Figure 5). For SINE
911 genes, however, the clusters have evolved similarly between vertebrates and spiders
912 following WGD. In spiders, one cluster is rearranged with a large distance separating
913 *Six4/5* from a cluster of *Six1/2* and *Six3/6*, and ohnologues of *Six1/2* and *Six3/6* are
914 linked on another chromosome (Figure 5). In vertebrates, *Six4*, *Six1* and *Six6* are
915 linked, as are *Six2* and *Six3*, while *Six5* is on a third chromosome (Kawakami et al.
916 2000; Ferrier 2016b). In each of these lineages, all three genes are still linked on one
917 chromosome, while the ohnologous cluster(s) have undergone losses and
918 rearrangements. The pattern we see between most ohnologous spider homeobox gene
919 clusters, wherein one copy is retained intact while the other has degenerated, is similar
920 to the pattern for only some human homeobox clusters, and not the human Hox or NK
921 clusters.

922 Besides conserved clusters, some homeobox genes have returned to the single copy in
923 both humans and spiders or scorpions, namely *Hlx*, *Hhex*, *Mnx*, *Dmbx*, *Otp*, *Prop*, and
924 *Mkx* (Holland et al. 2007; Leite et al. 2018). Genes often found in the single copy in both
925 spiders and horseshoe crabs are *Bari*, *Barx*, *CG11294*, *Dmbx*, *Hhex*, *Meox*, *Mkx*, *Msxlx*,
926 and *NK4* (Nong et al. 2021). The fact that there are single copy genes in common
927 between arachnopulmonates and vertebrates or horseshoe crabs suggests that certain
928 genes may be constitutive singletons, and either they are disadvantageous in multiple
929 copies, or unable to subfunctionalise so redundant copies are lost to nonfunctionalising
930 mutations. However, these comparisons are only between three distantly related
931 species, and humans and horseshoe crabs have each undergone multiple WGDs, while
932 there was only one in arachnopulmonates. Wider sampling of WGD lineages may reveal
933 clearer patterns of genes that tend to return to the single copy, and functional and
934 genomic investigations may reveal the underlying reasons.

935 **CONCLUSIONS**

936 It is clear from analyses of WGDs in arachnopulmonates, horseshoe crabs, and
937 vertebrates that homeobox genes are commonly retained in duplicate, as seems to be
938 the case for other regulatory genes (Cañestro et al. 2013; Berthelot et al. 2014; Leite et
939 al. 2016; Harper et al. 2021; Janssen et al. 2021). Recent advances in sequencing
940 technologies and reduction of costs, as well as endeavours like the Darwin Tree of Life
941 project (Blaxter et al. 2022) have generated a growing number of chromosome-level
942 genome assemblies for a more diverse array of species. These resources now provide
943 the opportunity to identify and compare the sets of retained ohnologues and consistently
944 single-copy genes to further explore the similarities and differences between the
945 consequences of WGDs in different animal lineages. In eight spiders, we find that the
946 majority of duplicated homeobox genes likely originated in the arachnopulmonate WGD
947 and that often, cluster or sub-cluster integrity is maintained in at least one ohnologous
948 copy. Across the eight genomes, we see varying levels of conserved macrosynteny, but
949 can identify genomic rearrangements affecting homeobox cluster synteny blocks.
950 Comparisons of expression patterns of selected homeobox genes support extensive
951 subfunctionalisation among spider ohnologues. Taken together, our results help us

952 understand the impact of WGD and tandem duplication on the evolution of homeobox
953 genes and genomes in spiders and provide an important comparison to WGD in other
954 animals.

955 **AUTHOR CONTRIBUTIONS**

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

961 **ACKNOWLEDGEMENTS**

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

964 **REFERENCES**

965 Aase-Remedios ME, Ferrier DEK. 2021. Improved Understanding of the Role of Gene
966 and Genome Duplications in Chordate Evolution With New Genome and
967 Transcriptome Sequences. *Front. Ecol. Evol.* 9:429.
968 https://doi.org/10.3389/fevo.2021.703163.
969 Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment
970 search tool. *J. Mol. Biol.* 215:403–410. https://doi.org/10.1016/S0022-
971 2836(05)80360-2.
972 Azpiazu N, Frasch M. 1993. tinman and bagpipe: two homeo box genes that determine
973 cell fates in the dorsal mesoderm of Drosophila. *Genes Dev.* 7:1325–1340.
974 https://doi.org/10.1101/gad.7.7b.1325.
975 Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Ostell J, Pruitt KD, Sayers EW.
976 2018. GenBank. *Nucleic Acids Res.* 46:D41–D47.
977 https://doi.org/10.1093/nar/gkx1094.
978 Berthelot C, Brunet F, Chalopin D, Juanchich A, Bernard M, Noël B, Bento P, Da Silva
979 C, Labadie K, Alberti A, et al. 2014. The rainbow trout genome provides novel
980 insights into evolution after whole-genome duplication in vertebrates. *Nat.*
981 *Commun.* 5:3657. https://doi.org/10.1038/ncomms4657.
982 Birkan M, Schaeper ND, Chipman AD. 2011. Early patterning and blastodermal fate
983 map of the head in the milkweed bug Oncopeltus fasciatus. *Evol. Dev.* 13:436–
984 447. https://doi.org/10.1111/j.1525-142X.2011.00497.x.
985 Blaxter M, Mieszkowska N, Di Palma F, Holland P, Durbin R, Richards T, Berriman M,
986 Kersey P, Hollingsworth P, Wilson W, et al. 2022. Sequence locally, think

987 globally: The Darwin Tree of Life Project. *Proc. Natl. Acad. Sci.* [Internet] 119.
988 https://doi.org/10.1073/pnas.2115642118. Available from:
989 http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2115642118/-
990 /DCSupplemental.
991 Brooke NM, Garcia-Fernàndez J, Holland PWH. 1998. The ParaHox gene cluster is an
992 evolutionary sister of the Hox gene cluster. *Nature* 392:920–922.
993 https://doi.org/10.1038/31933.
994 Bürglin TR. 2011. Homeodomain Subtypes and Functional Diversity. In: Hughes TR,
995 editor. A Handbook of Transcription Factors. Subcellular Biochemistry.
996 Dordrecht: Springer Netherlands. p. 95–122. Available from:
997 https://doi.org/10.1007/978-90-481-9069-0_5
998 Butts T, Holland PWH, Ferrier DEK. 2008. The Urbilaterian Super-Hox cluster. *Trends*
999 *Genet.* 24:259–262. https://doi.org/10.1016/j.tig.2007.09.006.
1000 Butts T, Holland PWH, Ferrier DEK. 2010. Ancient homeobox gene loss and the
1001 evolution of chordate brain and pharynx development: deductions from
1002 amphioxus gene expression. *Proc. R. Soc. B Biol. Sci.* 277:3381–3389.
1003 https://doi.org/10.1098/rspb.2010.0647.
1004 Cañestro C, Albalat R, Irimia M, Garcia-Fernàndez J. 2013. Impact of gene gains,
1005 losses and duplication modes on the origin and diversification of vertebrates.
1006 *Semin. Cell Dev. Biol.* 24:83–94. https://doi.org/10.1016/j.semcdb.2012.12.008.
1007 Cavodeassi F, Modolell J, Gómez-Skarmeta JL. 2001. The Iroquois family of genes:
1008 from body building to neural patterning. *Development* 128:2847–2855.
1009 https://doi.org/10.1242/dev.128.15.2847.
1010 Chan C, Jayasekera S, Kao B, Páramo M, von Grotthuss M, Ranz JM. 2015.
1011 Remodelling of a homeobox gene cluster by multiple independent gene reunions
1012 in Drosophila. *Nat. Commun.* 6:6509. https://doi.org/10.1038/ncomms7509.
1013 Charlesworth D. 2017. Evolution of recombination rates between sex chromosomes.
1014 *Philos. Trans. R. Soc. B Biol. Sci.* 372:20160456.
1015 https://doi.org/10.1098/rstb.2016.0456.
1016 Chipman AD, Ferrier DEK, Brena C, Qu J, Hughes DST, Schröder R, Torres-Oliva M,
1017 Znassi N, Jiang H, Almeida FC, et al. 2014. The First Myriapod Genome
1018 Sequence Reveals Conservative Arthropod Gene Content and Genome
1019 Organisation in the Centipede Strigamia maritima.Tyler-Smith C, editor. *PLoS*
1020 *Biol.* 12:e1002005. https://doi.org/10.1371/journal.pbio.1002005.
1021 Cordellier M, Schneider JM, Uhl G, Posnien N. 2020. Sex differences in spiders: from
1022 phenotype to genomics. *Dev. Genes Evol.* 230:155–172.
1023 https://doi.org/10.1007/s00427-020-00657-6.
1024 David R, Wedlich D. 2001. PCR-Based RNA Probes: A Quick and Sensitive Method to
1025 Improve Whole Mount Embryo In Situ Hybridizations. *BioTechniques* 30:768–
1026 774. https://doi.org/10.2144/01304st02.
1027 De S, Kingan SB, Kitsou C, Portik DM, Foor SD, Frederick JC, Rana VS, Paulat NS,
1028 Ray DA, Wang Y, et al. 2023. A high-quality Ixodes scapularis genome advances
1029 tick science. *Nat. Genet.* 55:301–311. https://doi.org/10.1038/s41588-022-01275-
1030 w.
1031 Duboule D. 2007. The rise and fall of Hox gene clusters. *Development* 134:2549–2560.
1032 https://doi.org/10.1242/dev.001065.

1033 Edgar RC. 2004. MUSCLE: Multiple sequence alignment with high accuracy and high
1034 throughput. *Nucleic Acids Res.* 32:1792–1797.
1035 https://doi.org/10.1093/nar/gkh340.
1036 Emms DM, Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for
1037 comparative genomics. *Genome Biol.* 20:238. https://doi.org/10.1186/s13059-
1038 019-1832-y.
1039 Escuer P, Pisarenco VA, Fernández-Ruiz AA, Vizueta J, Sánchez-Herrero JF, Arnedo
1040 MA, Sánchez-Gracia A, Rozas J. 2022. The chromosome-scale assembly of the
1041 Canary Islands endemic spider Dysdera silvatica (Arachnida, Araneae) sheds
1042 light on the origin and genome structure of chemoreceptor gene families in
1043 chelicerates. *Mol. Ecol. Resour.* 22:375–390. https://doi.org/10.1111/1755-
1044 0998.13471.
1045 Espinosa-Cantú A, Ascencio D, Barona-Gómez F, De Luna A. 2015. Gene duplication
1046 and the evolution of moonlighting proteins. *Front. Genet.* 6:227.
1047 https://doi.org/10.3389/fgene.2015.00227.
1048 Fan Z, Yuan T, Liu P, Wang LY, Jin JF, Zhang F, Zhang ZS. 2021. A chromosome-level
1049 genome of the spider Trichonephila antipodiana reveals the genetic basis of its
1050 polyphagy and evidence of an ancient whole-genome duplication event.
1051 *GigaScience* 10:1–15. https://doi.org/10.1093/GIGASCIENCE/GIAB016.
1052 Ferrier DEK. 2016a. The origin of the Hox/ParaHox genes, the Ghost Locus hypothesis
1053 and the complexity of the first animal. *Brief. Funct. Genomics* 15:333–341.
1054 https://doi.org/10.1093/bfgp/elv056.
1055 Ferrier DEK. 2016b. Evolution of Homeobox Gene Clusters in Animals: The Giga-
1056 Cluster and Primary vs. Secondary Clustering. *Front. Ecol. Evol.* 4:1–13.
1057 https://doi.org/10.3389/fevo.2016.00036.
1058 Ferrier DEK. 2019. Space and time in Hox/ParaHox gene cluster evolution. In: Fusco G,
1059 editor. Perspectives on Evolutionary and Developmental Biology: Essays for
1060 Alessandro Minelli. Padova: Padova University Press. p. 245–274.
1061 Ferrier DEK, Minguillón C. 2003. Evolution of the Hox/ParaHox gene clusters. *Int. J.*
1062 *Dev. Biol.* 47:605–611. https://doi.org/10.1387/ijdb.14756336.
1063 FigTree. 2018. FigTree v1.4.4. Available from: http://tree.bio.ed.ac.uk/software/figtree/
1064 Force A, Lynch M, Pickett FB, Amores A, Yan Y, Postlethwait J. 1999. Preservation of
1065 Duplicate Genes by Complementary, Degenerative Mutations. *Genetics*
1066 151:1531–1545.
1067 Gainett G, Crawford AR, Klementz BC, So C, Baker CM, Setton EVW, Sharma PP.
1068 2022. Eggs to long-legs: embryonic staging of the harvestman Phalangium opilio
1069 (Opiliones), an emerging model arachnid. *Front. Zool.* 19:11.
1070 https://doi.org/10.1186/s12983-022-00454-z.
1071 Gainett G, González VL, Ballesteros JA, Setton EVW, Baker CM, Barolo Gargiulo L,
1072 Santibáñez-López CE, Coddington JA, Sharma PP. 2021. The genome of a
1073 daddy-long-legs (Opiliones) illuminates the evolution of arachnid appendages.
1074 *Proc. R. Soc. B Biol. Sci.* 288:20211168. https://doi.org/10.1098/rspb.2021.1168.
1075 Gallardo ME, Lopez-Rios J, Fernaud-Espinosa I, Granadino B, Sanz R, Ramos C,
1076 Ayuso C, Seller MJ, Brunner HG, Bovolenta P, et al. 1999. Genomic cloning and
1077 characterization of the human homeobox gene SIX6 reveals a cluster of SIX
1078 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.

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

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

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

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

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

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

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

Hui JH, Raible F, Korchagina N, Dray N, Samain S, Magdelenat G, Jubin C, Segurens B, Balavoine G, Arendt D, et al. 2009. Features of the ancestral bilaterian inferred from Platynereis dumerilii ParaHox genes. *BMC Biol.* 7:43. https://doi.org/10.1186/1741-7007-7-43.

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

1125 Irimia M, Maeso I, Garcia-Fernàndez J. 2008. Convergent Evolution of Clustering of
1126 Iroquois Homeobox Genes across Metazoans. *Mol. Biol. Evol.* 25:1521–1525.
1127 https://doi.org/10.1093/molbev/msn109.
1128 Irimia M, Tena JJ, Alexis MS, Fernandez-Miñan A, Maeso I, Bogdanovic O, de la Calle-
1129 Mustienes E, Roy SW, Gómez-Skarmeta JL, Fraser HB. 2012. Extensive
1130 conservation of ancient microsynteny across metazoans due to cis-regulatory
1131 constraints. *Genome Res.* 22:2356–2367. https://doi.org/10.1101/gr.139725.112.
1132 Jagla K, Bellard M, Frasch M. 2001. A cluster of Drosophila homeobox genes involved
1133 in mesoderm differentiation programs. *BioEssays* 23:125–133.
1134 https://doi.org/10.1002/1521-1878.
1135 Janssen R. 2017. Comparative analysis of gene expression patterns in the arthropod
1136 labrum and the onychophoran frontal appendages, and its implications for the
1137 arthropod head problem. *EvoDevo* 8:1. https://doi.org/10.1186/s13227-016-0064-
1138 4.
1139 Janssen R, Andersson E, Betnér E, Bijl S, Fowler W, Höök L, Leyhr J, Mannelqvist A,
1140 Panara V, Smith K, et al. 2018. Embryonic expression patterns and phylogenetic
1141 analysis of panarthropod sox genes: insight into nervous system development,
1142 segmentation and gonadogenesis. *BMC Evol. Biol.* 18:88.
1143 https://doi.org/10.1186/s12862-018-1196-z.
1144 Janssen R, Damen WGM. 2008. Diverged and conserved aspects of heart formation in
1145 a spider. *Evol. Dev.* 10:155–165. https://doi.org/10.1111/j.1525-
1146 142X.2008.00223.x.
1147 Janssen R, Pechmann M, Turetzek N. 2021. A chelicerate Wnt gene expression atlas:
1148 novel insights into the complexity of arthropod Wnt-patterning. *EvoDevo* 12:12.
1149 https://doi.org/10.1186/s13227-021-00182-1.
1150 Jiménez-Delgado S, Pascual-Anaya J, Garcia-Fernàndez J. 2009. Implications of
1151 duplicated cis-regulatory elements in the evolution of metazoans: the DDI model
1152 or how simplicity begets novelty. *Brief. Funct. Genomic. Proteomic.* 8:266–275.
1153 https://doi.org/10.1093/bfgp/elp029.
1154 Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. 2017.
1155 ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat.*
1156 *Methods* 14:587–589. https://doi.org/10.1038/nmeth.4285.
1157 Kawakami K, Sato S, Ozaki H, Ikeda K. 2000. Six family genes-structure and function
1158 as transcription factors and their roles in development. *BioEssays* 22:616–626.
1159 https://doi.org/10.1002/1521-1878(200007)22:7<616::AID-BIES4>3.0.CO;2-R.
1160 Kenny NJ, Chan KW, Nong W, Qu Z, Maeso I, Yip HY, Chan TF, Kwan HS, Holland
1161 PWH, Chu KH, et al. 2016. Ancestral whole-genome duplication in the marine
1162 chelicerate horseshoe crabs. *Heredity* 116:190–199.
1163 https://doi.org/10.1038/hdy.2015.89.
1164 Kerner P, Ikmi A, Coen D, Vervoort M. 2009. Evolutionary history of the iroquois/Irx
1165 genes in metazoans. *BMC Evol. Biol.* 9:74. https://doi.org/10.1186/1471-2148-9-
1166 74.
1167 Kořínková T, Král J. 2013. Karyotypes, Sex Chromosomes, and Meiotic Division in
1168 Spiders. In: Nentwig W, editor. Spider Ecophysiology. Berlin, Heidelberg:
1169 Springer. p. 159–171. Available from: https://doi.org/10.1007/978-3-642-33989-
1170 9_12

1171 Krumlauf R. 2018. Hox genes, clusters and collinearity. *Int. J. Dev. Biol.* 62:659–663.
1172 https://doi.org/10.1387/ijdb.180330rr.
1173 Kumar JP. 2009. The sine oculis homeobox (SIX) family of transcription factors as
1174 regulators of development and disease. *Cell. Mol. Life Sci.* 66:565–583.
1175 https://doi.org/10.1007/s00018-008-8335-4.
1176 Larroux C, Fahey B, Degnan SM, Adamski M, Rokhsar DS, Degnan BM. 2007. The NK
1177 Homeobox Gene Cluster Predates the Origin of Hox Genes. *Curr. Biol.* 17:706–
1178 710. https://doi.org/10.1016/j.cub.2007.03.008.
1179 Larroux C, Luke GN, Koopman P, Rokhsar DS, Shimeld SM, Degnan BM. 2008.
1180 Genesis and Expansion of Metazoan Transcription Factor Gene Classes. *Mol.*
1181 *Biol. Evol.* 25:980–996. https://doi.org/10.1093/molbev/msn047.
1182 Leite DJ, Baudouin-Gonzalez L, Iwasaki-Yokozawa S, Lozano-Fernandez J, Turetzek
1183 N, Akiyama-Oda Y, Prpic N-M, Pisani D, Oda H, Sharma PP, et al. 2018.
1184 Homeobox Gene Duplication and Divergence in Arachnids. *Mol. Biol. Evol.*
1185 35:2240–2253. https://doi.org/10.1093/molbev/msy125.
1186 Leite DJ, Ninova M, Hilbrant M, Arif S, Griffiths-Jones S, Ronshaugen M, McGregor AP.
1187 2016. Pervasive microRNA Duplication in Chelicerates: Insights from the
1188 Embryonic microRNA Repertoire of the Spider Parasteatoda tepidariorum.
1189 *Genome Biol. Evol.* 8:2133–2144. https://doi.org/10.1093/gbe/evw143.
1190 Li Y, Nong W, Baril T, Yip HY, Swale T, Hayward A, Ferrier DEK, Hui JHL. 2020.
1191 Reconstruction of ancient homeobox gene linkages inferred from a new high-
1192 quality assembly of the Hong Kong oyster (Magallana hongkongensis) genome.
1193 *BMC Genomics* 21:713. https://doi.org/10.1186/s12864-020-07027-6.
1194 Luke GN, Castro LFC, McLay K, Bird C, Coulson A, Holland PWH. 2003. Dispersal of
1195 NK homeobox gene clusters in amphioxus and humans. *Proc. Natl. Acad. Sci.*
1196 100:5292–5295. https://doi.org/10.1073/pnas.0836141100.
1197 Lynch VJ, Roth JJ, Wagner GP. 2006. Adaptive evolution of Hox-gene homeodomains
1198 after cluster duplications. *BMC Evol. Biol.* 6:86. https://doi.org/10.1186/1471-
1199 2148-6-86.
1200 Marlétaz F, Firbas PN, Maeso I, Tena JJ, Bogdanovic O, Perry M, Wyatt CDR, de la
1201 Calle-Mustienes E, Bertrand S, Burguera D, et al. 2018. Amphioxus functional
1202 genomics and the origins of vertebrate gene regulation. *Nature* 564:64–70.
1203 https://doi.org/10.1038/s41586-018-0734-6.
1204 Mazza ME, Pang K, Reitzel AM, Martindale MQ, Finnerty JR. 2010. A conserved cluster
1205 of three PRD-class homeobox genes (homeobrain, rx and orthopedia) in the
1206 Cnidaria and Protostomia. *EvoDevo* 1:3. https://doi.org/10.1186/2041-9139-1-3.
1207 McLysaght A, Hokamp K, Wolfe KH. 2002. Extensive genomic duplication during early
1208 chordate evolution. *Nat. Genet.* 31:200–204. https://doi.org/10.1038/ng884.
1209 Meyer A, Schartl M. 1999. Gene and genome duplications in vertebrates: the one-to-
1210 four (-to-eight in fish) rule and the evolution of novel gene functions. *Curr. Opin.*
1211 *Cell Biol.* 11:699–704. https://doi.org/10.1016/S0955-0674(99)00039-3.
1212 Miller J, Zimin AV, Gordus A. 2023. Chromosome-level genome and the identification of
1213 sex chromosomes in Uloborus diversus. *GigaScience* 12:giad002.
1214 https://doi.org/10.1093/gigascience/giad002.
1215 Mittmann B, Wolff C. 2012. Embryonic development and staging of the cobweb spider
1216 Parasteatoda tepidariorum C. L. Koch, 1841 (syn.: Achaearanea tepidariorum;

1217 Araneomorphae; Theridiidae). *Dev. Genes Evol.* 222:189–216.
1218 https://doi.org/10.1007/s00427-012-0401-0.
1219 Muffato M, Louis A, Nguyen NTT, Lucas J, Berthelot C, Roest Crollius H. 2023.
1220 Reconstruction of hundreds of reference ancestral genomes across the
1221 eukaryotic kingdom. *Nat. Ecol. Evol.* 7:355–366. https://doi.org/10.1038/s41559-
1222 022-01956-z.
1223 Mulhair P, Crowley L, Boyes DH, Harper A, Lewis OT, Consortium DT of L, Holland
1224 PWH. 2022. Diversity, duplication, and genomic organization of homeobox genes
1225 in Lepidoptera. *Genome Res.*:gr.277118.122.
1226 https://doi.org/10.1101/gr.277118.122.
1227 Negre B, Ruiz A. 2007. HOM-C evolution in Drosophila: is there a need for Hox gene
1228 clustering? *Trends Genet.* 23:55–59. https://doi.org/10.1016/j.tig.2006.12.001.
1229 Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: A Fast and
1230 Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies.
1231 *Mol. Biol. Evol.* 32:268–274. https://doi.org/10.1093/molbev/msu300.
1232 Nong W, Qu Z, Li Y, Barton-Owen T, Wong AYP, Yip HY, Lee HT, Narayana S, Baril T,
1233 Swale T, et al. 2021. Horseshoe crab genomes reveal the evolution of genes and
1234 microRNAs after three rounds of whole genome duplication. *Commun. Biol.* 4:83.
1235 https://doi.org/10.1038/s42003-020-01637-2.
1236 Nossa CW, Havlak P, Yue J-X, Lv J, Vincent KY, Brockmann HJ, Putnam NH. 2014.
1237 Joint assembly and genetic mapping of the Atlantic horseshoe crab genome
1238 reveals ancient whole genome duplication. *GigaScience* 3:2047-217X-3–9.
1239 https://doi.org/10.1186/2047-217X-3-9.
1240 Ontano AZ, Gainett G, Aharon S, Ballesteros JA, Benavides LR, Corbett KF, Gavish-
1241 Regev E, Harvey MS, Monsma S, Santibáñez-López CE, et al. 2021. Taxonomic
1242 Sampling and Rare Genomic Changes Overcome Long-Branch Attraction in the
1243 Phylogenetic Placement of Pseudoscorpions. *Mol. Biol. Evol.* 38:2446–2467.
1244 https://doi.org/10.1093/molbev/msab038.
1245 Pollard SL, Holland PWH. 2000. Evidence for 14 homeobox gene clusters in human
1246 genome ancestry. *Curr. Biol.* 10:1059–1062. https://doi.org/10.1016/S0960-
1247 9822(00)00676-X.
1248 Prpic N-M, Schoppmeier M, Damen WGM. 2008. Collection and Fixation of Spider
1249 Embryos. *Cold Spring Harb. Protoc.* 2008:pdb.prot5067.
1250 https://doi.org/10.1101/pdb.prot5067.
1251 Putnam NH, Butts T, Ferrier DEK, Furlong RF, Hellsten U, Kawashima T, Robinson-
1252 Rechavi M, Shoguchi E, Terry A, Yu J-K, et al. 2008. The amphioxus genome
1253 and the evolution of the chordate karyotype. *Nature* 453:1064–1071.
1254 https://doi.org/10.1038/nature06967.
1255 Sánchez-Herrero JF, Frías-López C, Escuer P, Hinojosa-Alvarez S, Arnedo MA,
1256 Sánchez-Gracia A, Rozas J. 2019. The draft genome sequence of the spider
1257 Dysdera silvatica (Araneae, Dysderidae): A valuable resource for functional and
1258 evolutionary genomic studies in chelicerates. *GigaScience* 8:giz099.
1259 https://doi.org/10.1093/gigascience/giz099.
1260 Saudemont A, Dray N, Hudry B, Le Gouar M, Vervoort M, Balavoine G. 2008.
1261 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.

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

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

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

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

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

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

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

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

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

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

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

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

1307 So WL, Nong W, Xie Y, Baril T, Ma H, Qu Z, Haimovitz J, Swale T, Gaitan-Espitia JD,
1308 Lau KF, et al. 2022. Myriapod genomes reveal ancestral horizontal gene transfer
1309 and hormonal gene loss in millipedes. *Nat. Commun.* 13:3010.
1310 https://doi.org/10.1038/s41467-022-30690-0.
1311 Stierwald M, Yanze N, Bamert RP, Kammermeier L, Schmid V. 2004. The Sine
1312 oculis/Six class family of homeobox genes in jellyfish with and without eyes:
1313 development and eye regeneration. *Dev. Biol.* 274:70–81.
1314 https://doi.org/10.1016/j.ydbio.2004.06.018.
1315 Suyama M, Torrents D, Bork P. 2006. PAL2NAL: robust conversion of protein sequence
1316 alignments into the corresponding codon alignments. *Nucleic Acids Res.*
1317 34:W609–W612. https://doi.org/10.1093/nar/gkl315.
1318 Tinti M, Johnson C, Toth R, Ferrier DEK, Mackintosh C. 2012. Evolution of signal
1319 multiplexing by 14-3-3-binding 2R-ohnologue protein families in the vertebrates.
1320 *Open Biol.* 2:120103. https://doi.org/10.1098/rsob.120103.
1321 Treffkorn S, Kahnke L, Hering L, Mayer G. 2018. Expression of NK cluster genes in the
1322 onychophoran Euperipatoides rowelli: implications for the evolution of NK family
1323 genes in nephrozoans. *EvoDevo* 9:17. https://doi.org/10.1186/s13227-018-0105-
1324 2.
1325 Turetzek N, Pechmann M, Janssen R, Prpic N-M. 2022. Hox genes in spiders: Their
1326 significance for development and evolution. *Semin. Cell Dev. Biol.*:in press.
1327 https://doi.org/10.1016/j.semcdb.2022.11.014.
1328 Walldorf U, Gehring W j. 1992. Empty spiracles, a gap gene containing a homeobox
1329 involved in Drosophila head development. *EMBO J.* 11:2247–2259.
1330 https://doi.org/10.1002/j.1460-2075.1992.tb05284.x.
1331 Wang Z, Zhu K, Li H, Gao L, Huang H, Ren Y, Xiang H. 2022. Chromosome-level
1332 genome assembly of the black widow spider Latrodectus elegans illuminates
1333 composition and evolution of venom and silk proteins. *GigaScience* 11:1–11.
1334 https://doi.org/10.1093/GIGASCIENCE/GIAC049.
1335 Wotton KR, Weierud FK, Juárez-Morales JL, Alvares LE, Dietrich S, Lewis KE. 2009.
1336 Conservation of gene linkage in dispersed vertebrate NK homeobox clusters.
1337 *Dev. Genes Evol.* 219:481–496. https://doi.org/10.1007/s00427-009-0311-y.
1338 Zhong YF, Butts T, Holland PWH. 2008. HomeoDB: A database of homeobox gene
1339 diversity. *Evol. Dev.* 10:516–518. https://doi.org/10.1111/j.1525-
1340 142X.2008.00266.x.
1341 Zhong YF, Holland PWH. 2011. HomeoDB2: Functional expansion of a comparative
1342 homeobox gene database for evolutionary developmental biology. *Evol. Dev.*
1343 13:567–568. https://doi.org/10.1111/j.1525-142X.2011.00513.x.
1344 Zhu B, Jin P, Hou Z, Li J, Wei S, Li S. 2022. Chromosomal-level genome of a sheet-
1345 web spider provides insight into the composition and evolution of venom. *Mol.*
1346 *Ecol. Resour.* 22:2333–2348. https://doi.org/10.1111/1755-0998.13601.
1347 Zhu B, Jin P, Zhang Y, Shen Y, Wang W, Li S. 2023. Genomic and transcriptomic
1348 analyses support a silk gland origin of spider venom glands. *BMC Biol.* 21:82.
1349 https://doi.org/10.1186/s12915-023-01581-7.
1350