Investigating the role of gene duplication in the evolution and divergence of Arachnids

Luís Miguel Baudouin Gonzalez

Degree awarded by
Oxford Brookes University

This thesis is submitted in partial fulfilment of the requirements of the award of Doctor of Philosophy

March 2020

Acknowledgements

First, I would like to thank Alistair for providing me with the amazing opportunity to work in his lab and for introducing me to the exciting field of spider evo-devo. This thesis would not have been possible without his constant support and advice. Two other people were also essential throughout my PhD: Anna, who taught me all the lab techniques I needed to successfully study spiders; and Daniel, who introduced me to the subject of gene duplication in spiders and was a tremendous help with all bioinformatics tools. To both of them, a big thank you. I would also like to thank all other current and past members of the McGregor lab – Alex, Jo, Chris, Cláudia, Sebastian, Michaela, Pedro, Franzi, Shamma and Javi – who have helped me in one way or another in the lab. It was a pleasure working with such an amazing group of people.

Without a doubt, the list would not be complete without mentioning the close group of friends I made over the course of the last four years. Sam, Ela, Natalie, Emily and James, your friendship and support kept me sane through the toughest of times, and turned what could be an ordeal into a somewhat pleasant journey.

At last, I am forever grateful to my parents and their unconditional love and support in every choice I made throughout the years, I would not have been here without them urging me to always strive harder to fulfil my dreams. I am also thankful to all the other members of my family, who have also always believed in me and gave me strength to follow the path I chose.

Abstract

Gene duplication underlies the origin of novel genes with potential new functions that can contribute to organismal divergence. Relaxed selection on duplicates allows for the accumulation of mutations in coding and/or regulatory sequences of duplicated genes leading to subfunctionalization and/or neofunctionalization. Recent studies of spiders and scorpions have revealed a high prevalence of duplicated genes, including two Hox clusters and approximately 40% of other homeodomain containing genes. This supports the finding that there was a whole genome duplication (WGD) event in the common ancestor of these animals (Arachnopulmonata) that is not found in outgroups like ticks, mites and harvestmen. Studies of other known animal WGDs have shown how these events are often associated with an increase in organismal complexity, which is frequently linked to the diversification and functional divergence of entire gene families.

To better understand the impact of this WGD event during arachnid evolution, I identified the homeobox and Sox gene repertoires of *Phalangium opilio* and compared the embryonic expression patterns of Sox and duplicated homeodomain genes in spiders, and their single copy orthologues in harvestmen. My results reveal striking patterns of sub and neofunctionalisation among ohnologs of homeodomain and Sox genes in spiders when compared to harvestmen. Taken together, my results suggest that the retention and subsequent divergence of ohnologs originated in the arachnopulmonate WGD likely contributed to the phenotypic diversification of this arachnid lineage.

Additionally, to further our knowledge on spider embryogenesis and the genetic basis of phenotypic variation, I investigated the role of a duplicated SoxB gene (*Pt-Sox21b-1*) in spider segmentation, the role of Wnt signalling during spider eye development and the expression patterns of *Pax6* and *atonal* duplicates in the developing head of *P. tepidariorum*. My results further support the role of *Pt-Sox21b-1* as a gap gene in spider segmentation, and conserved roles for Wnt signalling and *atonal* in spider eye development, similar to those observed in *Drosophila melanogaster*.

Table of Contents

1. General Introduction

Understanding how evolution generates extensive species diversity has long been one of the major questions in biology. Though diversity presents itself at the level of morphology or physiology, the first change has to occur at the molecular level, more specifically a change in the DNA. Several mechanisms of DNA modification through which evolution can act have been proposed over the last century. Of these, gene duplication has been consistently suggested as a major driving force of phenotypic diversification[1–4]. First popularised by Susumu Ohno's classic work[4], gene duplication is generally assumed to be the major process by which new genes arise, as opposed to *de novo* generation of a protein-coding gene, assimilation of parasitic genetic material or horizontal gene transfer[5]. This helps explain how a unicellular life form, with just a small subset of genes, evolved into complex multicellular organisms with thousands of genes. As such, the processes by which gene duplication drives organismal divergence has been a much-debated subject because it contributes to the evolution of all living beings, from the simplest bacteria to the most complex of animals.

1.1. Evolution by Gene Duplication

Gene duplication is a relatively common molecular event (i.e. duplications arise and are fixed in populations at a rate of approximately 1 gene^{-1} 100 million years^{-1}) in which a copy of one or more genes is generated[3]. Ohno hypothesised that the presence of two exact copies of the same gene would cause functional redundancy, relaxing selection on both duplicates[4]. This creates a unique opportunity for the accumulation of mutations in both protein coding and regulatory sequences, which may lead to changes in either protein structure and/or gene expression. These changes can give rise to a novel function or subdivide the different pre-existing functions between paralogs, allowing functional specialization[1,3,4]. For this reason, gene duplication is thought to be the main molecular mechanism behind the birth of new genes with novel functions in the animal kingdom, with other mechanisms (e.g. *de novo* genesis) contributing at a much lower rate[5,6].

1.1.1. Origin of gene duplications

There are three main molecular mechanisms through which gene duplication can occur: DNA-mediated duplication mechanisms, retroposition and whole genome duplication (WGD)[1,3,5]. Retroposition takes place when a strand of complementary DNA (cDNA), reverse transcribed from a messenger RNA (mRNA), is randomly inserted into the genome (Fig. 1.1 C)[7]. These insertions are identified by their lack of introns and the presence of a poly A tract,

and, due to the lack of regulatory sequences, they rarely give rise to fully functional genes (e.g. out of 3,771-18,700 retrocopies found in the human genome, only 120-692 are retrogenes)[1,3,7]. Nevertheless, a considerable amount of functional retrogenes has been found in several species, demonstrating that this type of gene duplication plays a small but significant role in the acquisition of new genes[5].

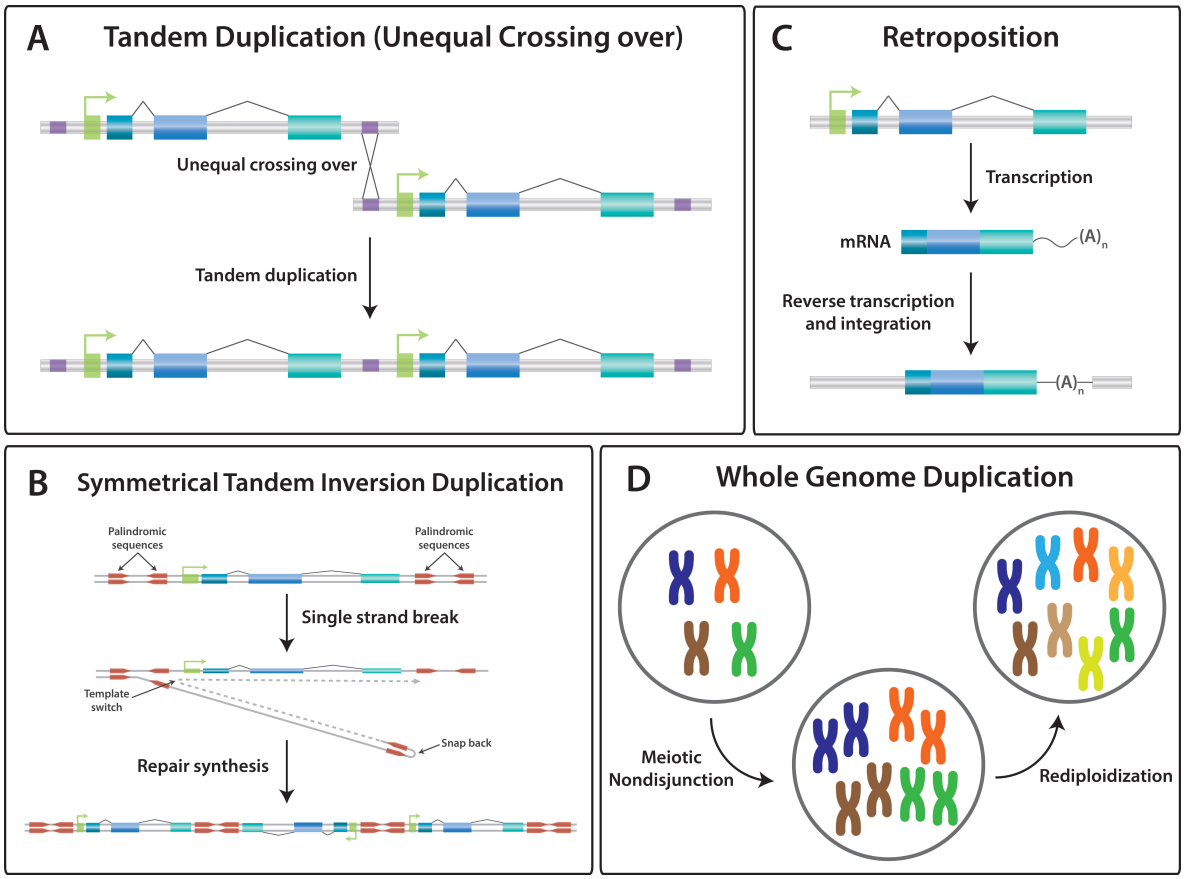

Figure 1.1: Origin of duplicated genes. (A) Tandem duplication by unequal crossing over that is caused by the recombination of two paralogous but highly similar sequences (purple boxes)[1]. (B) Formation of a symmetrical tandem inversion duplication by incorrect repair synthesis after a DNA break. This is caused by palindromic sequences (red boxes) which can snap back to prime repair synthesis. A second pair of palindromic sequences facilitates template switching[6]. (C) Creation of an intron-less gene retrocopy by the random insertion of a reverse transcribed mRNA[7]. (D) Whole genome duplication caused by meiotic nondisjunction and subsequent rediploidization[9].

DNA-mediated duplication mechanisms represent the most common origin of duplicated genes (e.g. ~80% of duplicated genes in four *Drosophila* species)[8]. There are several different DNA-mediated mechanisms through which a duplicate gene can arise, however, most of them are the result of incorrect recombination processes during meiosis or the result of incorrect DNA repair (Fig. 1.1 A and B)[5,6]. Furthermore, the outcome of this type of duplication mechanism is often the creation of tandemly duplicated sequences, which can have the same or inverted orientation and contain part of a gene, one gene or several genes[6].

WGD is the product of chromosomal nondisjunction during meiosis, an event where every chromosome is duplicated (Fig. 1.1 D)[9]. This is followed rediploidization, a process in which dramatic DNA changes reinstate correct homologous chromosome pairing and resolve gene dosage imbalances[10]. The retained duplicated genes that originated from WGD are commonly termed ohnologs[11]. WGDs were thought to be rare events in animals, however, with the enrichment of genomic data over the last two decades, several instances of WGD have now been identified in different animal lineages, suggesting these events have played important roles in the evolution of entire groups of animals as well as plants[12–14].

1.1.2. Fate of duplicated genes

Although gene duplication is a common molecular event, the probability of a duplicated gene being fixed in a population is small (one gene per MY)[1], due to the fact that most are selectively neutral and occur in a single individual[2,3]. Once fixed, a duplicated gene can still be lost over time. On average, the estimated half life of newly fixed duplicated genes can range from 2.9 MY in *Drosophila* to 7.3 MY in mammals[15]. Therefore, the most common fate of duplicated genes is that of non-functionalization or pseudogenization (~95% of duplicated genes)[15]. The lack of purifying selection allows the accumulation of mutations that can be deleterious to gene function, such as the insertion of a stop codon, disrupting important protein domains or causing loss of gene expression[2]. Over time, one of the copies loses all function and becomes a pseudogene[16].

In some cases, having another copy of a gene can be beneficial, especially if an increase of the gene product would lead to an increase in fitness. This can result in conservation of both copies, which is maintained by strong purifying selection[2,3]. Alternatively, to escape pseudogenization, one or both copies can diverge in function during the relatively short period of their half life that follows duplication[15]. This can be achieved by subdividing the original functions of the parental gene between paralogs, a process termed subfunctionalization, and/or by the acquisition of a new function by one of the copies, a process termed neofunctionalization (Fig. 1.2)[1–3].

1.1.2.1. Subfunctionalization

It is not unusual for a gene to carry out more than one function. Gene duplication creates an opportunity to partition these functions between each duplicate, lifting functional constraints and allowing specialization. This process of subfunctionalization can be accomplished at the protein level and/or at the level of regulation of gene expression (Fig. 1.2)[3]. An example of the former is the duplication and specialization of a digestive RNase in the colobine monkeys[17]. Colobine monkeys have two copies of this enzyme (RNase1 and RNase1B)

while other primates only have one, which have roles in both digestion and degradation of double-stranded RNA. The colobine monkey RNase1 retained the dsRNA degradation function while RNase1B specialised at facilitating the digestion of the leaf-based diet of these monkeys[17].

Nevertheless, subfunctionalization through the partition of ancestral expression patterns was several times suggested to be the most prevalent fate of duplicated genes, affecting both spatial and temporal patterns of duplicate expression[18–20]. One of the main proposed models by which this outcome is accomplished is the Duplication-Degeneration-Complementation (DDC) model[21]. The DDC model hypothesises that degenerative mutations are accumulated in the regulatory sequences of both copies, complementing each other in such a manner that the original expression domains are partitioned between duplicates (Fig. 1.2 B)[1,21]. This model demonstrates how duplicates can be maintained in a scenario of neutral selection, after which positive selection can optimize these subfunctions[21]. One example of this type of subfunctionalization, which is often termed specialization, is the temporal subdivision of expression of the α-globin and β-globin gene clusters, which are the product of several rounds of gene duplication[22]. Genes from each cluster are differentially expressed throughout developmental stages to form embryonic-, foetal- and adult-specific haemoglobins. This temporal subdivision of expression allowed for the optimization of O_2 binding affinity of these proteins for each developmental stage[1,22].

1.1.2.2. Neofunctionalization

Another way a duplicated gene can be retained is by gaining a new function, which can be achieved through the accumulation of rare beneficial mutations that alter gene expression and/or function (Fig. 1.2)[3]. In this scenario, one copy retains its original function while the other acquires a novel function. In theory, this new function needs to be immediately beneficial, in such a way that the initial gain in fitness is high enough to reinstate purifying selection against deleterious mutations[4]. Positive selection subsequently takes hold, resulting in an accelerated rate of non-synonymous mutations (in the case of coding sequences) that optimize the newly acquired function[1,4]. Few clear cases of neofunctionalization are known, possibly due to the fact that beneficial mutations are expected to be rare, making it less probable that such an event can occur[4,21]. However, a few genome wide studies of particular lineages suggest neofunctionalization could have a higher prevalence[23,24]. A well studied case of neofunctionalization is the gain of antibacterial activity by the ECP (eosinophil cationic protein) gene upon tandem duplication of the EDN (eosinophil-derived neurotoxin) gene in the lineage of hominoids and Old World monkeys, which was conferred by a large number of arginine substitutions[25].

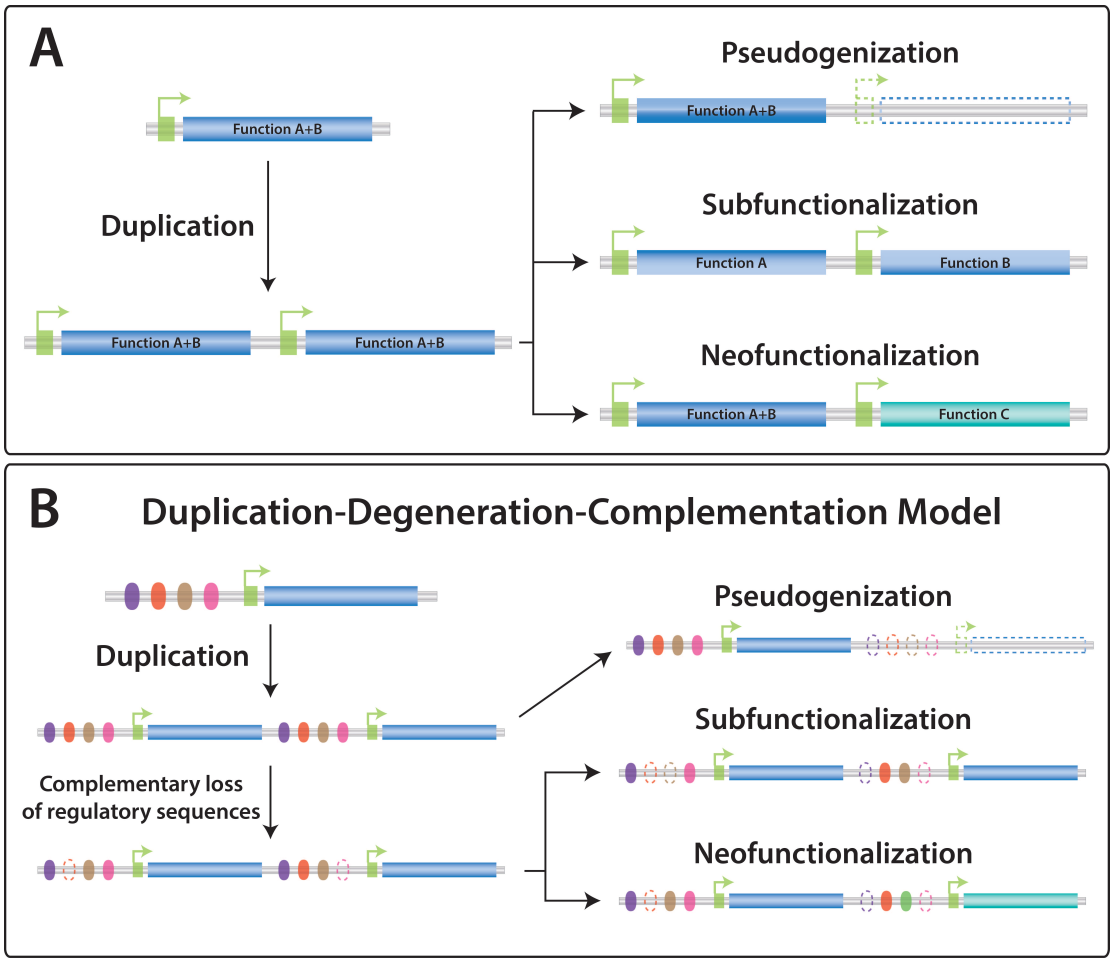

Figure 1.2: Fate of duplicated genes. (A) A duplicated gene with functions A and B can follow one out of three paths: Pseudogenization, caused by the accumulation of degenerative mutations; Subfunctionalization, resulting in the partition of functions A and B between each duplicate; Neofunctionalization, one paralog acquires the ancestral functions while the other copy acquires a new function. (B) The Duplication-Degeneration-Complementation model of duplicate gene divergence. In this model a gene with multiple regulatory sequences (coloured ovals) is duplicated along with these sequences. Pseudogenization is still a possible outcome. In the next step of this model, each copy acquires degenerative mutations in complementary regulatory sequences. This causes both copies to be essential for complete gene expression, which causes the preservation of both copies. Subsequent changes to the remaining regulatory sequences can either completely partition the ancestral expression between each duplicate, leading to subfunctionalization, or modify one of these sequences to create a new domain of expression that can resolve into a new function, leading to neofunctionalization. Adapted from Zhang [3] and Force *et al.* [21].

1.1.3. Role of ancient WGDs in evolution

WGD events are generally regarded as major contributors of organismal diversification[12,26]. Polyploidy is particularly abundant in the plant lineage, playing an important role in the diversification of vascular and flowering plants[13]. In contrast, WGD in the animal lineage appears to be much less common, with the most well studied events being the two rounds of WGD at the base of the vertebrate lineage and the teleost-specific WGD[11,14,27]. Nevertheless, other instances of WGD have now been identified in other animal lineages, such

as molluscs[28], chelicerates[29–31] and hexapods[32]. These ancient WGDs seem to be associated with highly diverse groups of animals, which supports the argument that polyploidy, rather than being an evolutionary dead end, provides a unique opportunity for phenotypic diversification[13,26]. However, not all WGD events are associated to an increase in species diversity (e.g. horseshoe crabs and paddlefish)[29,31,33] and post-WGD species radiations usually arise after an extensive time lag[34]. Lineage-specific rediploidization has been proposed as a model to explain these time lags[34].

The impact of WGD in the evolution of animal lineages was mostly studied in vertebrates. Two rounds of WGD were identified at the base of vertebrates (1R and 2R)[11], an animal lineage that has several morphological innovations like vertebrae, the skull, the adaptive immune system, and complex nervous, endocrine and circulatory systems[35,36]. These innovations are only possible through the coordinated evolution of gene regulatory networks (e.g. utilising homeobox and T-box genes) and signal transduction pathways (e.g. Wnt and Toll pathways), a process only achievable through the duplication of entire gene families[37]. Examples of this are the TGF-β pathway, which underwent a dramatic expansion of its core genes in vertebrates after the 2R-WGD[38], and the homeobox gene Superclass[39], which will be discussed in more detail in Chapter I. It has also been suggested that the duplication of key genes in the apoptosis pathway, alongside the apparently biased retention of genes expressed in neuronal tissue, could have been a decisive step for the evolution of the complex vertebrate nervous system[27,40,41]. Lastly, a recent study found evidence for the enrichment of open chromatin regions in gene families with multiple ohnologs, as well as a prevalence of specialisation in retained ohnologs, when comparing genomic data from amphioxus to that of vertebrate species[42].

The impact of WGD in animal evolution has mainly been studied on vertebrate lineages, however, extending the investigation to other independent WGD events, such as the one recently identified in the arachnid lineage[30], could give us some insight on possible general rules of gene divergence and its consequences upon WGD.

1.2. Arachnids

1.2.1. Phylogeny and relevance

Arachnids are a class of arthropods belonging to the subphylum Chelicerata, a sister group to Myriapoda and Pancrustacea in the arthropod phylogeny (Fig. 1.3)[43]. The class Arachnida harbours several well-known orders like spiders (Araneae), scorpions (Scorpiones) and ticks (Parasitiformes), as well as other less well-known orders like hooded tickspiders (Ricinulei), whip scorpions (Thelyphonida) and harvestmen (Opiliones)[44]. While not all relationships in the arachnid phylogeny are resolved, there is strong support for the grouping

of Araneae with Pedipalpi (Amblypygi, Schizomida and Thelyphonida), forming the Tetrapulmonata clade, and the addition of Scorpiones to this group is also well supported, establishing the Arachnopulmonata clade (Fig. 1.3)[45]. Relationships between other orders remain highly inconsistent, possibly due to long branch attraction caused by fast evolving lineages in Acariformes, Parasitiformes and Pseudoscorpiones. Nevertheless, there is some support for a clade comprised of Ricinulei, Solifugae and Opiliones, as well as the grouping of Pseudoscorpiones with Arachnopulmonata[45].

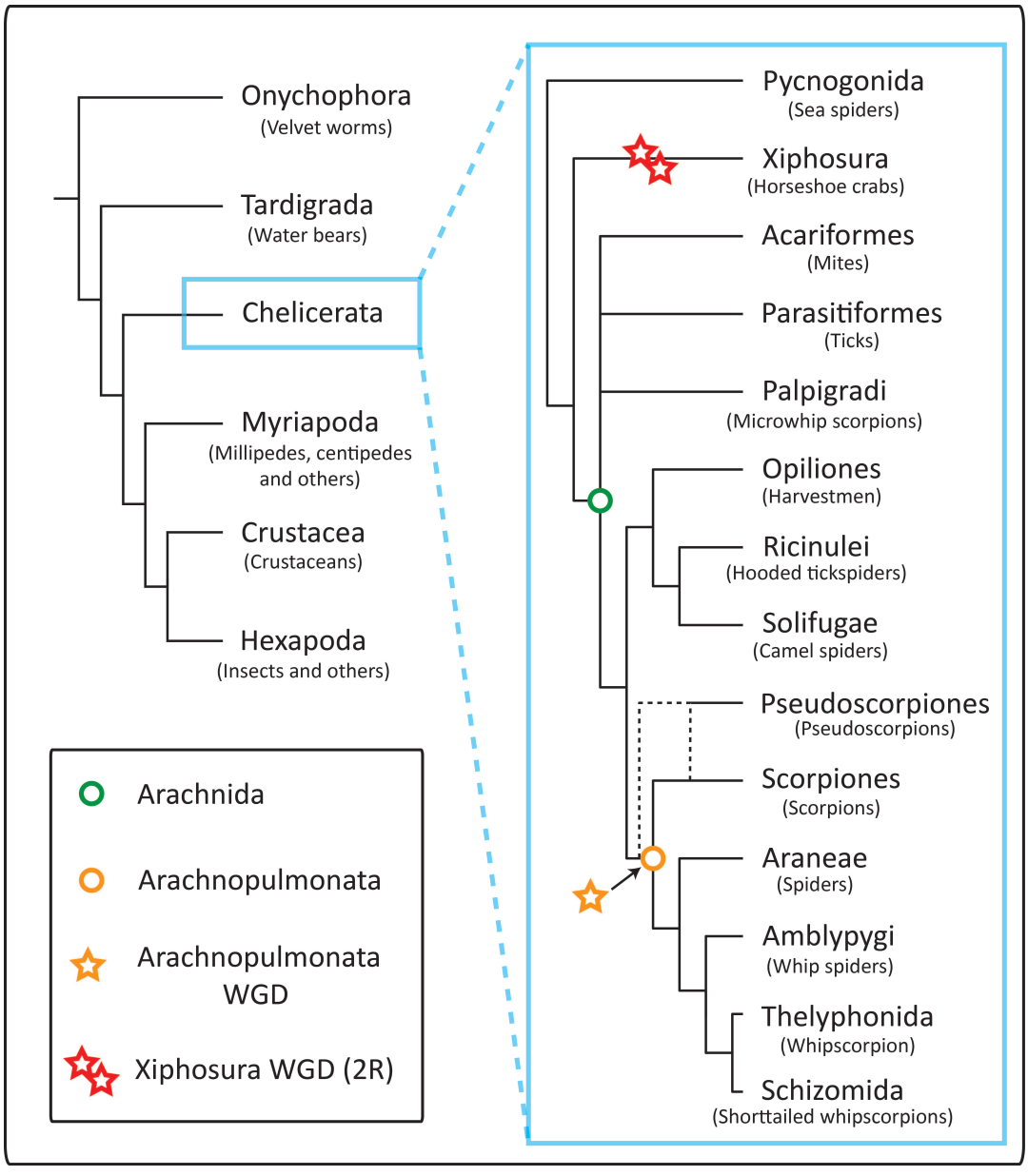

Figure 1.3: Panarthropod and Chelicerate phylogenies. Panarthropod phylogeny including all major phyla (left), as proposed by Regier *et al.* [43], and Chelicerate phylogeny, as proposed by Sharma *et al.* [45]. Dashed lines represent two alternative relationships between Pseudoscorpiones and Scorpiones. Circles mark the nodes encompassing all orders within the arachnid (green) and arachnopulmonata (yellow) lineages. Stars represent the two known WGDs in the horseshoe crab lineage (red) and the arachnopulmonata WGD (yellow).

Encompassing some of the most specious and successful orders of animals (e.g. over 45,000 described spider species with a wide range of habitats), arachnids display a vast morphological diversity[44]. Their segmented body can be subdivided into the anterior prosoma, which bears six pairs of appendages (i.e. chelicerae, pedipalps and four pairs of walking legs), and the posterior opisthosoma, which can vary in segment number and appendage composition[44]. There are several lineage specific traits, such as the book lungs, which are present in all arachnopulmonata, the telson sting in scorpions and the spinnerets in spiders. Recent studies have found evidence to support the existence of a WGD event in the common ancestor of spiders and scorpions, which led to the retention of ohnologs of several developmentally important genes[30], as well as a dramatic expansion of gene families associated with venom and silk production[46–51]. Taking all into consideration, arachnids provide a unique system to systematically study the role of WGD in evolution of morphological innovation, as well as a separate event from the vertebrate WGDs, allowing for the inference of possible general rules of the molecular and phenotypic consequences of WGD.

1.2.2. The spider model *Parasteatoda tepidariorum*

With available genome and transcriptome sequencing data, and an increasing array of molecular tools, the spider *Parasteatoda tepidariorum* has emerged as one of the main arachnid models for evolutionary and developmental biology[30,52]. This spider species is easy to maintain under lab conditions, with females providing cocoons containing up to 300 embryos on a weekly basis, which develop more or less synchronously, making it an excellent model for studies on embryonic development[52].

P. tepidariorum embryogenesis has been previously described, with a clear and detailed staging system (Fig. 1.4 and 1.5)[53]. A working *in situ* hybridization (ISH) protocol provides a base for the expression pattern analysis of any developmental gene of interest[54] and parental or embryonic RNAi (pRNAi or eRNAi) provide powerful methods to disrupt gene expression in this spider[55,56]. Together, these techniques have been key to the study of important aspects of arthropod evolution and development, such as anterior and posterior segmentation[55,57,58], appendage development and morphological diversity[56,59].

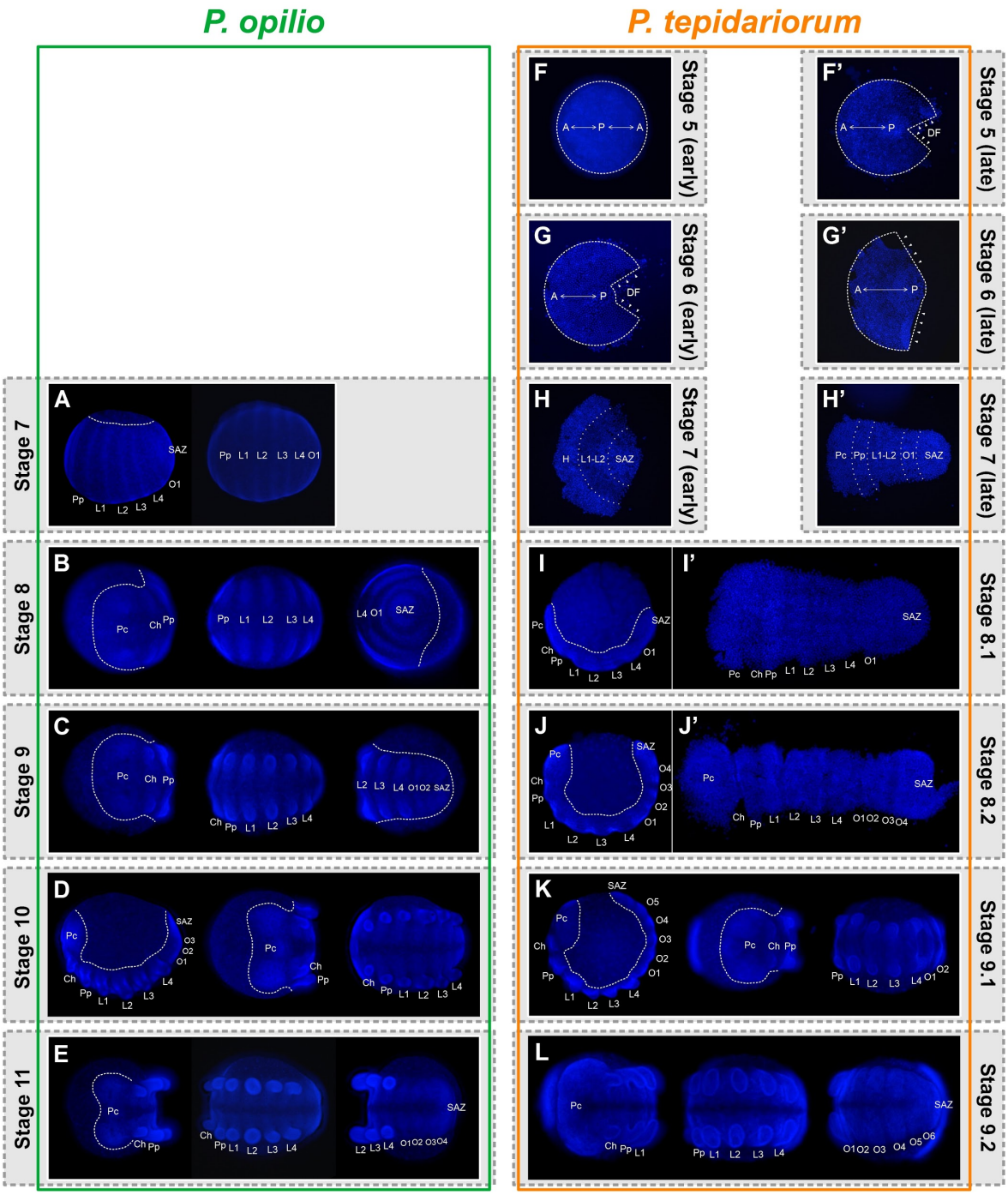

Figure 1.4: *Parasteatoda tepidariorum* and *Phalangium opilio* early embryogenesis. Early stages of embryonic development of *P. opilio* (A-E) and *P. tepidariorum* (F-L) analysed in this thesis. Anterior is to the left in all images except F-G'. Arrowheads mark the direction of the expanding dorsal field. Dashed lines represent the edge of the germ disc/band. Dotted lines in H and H' mark the borders of each germ band region. A, anterior; P, posterior; DF, dorsal field; H, presumptive head region; Pc, pre-cheliceral region; Ch, cheliceral segment; Pp, pedipalpal segment; L1-L4, leg segments 1-4; O1-O6, opisthosomal segments 1-6; SAZ, segment addition zone.

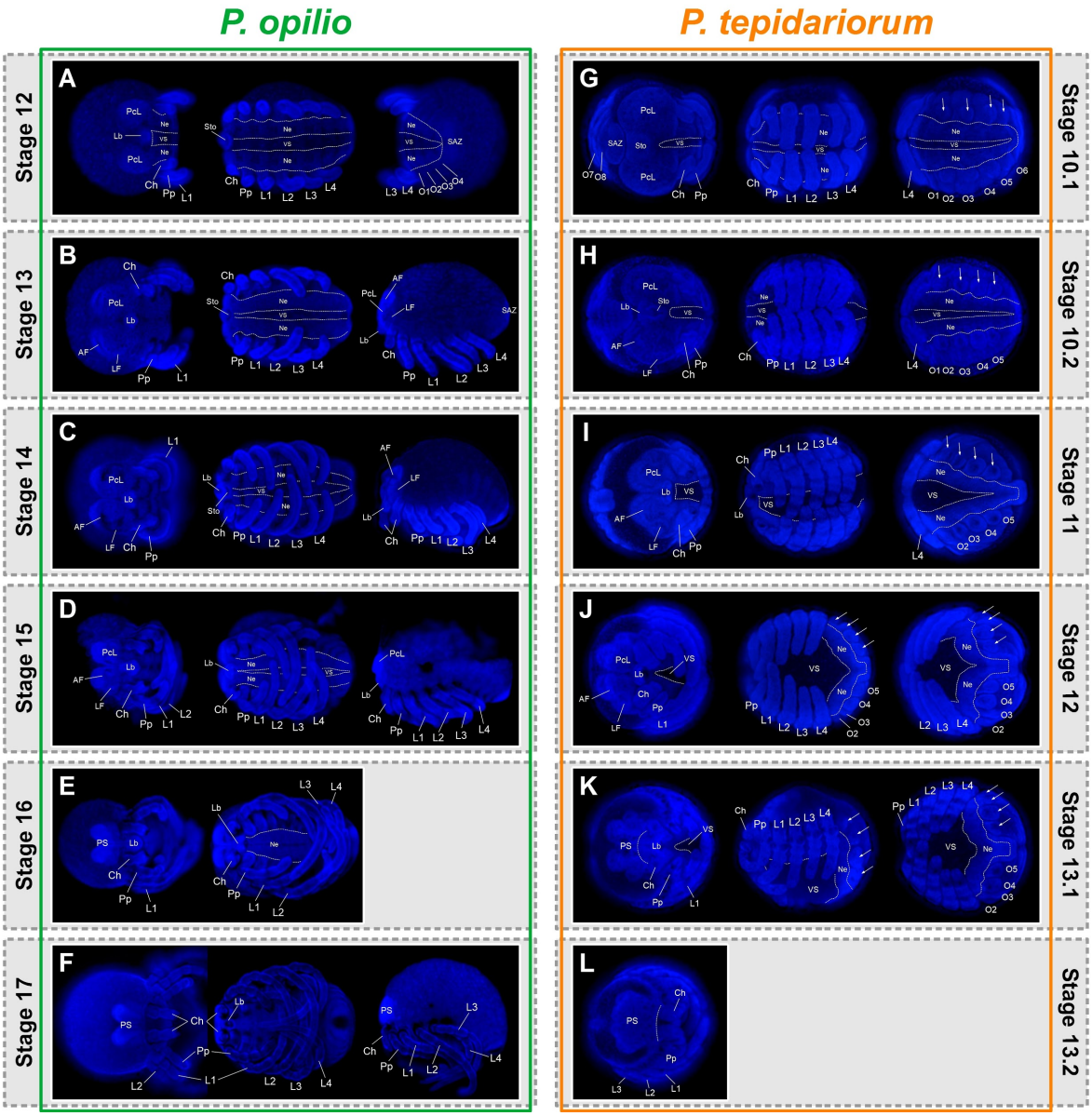

Figure 1.5: *Parasteatoda tepidariorum* and *Phalangium opilio* early embryogenesis. Early stages of embryonic development of *P. opilio* (A-F) and *P. tepidariorum* (G-L) analysed in this thesis. Anterior is to the left in all images. Arrows mark the four pairs of opisthosomal limb buds that represent the developing the tracheal system (O2), book lungs (O3) and spinnerets (O4 and O5). Dashed lines mark mark the dorsal and ventral borders of the developing neuroectoderm. PcL, Precheliceral lobes; Sto, stomodeum; Lb, labrum; Ne, neuroectoderm; VS, ventral sulcus; Ch, chelicerae; Pp, pedipalps; L1-L4, leg pairs 1-4; O1-O6, opisthosomal segments 1-6; AF, anterior furrow; LF, lateral furrow; PS, prosomal shield.

1.2.3. The harvestmen model *Phalangium opilio*

Recent evolutionary, developmental and phylogenetic studies in harvestmen have used *Phalangium opilio* as a model[45,60–63]. Occupying a pivotal position in the arachnid phylogeny (Fig. 1.3), the addition of this species as a chelicerate model was an important step for evolutionary developmental biology studies[45]. As with *P. tepidariorum*, ISH and RNAi protocols, as well as transcriptome sequencing data, are available for this species[61,63]. Maintenance of a

laboratory culture is feasible, and it is possible to obtain a considerable number of embryos from a healthy colony[64]. However, a clear staging system has yet to be published, so the staging system used in this thesis is based on recent publications (Fig. 1.4 and 1.5). Additionally, the early stages of embryonic development are not easily accessible to gene expression analysis due to technical limitations. *P. opilio* offers an excellent model to infer the state of the unduplicated ancestor, representing a closely related taxon to scorpions and spiders (i.e. Opiliones) that is thought to not have undergone WGD (Fig 1.3).

1.3. Aims

The overall aim of my PhD project was to elucidate the impact of gene duplication during arachnid development and evolution. Arachnids offer a unique opportunity to study the implications of whole genome duplication in evolution and diversification of animal lineages, with an established model organism in the spider *P. tepidariorum*. The harvestmen *P. opilio* provides a closely related species from which to infer the ancestral state of arachnopulmonate ohnologs, allowing us to distinguish between conserved and novel patterns of expression. At a more functional level, the investigation of duplicated gene functions in key developmental processes, such as segmentation and eye development, would help us better understand the role of gene duplication in the evolution of novel traits. Specialised introductions are provided at the beginning of each chapter.

1.3.1. Impact of gene duplication on arachnid Homeobox genes

I first aimed to infer the extent of functional divergence of ohnologs from a developmentally important gene set known as the homeobox superclass. To do this by compared the expression patterns of single-copy genes in the harvestmen *P. opilio* to those of duplicated genes in the spider *P. tepidariorum*. Therefore, the aims of Chapter I included:

– Identification and classification of homeobox genes in the harvestmen *P. opilio*, using available transcriptome sequence data.

– Gene expression pattern characterisation and comparison of selected duplicated homeobox genes in *P. tepidariorum* and their single-copy orthologs in *P. opilio*, identifying potential sub or neofunctionalisation through the identification of divergent patterns of expression.

1.3.2. Impact of gene duplication on arachnid Sox genes

To further understand the impact of gene duplication on key developmental gene families, I extended my analysis to Sox genes. A previous study of Sox genes in *P. tepidariorum* found that several of them were duplicated, although it did not find extensive expression divergence[65]. However, a more systematic analysis of gene expression and comparison to single-copy orthologs of *P. opilio* was needed. For this purpose, the aims of Chapter II were:

– Identification and classification of Sox genes in *C. sculpturatus*, *P. opilio*, *I. scapularis* and *Strigamia maritima*.

– Phylogenetic analysis of chelicerate Sox gene repertoires through the construction of a phylogenetic tree.

– Gene expression pattern characterisation and comparison of duplicated Sox genes in *P. tepidariorum* and their single-copy orthologs in *P. opilio*, identifying potential sub or neofunctionalisation through the identification of divergent patterns of expression.

1.3.3. Functional analysis of a duplicated Sox gene

A duplicated SoxB gene in the spider *P. tepidariorum* (*Pt-Sox21b-1*) was previously reported to have a role in segmentation[66]. Moreover, it appears to affect both anterior and posterior segmentation. Previous analysis was focused on posterior segmentation. In addition, due to the highly disruptive effect of *Pt-Sox21b-1* knockdown, a more systematic analysis of its interaction with other segmentation genes was necessary. To better understand the role of this gene in spider segmentation, in Chapter III I aimed to:

– Investigate the interaction of *Pt-Sox21b-1* with anterior segmentation genes by analysing their expression upon knockdown of *Pt-Sox21b-1* by means of pRNAi.

– Further elucidate the role of *Pt-Sox21b-1* on spider segmentation by generating knockdown clones through means of eRNAi, followed by expression analysis of segmentation genes.

1.3.4. Spider eye development: gene duplication and Wnt signalling

Previous work on the retinal determination gene network of two spiders found several members of this key developmental network to be duplicated[67,68]. Additionally, despite the fact that Wnt signalling plays a crucial role in eye development of both insects[69] and vertebrates[70], the question of whether or not this is also true in arachnids remains to be addressed.

A total of five Wnt genes are expressed in the developing head of *P. tepidariorum*, one of which is duplicated (*Pt-Wnt7.2*)[71]. Furthermore, expression of *atonal* orthologs has yet to be characterised in *P. tepidariorum*, along with two additional copies of *Pax6* genes that were recently identified[72]. All this taken into account, WGD could have played a role in the evolution of spider visual systems, which display highly variable morphology and function[73]. To address these questions, in Chapter IV I aimed to:

– Characterise the expression patterns of two additional *Pax6* genes identified in *P. tepidariorum*.

– Identify and characterise the expression pattern of *atonal* orthologs in *P. tepidariorum*.

– Characterise the expression patterns of Wnt genes in the developing head of *P. tepidariorum* in relation to the developing eye primordia (which can be visualised by *Pt-so1* staining) through the means of dFISH.

2. Methods

2.1. *P. tepidariorum* and *P. opilio* cultures

P. tepidariorum spiders were maintained at 25°C with a 12:12 h light:dark cycle. Spiderlings and adult males were fed on a diet of *vestigial* mutant flies (*Drosophila melanogaster*), and juvenile and adult females on a diet of small crickets (*Gryllodes sigillatus*). *P. opilio* were maintained at 20°C with a 14:10 h light:dark cycle, and fed on a diet of fish flakes supplemented with *Acheta domesticus* nymphs.

2.2. Bioinformatics

2.2.1. Identification of *P. opilio* homeobox genes

To identify homeobox genes in *P. opilio*, I performed a TBLASTP search (e-value 0.05) against the available *P. opilio* transcriptome (PRJNA236471)[61], using all the homeodomain protein sequences available from HomeoDB[74,75] combined with the homeodomain sequences from *P. tepidariorum* as the query. TBLASTP hits were then individually analysed in the ORFfinder NCBI online tool (https://www.ncbi.nlm.nih.gov/orffinder/; default settings except in 'ORF start codon to use' setting, where I used the 'Any sense codon' option to retrieve gene fragments lacking a start codon) to obtain predicted protein sequences. The longest obtained ORFs were then queried in the SMART BLAST NCBI online tool (https://blast.st-va.ncbi.nlm.nih.gov/blast/smartblast/; used default settings) for identification (best hit) and verification of the presence of a homeodomain. Transcripts lacking a homeodomain were discarded and highly similar sequences were treated as isoforms. Note that partial transcripts of homeobox genes or transcripts with highly divergent homeodomains might have been missed using this approach, as well as any gene not expressed during embryogenesis of *P. opilio*.

2.2.2. Identification of *C. sculpturatus*, *P. opilio*, *I. scapularis* and *S. maritima* Sox genes

A similar approach to that above was used to identify the Sox gene repertoires of *C. sculpturatus*, *P. opilio*, *I. scapularis* and *S. maritima*. HMG domain sequences of *P. tepidariorum*, *Stegodyphus mimosarum* and *D. melanogaster* available from Paese *et al.* [65] were used as queries against the available genomic and transcriptomic resources (*C. sculpturatus* – PRJNA422877; *P. opilio* – PRJNA236471; *I. scapularis* – PRJNA357111; *S. maritima* – PRJNA20501). Predicted proteins are available for all species besides *P. opilio*, which allowed me to circumvent the protein prediction step by using a BLASTP search (e-value 0.05), rather than a TBLASTP search. An additional TBLASTP search for sequences lacking a HMG domain

was performed using the complete Sox protein sequences of *P. tepidariorum* against the available *P. opilio* transcriptome.

2.2.3. Phylogenetic analysis

Protein sequence alignments of HMG domains from *P. tepidariorum*, *C. sculpturatus*, *P. opilio*, *I. scapularis*, *S. maritima*, *T. castaneum* and *D. melanogaster* were generated in MEGAv7 using MUSCLE (default settings). Phylogenetic analysis was performed using the RAxML tool[76] with the PROTGAMMALG amino acid substitution model (1000 bootstraps, rapid bootstrap algorithm). The resulting tree was visualised and processed in FigTree v1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/).

2.3. RNA extraction and cDNA synthesis

Total RNA was extracted using QiAzol (Qiagen) from stage 1-14 *P. tepidariorum* and stage 7-17 *P. opilio* embryos following the manufacturer's protocol. Obtained RNA was then used to generate cDNA using the QuantiTect reverse transcription kit (Qiagen), according to manufacturer's guidelines.

2.4. Primer design, PCR and Cloning

2.4.1. Primer design

Gene specific primers were designed with Primer3 (http://primer3.ut.ee), and a linker sequence (CCCGGGGC) was added to the 5' end of the reverse primer when carrying out direct synthesis of ISH probes. A complete list of primers used is provided in Tables S1-S4.

2.4.2. Purification of PCR products for direct synthesis of ISH probes

The desired gene fragments were amplified from cDNA (mixed stages, 1-14 for P. tepidariorum and 7-17 for P. opilio) by a standard PCR method using One*Taq*® 2x Master Mix (New England Biolabs, NEB). A second PCR was performed using the purified PCR product from the first PCR (NucleoSpin® Gel and PCR Clean-up kit, Macherey-Nagel), the gene specific forward primer and a 3' T7 universal reverse primer targeting the linker sequence (Table S4). The resulting PCR products were run on an agarose gel (1-2%), and the band with the expected size cut out and purified using the NucleoSpin® Gel and PCR Clean-up kit (Macherey-Nagel). The 2nd PCR products were sequenced using the services of Eurofins Genomics (Germany) and aligned with the MUSCLE online tool (EMBL-EBI). The correct 2nd PCR products were used for ISH probe synthesis.

2.4.3. Cloning of PCR products for ISH probe and dsRNA synthesis

The desired gene fragments were amplified from cDNA by a standard PCR method using One*Taq*® 2x Master Mix (New England Biolabs, NEB). Purified PCR products (NucleoSpin® Gel and PCR Clean-up kit, Macherey-Nagel) were cloned using the TOPO® TA cloning kit for sequencing (ThermoFisher Scientific). Colony PCR was performed using One*Taq*® 2x Master Mix (NEB) and M13 universal primers (Table S4). Purified plasmids were sequenced using the services of Eurofins Genomics (Germany) and aligned with the MUSCLE online tool (EMBL-EBI). Information on whether PCR was successful for each gene analysed is provided in Tables S5 and S6.

For ISH probe synthesis, a PCR product was amplified from the purified plasmids using M13 universal primers and One*Taq*® 2x Master Mix (NEB), and purified using the NucleoSpin® Gel and PCR Clean-up kit (Macherey-Nagel).

For dsRNA synthesis, the PCR product was amplified from a plasmid using the T7 forward, the T7/T3 reverse universal primers (TableS4) and One*Taq*® 2x Master Mix (NEB), and purified using the NucleoSpin® Gel and PCR Clean-up kit (Macherey-Nagel).

2.5. Synthesis of *in situ* hybridization probes

Sense and antisense RNA probe synthesis was performed using T3 (11031163001, Roche) or T7 polymerase (10881775001, Roche) with either DIG RNA labelling mix (11277073910, Roche) or Fluorescein RNA labelling mix (11685619910, Roche), according to manufacturer's guidelines.

2.6. Double-stranded RNA (dsRNA) synthesis

Synthesis of dsRNA was carried out using the MegaScript T7 transcription kit (Invitrogen), followed by annealing of both strands in a water bath starting at 95°C and slowly cooled down to room temperature. Purified dsRNA concentration was adjusted to 1.5-2.0 µg/µl for injections.

2.7. *In situ* hybridization in *P. tepidariorum* and *P. opilio* embryos

P. tepidariorum embryos were staged according to Mittmann & Wolff [53] and fixed following the protocol in Akiyama-Oda [54] with minor modifications. After transition to methanol, the embryos were left in this solution for at least 30 min at room temperature, followed by overnight incubation at -20°C. *P. opilio* embryos were fixed as previously described[62]. The vitelline membranes of fixed embryos were removed by dissection with Dumont 5 forceps in

methanol. For each staining, approximately 10 embryos per embryonic stage analysed were used. Staining with sense probes was performed as a negative control for all genes analysed and never produced any visible signal.

Colourimetric ISH was performed following the whole-mount protocol described in Prpic *et al.* [77], with minor modifications: steps 4-8 were replaced by two 10 minute washes in PBS-Tween-20 (0.02%) (PBS-T), and at step 18 the embryos were incubated for 30 minutes. Post-fixation was carried out followed by ethanol treatment to decrease background. For the ethanol treatment, embryos were incubated for 10 minutes in inactivation buffer (75 g glycine, 600 µL 1N HCl, 50 µL 10% Tween-20 and dH2O to 10 mL), followed by 3 wash steps with PBS-T, washed 5 min in 50% ethanol in PBS-T, washed in 100% ethanol until background had decreased, washed for 5 minutes in 50% ethanol in PBS-T and finally washed twice with PBS-T. Embryos were then counterstained with DAPI (1:2000; 10236276001, Roche) for ~20 minutes and stored in 80% glycerol in 1x PBS at 4°C. Imaging was performed using a Zeiss Axio Zoom V.16 and a Nikon SMZ25. DAPI overlays were generated in Photoshop CS6.

2.8. Double Fluorescent ISH (dFISH) in *P. tepidariorum*

P. tepidariorum embryos were staged and fixed as above. dFISH was performed following a protocol modified from Clark *et al.* [78]. Embryos were gradually moved from methanol to PBS-T and washed for 15 minutes twice in PBS-T. Embryos were then transferred to hybridization buffer, hybridized overnight at 65°C and washed post-hybridization as detailed in Prpic *et al.* [77]. 2 µL of each probe (DIG- and FITC-labelled) were used in the hybridization step. Incubation in AP-conjugated anti-DIG (1:2000; 11093274910, Roche) and POD-conjugated anti-FITC (1:2000; 11426346910, Roche) was carried out for two hours, following 30 minutes incubation in 1x Blocking solution[77]. Tyramide biotin amplification (TSA Plus Biotin Kit, NEL749A001KT, Perkin Elmer) was performed for 10 minutes, followed by incubation for 90 minutes in streptavidin Alexa Fluor 488 conjugate (1:500; S11223, ThermoFisher Scientific). AP signal was visualised by Fast Red staining (1/2 Fast Red tablet; 4210, Kem En Tec Diagnostics). Counterstaining with DAPI (1:2000; 10236276001, Roche) was carried out for 5-10 minutes.

Heads of stained embryos were dissected and flat mounted prior to imaging. Imaging was performed using a Zeiss LSM800 confocal with Airyscan. Photos were processed using the FIJI software[79].

2.9. Gene knockdown by RNA interference (RNAi)

2.9.1. Parental RNAi (pRNAi)

Five virgin adult female spiders were injected according to the protocol described in Akiyama-Oda & Oda [55]. Each spider was injected in the opisthosoma with 2 µl of dsRNA every 2 days, to a total of five injections. Two additional virgin adult females were injected with 2µl of dsRNA targeting GFP, as a negative control. A male was added to the vial for mating after the second injection. Embryos from injected females were fixed at stages 5 to 8.2 as described above. Embryos from GFP-injected control females were generated and treated as described above.

2.9.2. Embryonic RNAi (eRNAi)

Embryonic injections were carried out as described in Schönauer *et al*. [80] with minor changes. Embryos were injected between the 8- and 16-cell stages with a mix of 5 µl of FITC-dextran, 5 µl of biotin-dextran and 2.5 µl of dsRNA. Embryos were subsequently fixed at stages 5-8.2 of development. Visualization of eRNAi clones was achieved by detecting the co-injected biotin-dextran with the Vectastain ABC-AP kit (Vector Laboratories) after ISH, according to the manufacturer's protocol.

3. Chapter I: Outcomes of gene duplication among arachnid Homeobox genes

3.1. Introduction

The Homeobox gene superclass is a large gene family that is mostly comprised of transcription factors (e.g. human CERS2 and CERS6 encode for transmembrane proteins), which are characterised by an ancient and conserved DNA-binding domain: the homeodomain[81]. They bind to DNA and interact with protein cofactors to regulate gene expression[39]. In animals, homeobox genes are further divided into 11 classes: ANTP, PRD, LIM, POU, SINE, TALE, CUT, HNF, ZF, CERS and PROS[81]. These are further subdivided into over 100 gene families, each composed of closely related genes descending from at least one ancestral copy present in the common ancestor of bilaterians[39]. With important roles in development of all animals and in some cases underlying human disease when mutated[82,83], homeobox genes are one of the most studied gene classes in developmental biology.

3.1.1. Homeobox Genes in Animal Development

Homeobox genes are key regulators of animal development, with roles ranging from body segment identity to cell type differentiation. Therefore, correct spatial and temporal expression of these genes is essential throughout embryogenesis, as their main role is the activation of downstream batteries of target genes, which commit cells to a specific cell type, tissue or organ (e.g. neuron, muscle or eyes)[39,84]. An example of this is the case of *eyeless* (*ey*) and *sine oculis* (*so*) genes in *Drosophila*, which are crucial factors for the regulation of eye development[85–87]. Absence of expression of any of these genes leads to eye abnormalities or even complete eye loss[85–87]. Furthermore, ectopic expression of *ey* or *so* can lead to the formation of structures similar to compound eyes in the location where they were incorrectly expressed[85–87].

One of the most studied groups of homeobox genes is the Hox cluster, a set of genes responsible for the establishment of identity along the anterior-posterior (AP) body axis, a role that is conserved between all bilaterians[84,88,89]. Usually located in a single genomic cluster, the expression pattern of each Hox gene along the AP axis is reflective of their position within the cluster[90]. In other words, from the 3' end to 5' end of the cluster, each gene is expressed from the anterior-most to the posterior-most region of the embryo[88,89,91]. In addition to this spatial collinearity, vertebrate Hox genes also exhibit a temporal collinearity, with 3' Hox genes being expressed earlier and 5' Hox genes being expressed last[92]. The correct expression domain boundaries are maintained through Hox gene interactions, with more posteriorly expressed

genes supressing the expression and function of more anteriorly expressed genes[89,93,94]. The collinear nature of Hox gene expression is critical to provide correct structural identity along the AP axis of the animal body plan, and any mutation disrupting the function or expression of any of these genes results in the transformation of one body part into another, a phenotypic outcome termed homeotic shift or transformation[84].

One of the earliest described examples of a homeotic shift was identified in the first functional studies of the *Antennapedia* (*Antp*) gene of *Drosophila melanogaster*. The role of *Antp* is to specify the identity of the second thoracic segment (T2), and as such, its expression pattern is confined to this domain. Ectopic expression of *Antp* in the head activates a set of genes that confer T2 segment identity in that location, promoting the formation of T2-like legs in place of antennae[95]. Moreover, generation of clones lacking *Antp* expression in the adult leg induces the formation of antenna-like structures[89].

Vertebrate Hox genes have been shown to function in a similar way[88,94,96]. Differential AP expression of Hox genes correlates with specific characteristics of the vertebrate skeleton, such as the cervical to thoracic transition and forelimb formation, which are both correlated with the anterior boundary of *Hoxc6* expression in mice and chick[88]. Furthermore, ectopic expression of *Hoxb6* in the presomitic region in mice induces the formation of ribs along the whole length of the axial skeleton, including the normally ribless lumbar region[97].

3.1.2. Homeobox genes and animal evolution

The identification of conserved homeodomain-containing genes in distantly related animals sparked the idea of a developmental genetic toolkit, a set of genes encoding for transcription factors and signalling proteins, conserved between different species, and responsible for the developmental regulation of different body plans and morphologies[39,84]. This has become one of the main principles in the field of evo-devo. Furthermore, the conserved function of homeobox genes as regulators of embryonic development was first demonstrated by expressing vertebrate orthologs of Hox genes in the fruit fly[98,99]. Ectopic expression in *Drosophila* of a mouse *Antp* ortholog resulted in antenna to leg transformation, and a similar experiment using a human *Deformed* (*Dfd*) ortholog was able to copy previously observed homeotic effects[98,99].

These findings contributed greatly to the idea that changes in expression, rather than changes in protein function, are the main mechanism by which a conserved set of genes is able to regulate the development of diverse morphological features[39,84,100]. Once again, Hox genes were the most promising candidates to test this hypothesis. One of the first studies addressing this compared Hox gene expression patterns during embryogenesis between different vertebrate species[88]. Anterior expression boundaries of Hox genes were revealed to be

variable between species, and this variability was tightly correlated to changes in skeletal morphology[88]. For example, the number of cervical vertebrae in mouse and chick, which are 7 and 14 in total respectively, correlate with changes in the anterior boundary of Hoxc6[88].

A similar trend is observed in arthropods, where Hox expression boundaries are closely related to segmental specialization[39,89]. In crustaceans, shifting of the anterior border of Ultrabithorax (Ubx) expression is associated to the presence or absence of maxillipeds[101], and in chelicerates, position and number of breathing organs vary in concert with the anterior expression boundary of Abdominal-B (Abd-B)[44]. Examples are not restricted to members of the Hox cluster; deletion mutations identified in a cis-regulatory element that controls the expression pattern of *Pitx1*, have been shown to be responsible for the variation in pelvic fin morphology between freshwater and marine populations of the three-spined stickleback[102].

3.1.3. Impact of gene duplication in Homeobox gene evolution

Homeobox gene number has expanded throughout evolution, ranging from less than 10 in fungi to over 300 in teleost fish[39]. The current hypothesis is that a small number of proto-homeobox genes, present in the pre-metazoan ancestor, went through several rounds of tandem gene duplications to give rise to each metazoan homeobox class, which explains their cluster-like distribution in the genomes of extant animals[103,104]. Consistent with changes in Hox expression associated with changes in body plans, as discussed above, certain ancient episodes of homeobox expansion are intriguingly correlated with major changes in animal evolution. For example, the origin of bilateral symmetry and triploblasty is closely related to the emergence of the ParaHox, Hox and NK clusters[103,105], and multicellularity appears to be linked with the emergence of the ANTP class[103]. Lineage specific duplications are also associated with morphological innovation[106]; for example, a group of duplicated paralogues of *Hox3* in lepidopterans have seemingly acquired new functions in the formation of extraembryonic tissue[107].

Although less frequent, WGDs had also a major role in homeobox gene evolution. There was a high retention rate of duplicated homeobox genes after the two rounds of WGD in the common ancestor of vertebrates[39]. One consequence of these WGDs was the retention of duplicated gene clusters, including the Hox clusters[90]. It is now commonly accepted that the generation of multiple Hox clusters, followed by lineage specific divergence and differential gene retention, may have contributed to phenotypic divergence in the vertebrate lineage, such as axial skeleton variations, fin-to-limb transition and heart complexity[94,96,108].

3.1.4. Homeobox genes in arachnids

Homeobox genes have long been a focus in arachnid evo-devo studies. In fact, it was the observation of a high prevalence of homeobox gene duplication in spiders[67,68,109–111] and scorpions[112], especially the presence of two Hox clusters[113], which led to the discovery of the Arachnopulmonata WGD (Fig. 3.1 A)[30]. Attempts to identify the whole repertoire of homeobox genes has now been performed for two spiders (*P. tepidariorum* and *Pholcus phalangioides*), two scorpions (*Centruroides sculpturatus* and *Mesobuthus martensii*) and a tick (*Ixodes scapularis*)[72]. Between 50-59% of all homeobox families are duplicated in the spiders and scorpions, a striking increase when compared to the number of duplicated families in non-WGD arachnids like *I. scapularis* (24%) or other arthropods like the centipede *Strigamia maritima* (19%)[72].

The expression patterns of Hox genes have also been studied in several arachnid species, including the spider *P. tepidariorum* and the harvestmen *P. opilio*[30,61,89,111,114]. As expected, these are collinearly expressed along the AP axis, with the first six genes of each cluster patterning the anterior prosoma (*labial*, *proboscipedia*, *Hox3*, *Dfd*, *Sex combs reduced* and *fushi tarazu*), and the remaining four pairs of duplicates patterning the posterior opisthosoma (*Antp*, *Ubx*, *abdominal-A* and *Abd-B*) (Fig. 3.1 B)[44]. Expression boundaries of prosomal Hox genes are apparently conserved between all the species studied so far, which is reflective of the highly conserved structure of this tagma within Arachnida[61]. However, for the opisthosomal Hox genes, a different pattern was observed, with the anterior boundaries of these genes found to be variable between species of different orders, a result that closely correlates with lineage specific differences in opisthosomal morphology (e.g. position of breathing organs) (Fig. 3.1 B)[114]. Functional analysis in *P. tepidariorum* has been carried out for *lab-A*, *Dfd-A*, *Antp-A* and *Ubx-A*. Knockdown of *lab-A* results in the loss of pedipalps, whereas loss of *Dfd-A* expression leads to the transformation of the L1 segment towards a pedipalpal identity[115]. As for *Antp-A* and *Ubx-A*, both seem to have a role in repressing leg formation in opisthosomal segments[116].

Outside the Hox cluster, the expression and function of other duplicated homeobox gene families has mainly been characterised in regard to their roles in segmentation, eye development and evolution of appendage morphology[44]. Genes belonging to the Pax3/7 and En families display expression patterns similar to their insect homologs, suggestive of their functional conservation as segmentation genes[61,109,117]. Functional analysis of *Pt-Otx1* has revealed its role as an organiser of anterior regionalization in spider embryogenesis[58]. Expression patterns of Otx, Pax4/6, Six1/2 and Six3/6 family members have been analysed with respect to eye development in two spiders and a harvestmen, providing some insight on the evolution of arachnid eye development[60,67,68]. Finally, expression patterns of genes from the Barhl, Tlx,

Arx, Lhx1/5, Meis and Pbx families have been analysed during arachnid appendage development, revealing both conserved and novel elements in the gene network regulating this process when compared with other arthropods[59,62,110,118].

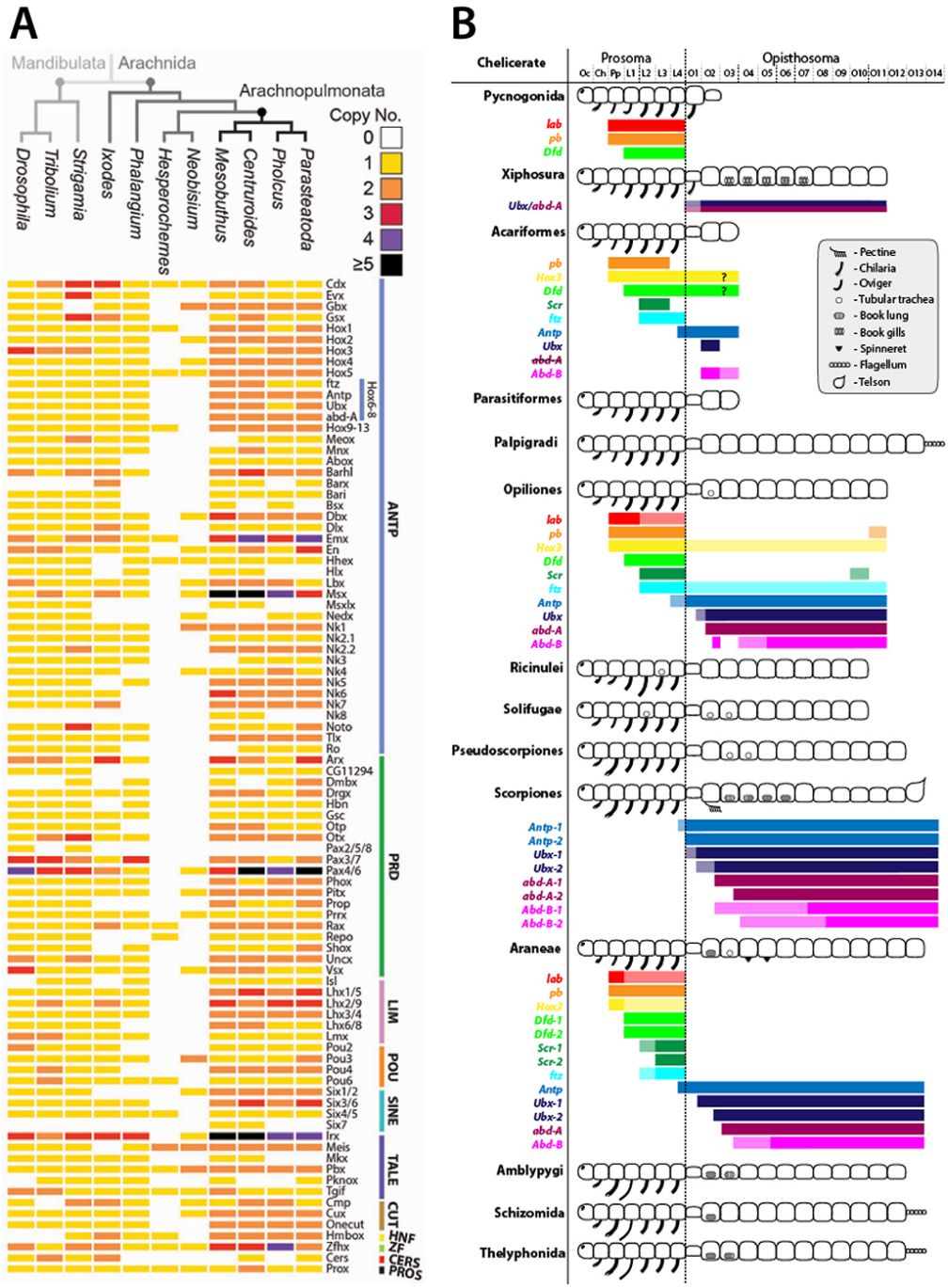

Figure 3.1: Homeobox repertoires in Arthropods and Hox cluster gene expression in Chelicerates. (A) Homeobox gene number by gene family for three Mandibulates and eight Arachnids. Adapted from Leite *et al.* [72]. There is a clear increase in gene number in Arachnopulmonata when compared to other Arthropods. Note that the *Phalangium* data is a result from this thesis. (B) Hox gene expression in relation to the variable body plans of Chelicerates. Variation in the anterior expression boundaries of posterior Hox genes closely correlates with posterior segment identity, such as the position of breathing organs (e.g. trachea and book lungs) and spinnerets. Image adapted from Schwager *et al.* [44].

Therefore, considerable work has already been carried out on duplicated homeobox genes of spiders and scorpions, although, a more systematic approach to address the impact of functional divergence upon WGD, and its role during arachnid evolution, remains to be done. For this purpose, I aimed to investigate the extent of expression divergence of homeobox ohnologs by characterising the expression patterns of duplicated homeobox genes in the spider *P. tepidariorum* and their single-copy orthologues in the harvestmen *P. opilio*. A copy of the article where this work was published is provided as an appendix.

3.2. Results

3.2.1. Identification of Homeobox genes in the harvestmen *P. opilio*

To characterise the homeobox gene repertoire of the harvestmen *P. opilio*, I surveyed the available transcriptome for homeodomain-containing sequences. A total of 69 genes belonging to 65 homeobox gene families were identified (Fig. 3.1 A). These numbers are considerably smaller than those previously found for two spiders and two scorpions in our lab (*P. tepidariorum* – 145/80, *P. phalangioides* – 132/78, *C. sculpturatus* – 156/82, and *M. martensii* – 156/82)[72]. However, this is likely to be an incomplete set of homeobox genes in *P. opilio*, because the total number of genes is substantially lower than that of other non-WGD arthropod species (*I. scapularis* – 96/70, *S. maritima* – 113/83, *Tribolium castaneum* – 105/80, and *D. melanogaster* – 104/80)[72].

Furthermore, there is a huge discrepancy in the number of duplicated gene families when compared to other non-WGD arthropods (only 3% in *P. opilio*, compared to 24% in *I. scapularis* and 19% in *S. maritima*)[72]. Pax3/7 and Irx were the only gene families for which duplicated sequences were found. This is very likely to be due to incomplete transcriptomic data and/or fragmented sequence data. Many of the transcripts identified were partial sequences, sometimes extending not much further than the sequence coding for the homeodomain. Therefore, some transcripts might have been missed if the sequenced fragment did not extend to the homeodomain.

Nevertheless, I was able to identify a significant number of homeobox gene sequences in the available transcriptome of *P. opilio*, providing me with enough sequence data for the construction of the ISH RNA probes needed for the subsequent gene expression analysis. Note that data produced from this analysis was also used in the construction of a phylogenetic tree by Daniel Leite that included several arthropod homeobox gene repertoires, and provided further validation to the identity of each homeobox sequence from *P. opilio*[72].

3.2.2. Expression pattern analysis of duplicated *P. tepidariorum* homeobox genes and *P. opilio* single-copy orthologs

To begin to infer the extent of divergence in the roles of homeobox genes following the arachnopulmonate WGD, I characterised the expression patterns of duplicated genes in *P. tepidariorum* and their corresponding single-copy orthologs in *P. opilio* by means of ISH. I focused mainly on gene families that lacked previously published expression pattern data. Nonetheless, I have also analysed gene families for which expression data was only available for one duplicate or is restricted to a specific domain or stage. In addition to this, I performed ISH on single-copy orthologs of three families previously analysed in the lab for *P. tepidario-*

rum, allowing me to infer the possible ancestral state of these duplicated gene families and ascertain what type of expression divergence the previously characterised expression patterns underwent.

The homeobox gene families selected were Gbx, Hox3, Dbx, Emx, Msx, Nk1 and Nk2.2 of the ANTP class, Otx and Pitx of the PRD class, Lhx2/9 of the Lim class, Irx of the TALE class, Cux of the CUT class and Zfh of the ZFH class. Note that some of these genes arose through tandem gene duplication that likely preceded WGD, namely Emx, Lhx2/9 and Irx genes[72]. Embryos ranging from stage 5 to stage 12 were used for *P. tepidariorum* and from stage7 to stage 17 for *P. opilio*, allowing a broad overview of homeobox gene expression during the stages of embryonic development where most cell types are specified, which is the major function of this family of transcription factors. A result summary is provided in Table S5.

3.2.2.1. Expression patterns of *Gbx* genes

Homologs of the Gbx homeobox gene family usually play a role in the development of the nervous system[119–121]. More specifically, *Gbx* appears to have a conserved role in establishing the midbrain/hindbrain boundary in both insects and vertebrates[119,120]. In *D. melanogaster*, the Gbx family member *unplugged* also plays a role in the development of the tracheal system[122].

Two copies of *Gbx* were previously identified in the *P. tepidariorum* (*Pt-Gbx1* and *Pt-Gbx2*) and I was able to identify one homologous sequence in *P. opilio* (*Po-Gbx*). *Po-Gbx* expression is first observed at stage 13 in the prosomal appendages and neuroectoderm (Fig. 3.2 A-A'''). Expression in the neuroectoderm is confined to small groups of cells along its ventral region (Fig 3.2 A''). Two more groups of cells were observed in the pre-cheliceral region (Fig. 3.2 A'). Expression in the walking legs and pedipalps is restricted to a posterior and an anterior row of ectodermal cells (Fig. 3.2 A). Expression is also present at the base of the chelicerae (Fig. 3.2 A').

ISH with a *Pt-Gbx1* probe produced no distinguishable signal, suggesting this gene has lost embryonic expression. *Pt-Gbx2* expression was observed in the prosomal appendages in a similar pattern of that of *Po-Gbx* (Fig 3.2 B and B''). However, neuroectoderm expression appeared to be missing from *Pt-Gbx2* compared to *Po-Gbx* (Fig. 3.2 B'). Interestingly, a potential novel expression domain was seen at the base of each opisthosomal limb bud (Fig. 3.2 B''').

In summary, the single-copy expression pattern observed in *Po-Gbx* appears to be conserved in *Pt-Gbx2*, with the exception of the neuroectoderm expression domain, and *Pt-Gbx1* appears not to be expressed during embryogenesis. Additionally, *Pt-Gbx2* acquired a potentially novel expression domain in the opisthosomal organs.

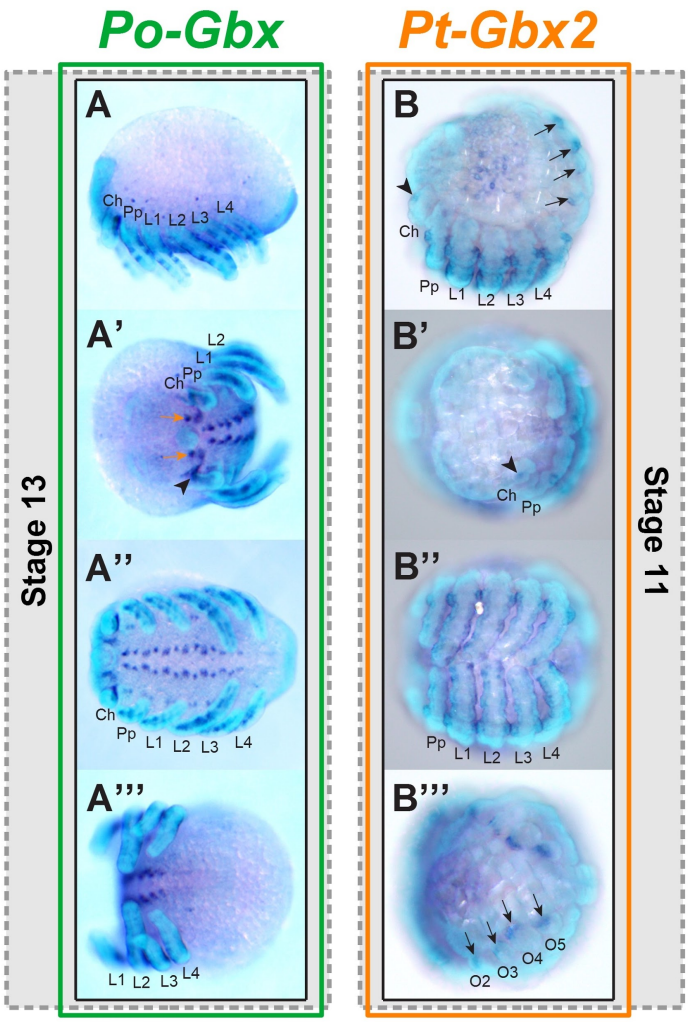

Figure 3.2: Expression patterns of *P. opilio* and *P. tepidariorum Gbx* genes. (A-A''') Expression of *Po-Gbx* in a stage 13 embryo. (B-B''') Expression of *Pt-Gbx2* in a stage 11 embryo. Anterior is to the left in all images. Arrowheads mark expression at the base of the chelicerae. Arrows mark expression of *Po-Gbx* in the pre-cheliceral region (orange) and expression of *Pt-Gbx2* in the opisthosomal organs (black). Ch, chelicerae; Pp, pedipalps; L1-L4, walking legs 1-4.

3.2.2.2. Expression patterns of *Hox3* genes

Unlike other members of the Hox cluster, *Hox3* genes have highly divergent functions throughout metazoans[89,94]. In vertebrates, *Hox3* genes are involved in limb, cardiovascular and pharyngeal organ development [123–125]. However, insect *Hox3* genes diverge substantially from vertebrates, with roles in segmentation and the formation of extraembryonic tissue[89,107].

The expression patterns of *Po-Hox3* and one out of two ohnologs in *P. tepidariorum* (*Pt-Hox3b*) have been previously studied[30,61]. However, expression of *Pt-Hox3a* remained to be characterised. *Po-Hox3* and *Pt-Hox3b* have very similar patterns, with expression in the pedipalps and walking legs, and an additional domain in the ventral neuroectoderm in later embryonic stages (Fig. 3.3 B-D')[30,61].

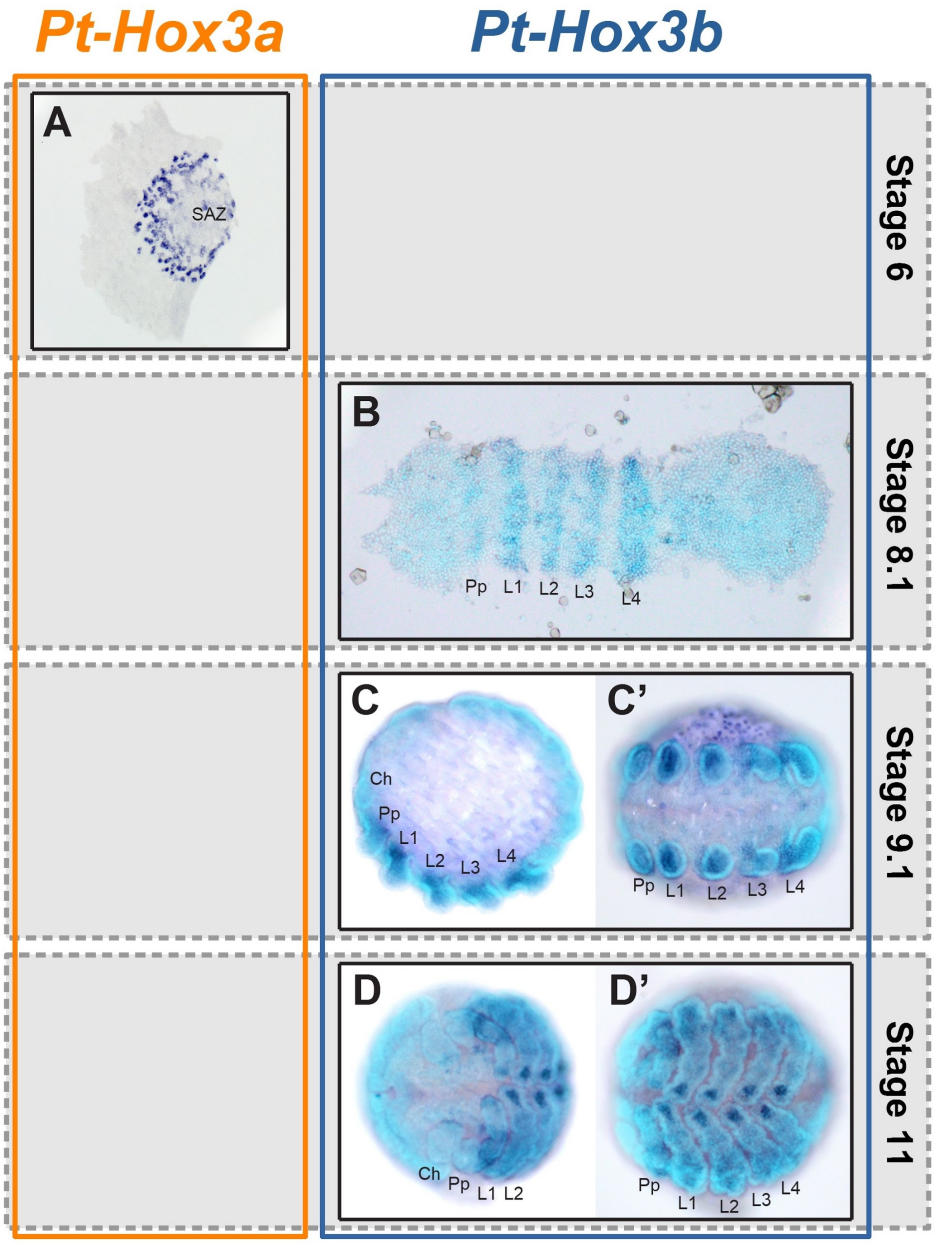

Figure 3.3: Expression patterns of *P. tepidariorum Hox3* genes. (A) Expression pattern of *Pt-Hox3a* in a stage 6 embryo (flatmount). (B-D') Expression pattern of *Pt-Hox3b* in stage 8.1 (B) (flatmount), 9.1 (C, C') and 11 (D, D') embryos. Anterior is to the left in all images. Ch, chelicerae; Pp, pedipalps; L1-L4, walking legs 1-4; SAZ, segment addition zone.

Pt-Hox3a is mainly expressed at stage 6, forming a band of salt-and-pepper-like expression surrounding the area corresponding to the presumptive posterior of the embryo (Fig. 3.3 A), which fades away during stage 7. This expression pattern appears to be specific to *Pt-Hox3a*, since *Pt-Hox3b* expression is only detected from stage 8.1 onwards (Fig. 3.3 B). However, it is possible *Po-Hox3* is also expressed in this domain, as analysis of *Po-Hox3* expression on similar stages during *P. opilio* embryogenesis was not performed. As it stands, the *Hox3*

ohnologs of *P. tepidariorum* either represent a partition of ancestral expression patterns, or *Pt-Hox3a* has gained a novel domain in earlier stages of development.

3.2.2.3. Expression patterns of *Dbx* genes

Dbx homeobox family members play roles in embryonic development of the central nervous system in both insects and vertebrates[126,127]. In *D. melanogaster*, *Dbx* controls the specification of a particular neuronal cell subtype[127].

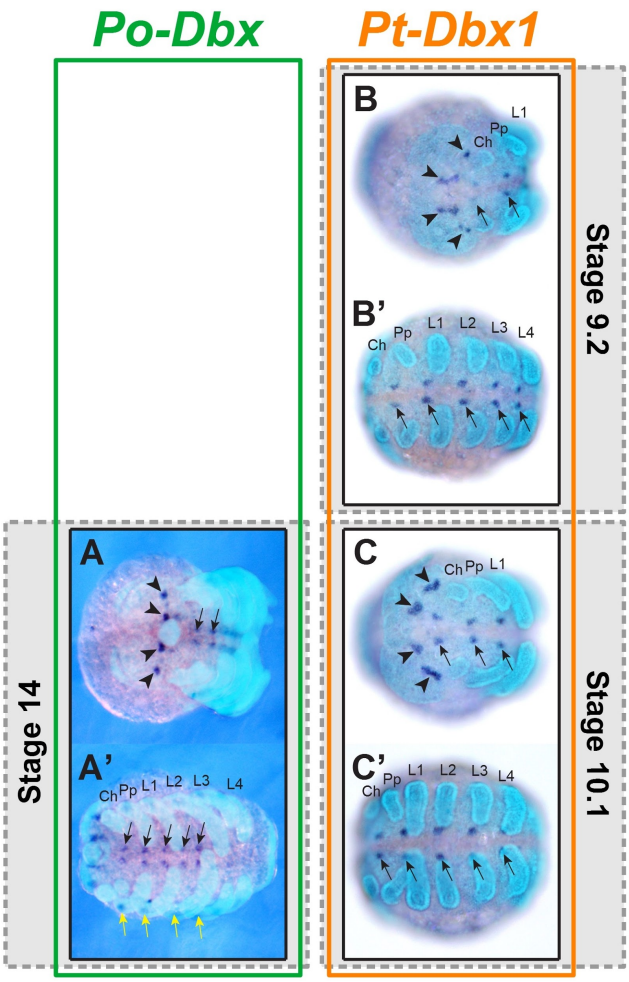

Figure 3.4: Expression patterns of *P. opilio* and *P. tepidariorum Dbx* genes. (A-A') Expression pattern of *Po-Dbx* in a stage 14 embryo. (B-C') Expression pattern of *Pt-Dbx1* in stage 9.2 (B, B') and 10.1 (C, C') embryos. Anterior is to the left in all images. Arrowheads mark expression in the pre-cheliceral region. Arrows mark expression in the ventral neuroectoderm (black) and *Po-Dbx* expression at the base of prosomal appendages (yellow). Ch, chelicerae; Pp, pedipalps; L1-L4, walking legs 1-4.

Analysis of *P. opilio* and *P. tepidariorum Dbx* homologues revealed a scenario similar to that of the Gbx family. *Po-Dbx* expression is confined to the base of the pedipalps and walking legs, and clusters of cells in the developing neuroectoderm (Fig. 3.4 A-A'). Expression in the

neuroectoderm is restricted to four clusters of cells in the pre-cheliceral region (Fig. 3.4 A) and two clusters of cells per prosomal segment in the ventral neuroectoderm (Fig. 3.4 A').

The same neuroectodermal expression pattern is seen for *Pt-Dbx1* (Fig. 3.4 B-C'), although, the domain of *Po-Dbx* expression at the base of the prosomal appendages was not observed for *Pt-Dbx1* (Fig. 3.4 B and C'). ISH for *Pt-Dbx2* produced no visible signal during embryogenesis.

These results suggest that *Po-Dbx* expression in the neuroectoderm is conserved in *Pt-Dbx1*, and that *Pt-Dbx2* appears not to be expressed during embryogenesis. Expression in the prosomal appendages was only detected in *Po-Dbx*, representing either a loss of this expression domain in both copies of *P. tepidariorum* or a lineage-specific gain in the harvestmen single-copy. Nevertheless, it appears that the involvement of *Dbx* genes in central nervous system development is conserved in both *P. opilio* and *P. tepidariorum*.

3.2.2.4. Expression patterns of *Emx* genes

Emx genes are involved in several developmental processes in both vertebrates and insects, although, they appear to be mainly conserved in respect to their roles in olfactory and nervous system development[128–130]. In *D. melanogaster*, the Emx family member *empty spiracles* is also involved in the tracheal system development[131].

There are a total of four *Emx* genes in *P. tepidariorum* genome, although, I was only able to identify one gene in *P. opilio*, *Po-Emx*, whose expression was previously reported[60]. However, here I provide a more detailed description. *Po-Emx* is first expressed during stage 10 in all prosomal appendages and in stripes in the forming opisthosomal segments (Fig. 3.5 A and A'''). The expression extends distally from base to the middle of each prosomal appendage (Fig. 3.5 A). At stage 13, expression in segmental clusters of cells is observed in the neuroectoderm (Fig. 3.5 B'), and in a group of cells in the pre-cheliceral region (Fig. 3.5 B). Expression in the prosomal appendages is then subdivided into a ring-like pattern and at the base of each appendage (Fig. 3.5 B). Segmental stripes are still present in the opisthosomal segments at stage 13 (Fig. 3.5 B").

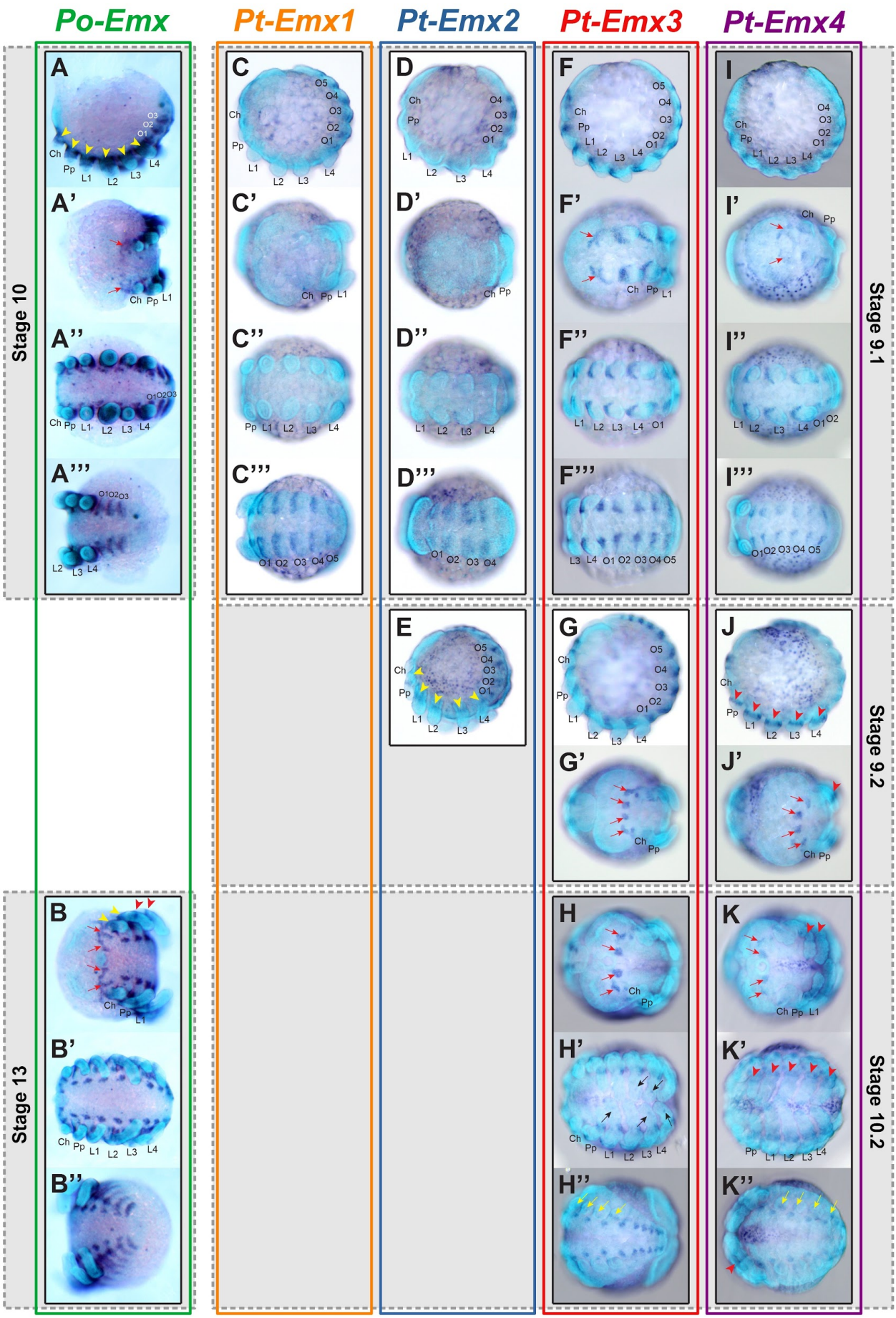
Po-Emx
Pt-Emx1
Pt-Emx2
Pt-Emx3
Pt-Emx4
Stage 10
Stage 13
Stage 9.1
Stage 9.2
Stage 10.2
A
A'
A''
A'''
B
B'
B''
C
C'
C''
C'''
D
D'
D''
D'''
E
F
F'
F''
F'''
G
G'
H
H'
H''
I
I'
I''
I'''
J
J'
K
K'
K''

Figure 3.5 (previous page): Expression patterns of *P. opilio* and *P. tepidariorum Emx* genes. (A-B'') Expression pattern of *Po-Emx* in stage 10 (A-A''') and 13 (B, B') embryos. (C-C''') Expression pattern of *Pt-Emx1* in a stage 9.1 embryo. (D-E) Expression pattern of *Pt-Emx2* in stage 9.1 (D-D''') and 9.2 (E) embryos. (F-H'') Expression pattern of *Pt-Emx3* in stage 9.1 (F-F'''), 9.2 (G, G') and 10.2 (H-H'') embryos. (I-K'') Expression pattern of *Pt-Emx4* in stage 9.1 (I-I'''), 9.2 (J, J'') and 10.2 (K-K'') embryos. Anterior is to the left in all images. Arrowheads mark expression at the base (yellow) and a ring-like domain (red) of prosomal appendages. Arrows mark expression in the pre-cheliceral region (red), *Pt-Emx3* dot-like expression in prosomal appendages (black), and *Pt-Emx3* and *Pt-Emx4* expression in the opisthosomal organs (yellow). Ch, chelicerae; Pp, pedipalps; L1-L4, walking legs 1-4; O1-O5, opisthosomal segments 1-5.

The *P. tepidariorum Emx* orthologues have overlapping domains, showing subdivision of the single-copy expression domains detected in *Po-Emx*. Early segmental stripe expression in the opisthosoma is conserved among all paralogs, and is the only domain in which *Pt-Emx1* is expressed (Fig. 3.5 C''', D''', F''' and I'''). Neuroectoderm expression is conserved between *Pt-Emx3* and *Pt-Emx4*, although the former appears to be stronger (Fig. 3.5 F-H'' and I'-K''). Prosomal appendage expression domains are subdivided between *Pt-Emx2*, *Pt-Emx3* and *Pt-Emx4*. *Pt-Emx2* expression was detected at the base of each appendage (Fig. 3.5 E), whilst *Pt-Emx3* and *Pt-Emx4* show ring-like expression patterns (Fig. 3.5 G-H and J-K''). Putatively novel domains of expression are visible in the opisthosomal limb buds for *Pt-Emx3* and *Pt-Emx4* (Fig. 3.5 H'' and K''), as well as a dotted expression of *Pt-Emx3* in the pedipalps and walking legs (Fig. 3.5 H').

Overall, the expression pattern observed for *Po-Emx* appears to have been subdivided between the four *Emx* ohnologs found in *P. tepidariorum*, along with few possibly novel expression domains found in *Pt-Emx3* and *Pt-Emx4*.

3.2.2.5. Expression patterns of *Msx* genes

The Msx homeobox gene family is composed of transcriptions factors with highly conserved roles in muscle and neuronal development of animals[132,133]. However, vertebrate ohnologs of *Msx* acquired additional roles in other developmental processes, such as craniofacial, branchial arch, limb and mammary gland development[134–136].

A single copy of *Msx* is present in *P. opilio* and three copies are present in *P. tepidariorum*. At stage 7, *Po-Msx* expression was observed as segmental stripes (Fig. 3.6 A). At later stages, expression becomes restricted to the dorsal region of the neuroectoderm and a dot-like expression along the ventral midline (Fig. 3.6 C' and D'). Expression is also present at the base of each prosomal appendage, which could possibly be associated to developing muscle tissue (Fig. 3.6 C-D). The stripe-like segmental expression can still be seen in opisthosomal segments at stage 16 (Fig. 3.6 D').

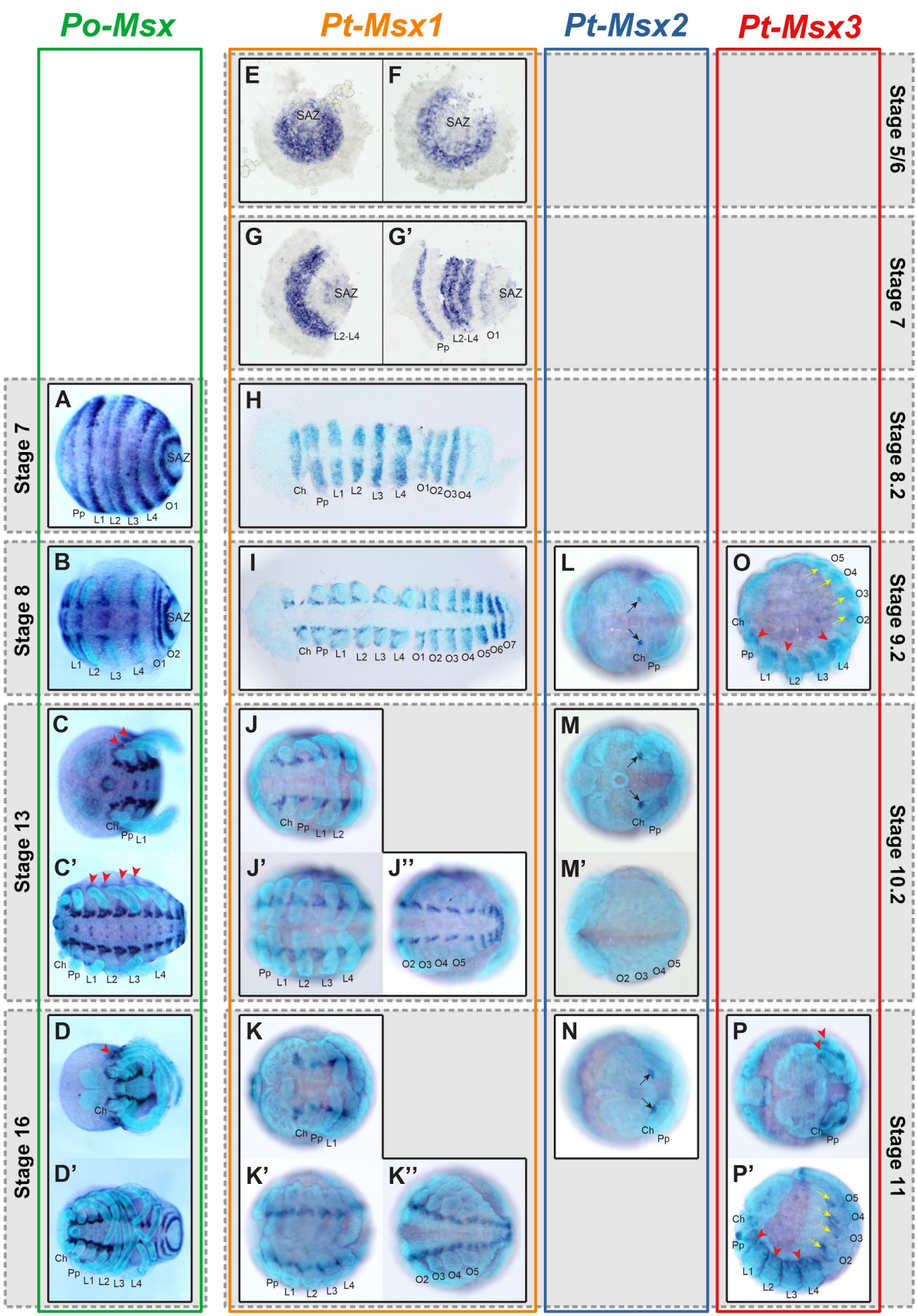
Po-Msx
Pt-Msx1
Pt-Msx2
Pt-Msx3
Stage 5/6
Stage 7
Stage 8.2
Stage 9.2
Stage 10.2
Stage 11
Stage 7
Stage 8
Stage 13
Stage 16
A
B
C
C'
D
D'
E
F
G
G'
H
I
J
J'
J''
K
K'
K''
L
M
M'
N
O
P
P'
SAZ
Ch
Pp
L1
L2
L3
L4
L2-L4
O1
O2
O3
O4
O5
O6
O7

Figure 3.6 (previous page): Expression patterns of *P. opilio* and *P. tepidariorum Msx* genes. (A-D') Expression pattern of *Po-Msx* in stage 7 (A), 8 (B), 13 (C, C') and 16 (D, D') embryos. (E-K'') Expression pattern of *Pt-Msx1* in stage 5 (E) (flatmount), 6 (F) (flatmount), 7(G. G') (flatmount), 8.2 (H), 9.1 (I) (flatmount), 10.1 (J-J'') and 11 (K-K'') embryos. (L-N) Expression pattern of *Pt-Msx2* in stage 9.2 (L), 10.2 (M, M') and 11 (N) embryos. (O-P') Expression pattern of *Pt-Msx3* in stage 9.2 (M) and 11 (N, N') embryos. Anterior is to the left in all images except E and F. Arrowheads mark expression in presumptive developing muscle. Arrows mark *Pt-Msx2* expression in the chelicerae (black) and *Pt-Msx3* expression in the opisthosomal organs (yellow). SAZ, segment addition zone; Ch, chelicerae; Pp, pedipalps; L1-L4, walking legs 1-4; O1-O7, opisthosomal segments 1-7.

The expression patterns of *Pt-Msx1*, *Pt-Msx2* and *Pt-Msx3* do not appear to overlap. Expression of *Pt-Msx1* starts at stage 5 in the centre of the germ disc (Fig. 3.6 E). From stages 6 to 8.2, the pattern becomes segmental, much like the expression observed at equivalent stages for *Po-Msx* (Fig. 3.6 F-H). At later stages, expression becomes restricted to the same neuroectodermal domain observed for the single-copy of *P. opilio* (Fig. 3.6 I-K''). *Pt-Msx2* expression is restricted to a cluster of cells at the mid-ventral region of the chelicerae that is first visible at stage 9.2 (Fig 3.6 L-N), an expression domain not detected for *Po-Msx*. Expression at the base of the prosomal appendages, similar to the one observed for *Po-Msx*, was detected for *Pt-Msx3* (Fig. 3.6 O-P'). *Pt-Msx3* expression is also present in the opisthosomal limb buds (Fig. 3.6 O-P').

In conclusion, *Po-Msx* is expressed in the developing nervous system and presumptive developing muscle, as expected of the known roles of this homeobox family in the development of these tissues. These two expression domains were partitioned between the spider ohnologs *Pt-Msx1* and *Pt-Msx3*, possibly representing a subdivision of the respective putative functions. In addition to this, I detected an early segmental-like expression pattern in *Po-Msx*, which was conserved in *Pt-Msx1*, hinting at a possible role of this gene in segmentation. Lastly, *Pt-Msx2* shows an interesting restricted domain of expression in the chelicerae, which could be associated to the development of an organ or structure specific to this appendage (e.g. venom gland).

3.2.2.6. Expression patterns of *Nk1* (*Slouch*) genes

Members of the Nk1 homeobox gene family appear to have different roles in insects and vertebrates. In *D. melanogaster*, the Nk1 family gene *Slouch* is necessary for the specification of a subset of muscle progenitor cells[137], while expression pattern analysis of *Nk1* genes in zebrafish is suggestive of a role in neurogenesis[138].

There are a total of two *Slouch* orthologs in *P. tepidariorum* and one in *P. opilio*. ISH for *Po-Slou* produced no visible staining, suggesting it is not expressed during embryogenesis. Of the two *P. tepidariorum* paralogs, only *Pt-Slou1* shows expression during embryonic stages. Expression was first visible at stage 9.1, in the mesendodermal layer of the developing pro-

somal appendages (Fig. 3.6 A). Later, this pattern remains unchanged in the chelicerae (Fig. 3.6 C and D), but expression in the pedipalps and walking legs becomes restricted to the base and two rows of cells from the middle to the tip of the appendages (Fig. 3.6 C', D' and D'''). At stage 12, small clusters of expression can be seen in each opisthosomal limb bud (Fig. 3.6 D'' and D''').

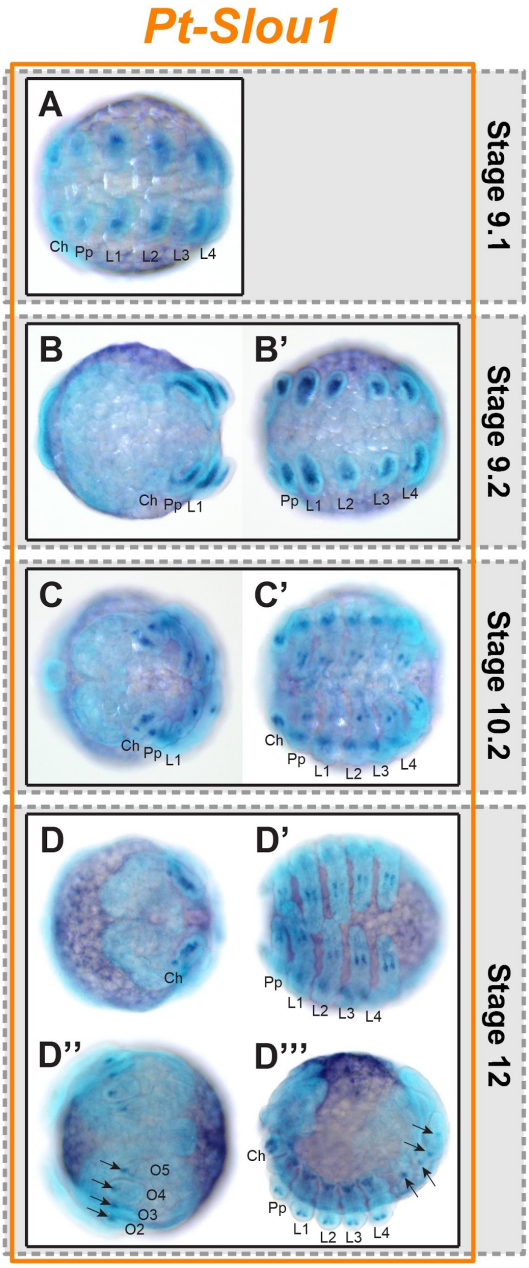

Figure 3.6: Expression pattern of *P. tepidariorum Slou1* gene. (A-D''') Expression pattern of *Pt-Slou1* in stage 9.1 (A), 9.2 (B, B'), 10.2 (C, C') and 12 (D-D''') embryos. Anterior is to the left in all images. Arrows mark expression in the opisthosomal organs. Ch, chelicerae; Pp, pedipalps; L1-L4, walking legs 1-4; O2-O5, opisthosomal segments 2-5.

These results suggest a role of *Pt-Slou1* in the development of both prosomal and opisthosomal appendages, possibly specifying muscle tissue associated with these structures, while neither *Po-Slou* nor *Pt-Slou2* appear not to be expressed during embryogenesis.

3.2.2.7. Expression patterns of *Nk2.2* (*Vnd*) genes

Genes of the Nk2.2 homeobox family have conserved roles in the development of the nervous systems of insects and vertebrates[133,139]. In *D. melanogaster*, the Nk2.2 family member *ventral nervous system defective* (*vnd*) is involved in the dorsoventral patterning of the neuroectoderm, as well as the formation and specification of ventral neuronal cell lineages[133]. A similar role was observed for the *Nk2.2* orthologs in mice, where they play a role in floor plate development (i.e. the ventral-most structure of the neural tube) and commissural axon guidance[139].

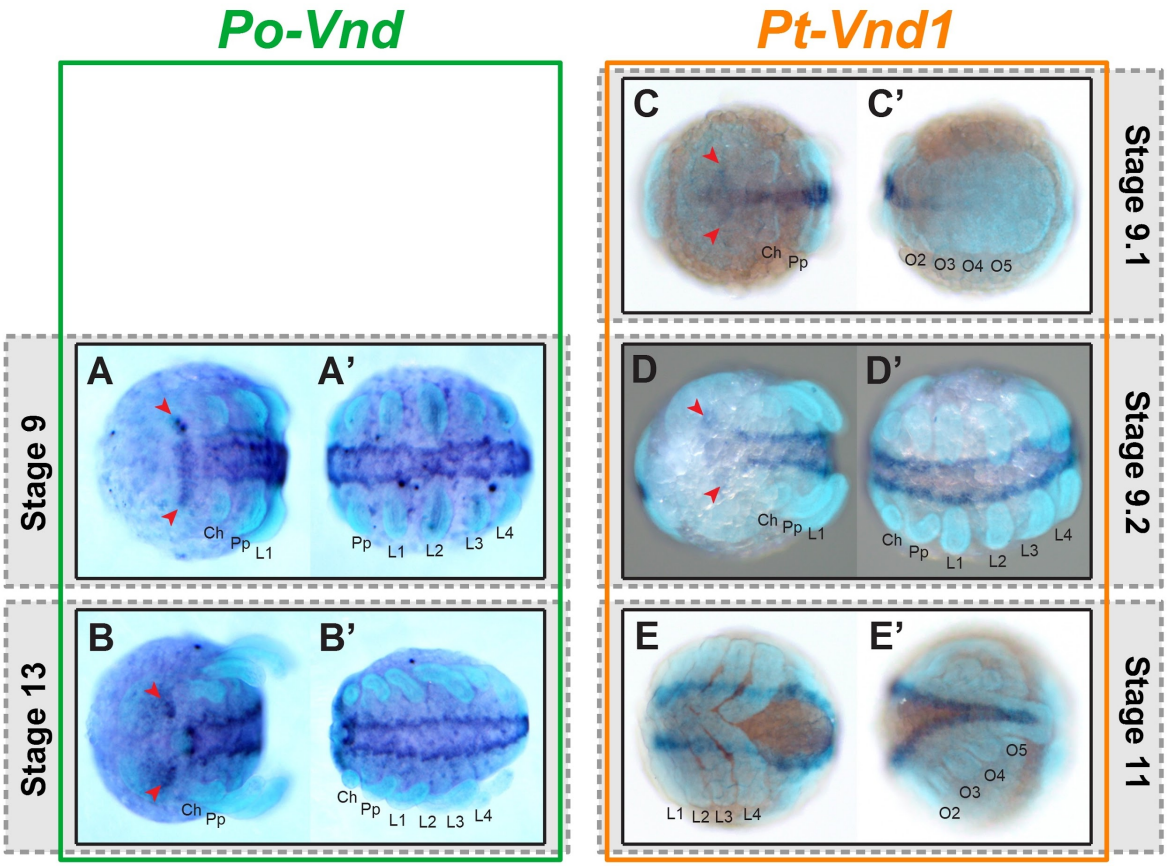

Figure 3.7: Expression patterns of *P. opilio* and *P. tepidariorum Vnd* genes. (A-B') Expression pattern of *Po-Vnd* in stage 9 (A, A') and 13 (B, B') embryos. (C-E') Expression pattern of *Pt-Vnd1* in stage 9.1 (C, C'), 9.2 (D, D') and 11 (E, E') embryos. Anterior is to the left in all images. Arrowheads mark expression in the pre-cheliceral region. Ch, chelicerae; Pp, pedipalps; L1-L4, walking legs 1-4; O2-O5, opisthosomal segments 2-5.

At stage 9, the single-copy of *P. opilio*, *Po-Vnd*, is mainly expressed in the ventral region of the neuroectoderm (Fig. 3.7 A-B') and in a stripe in the pre-cheliceral region (Fig. 3.7 A). Lat-

er at stage 13, the pre-cheliceral expression becomes subdivided into two cell clusters in the head lobes (Fig. 3.7 B). *Pt-Vnd1* expression completely recapitulates the expression of the single-copy gene (Fig. 3.7 C-E'). ISH for *Pt-Vnd2* produced no clear signal, suggesting this gene has lost its embryonic expression.

My results suggest a conserved role of *Po-Vnd* and *Pt-Vnd1* in patterning the ventral region of the neuroectoderm, with *Pt-Vnd2* apparently not being expressed during embryogenesis, presumably having a role during juvenile of adult stages.

3.2.2.8. Expression patterns of *Otx/otd* genes

Otx/otd genes are mainly characterised by their roles in eye and nervous system development, as well as segmentation[130,140,141]. The Otx homeobox gene family is represented by one copy in *P. opilio* where it is expressed along the ventral midline, two clusters of cells in the head lobes and along the edge of the pre-cheliceral region, which extends into two clusters of expression in the eye fields[60].

Functional analysis of *Pt-Otx1*, one of the two orthologs found in *P. tepidariorum*, demonstrated the role of this gene in head segmentation of this spider[58]. Furthermore, expression analysis of *Pt-Otx2* in the developing head of *P. tepidariorum* is suggestive of a role in eye development[67]. While the expression patterns of these genes have been previously characterised[58,67], here I provide a more detailed comparison of their expression patterns.

Pt-Otx1 expression is first detected at stage 5 at the rim of the germ disc[58]. At stage 7, expression becomes restricted to the anterior edge of the germ band (Fig. 3.8 A). At stage 8.2, this expression domain now forms a band just above the cheliceral segment (Fig. 3.8 B and B'). At stage 9.1, the band of expression becomes restricted to two clusters of cells in the centre of each head lobe and *Pt-Otx1* expression is also present in the ventral midline (Fig. 3.8 C and C'). This pattern is maintained is still present in the next two stages but at much lower levels (Fig. 3.8 D-E'), fading away at later stages. I was also able to detect an additional domain of expression in the opisthosomal limb buds from stage 9.2 onwards (Fig. 3.8 D' and E').

Pt-Otx2 expression is first detected at stage 8.2, in the same pre-cheliceral band of expression as *Pt-Otx1* (Fig.3.8 F and F'). At stage 9.1, expression in the pre-cheliceral region separates into two distinguishable groups of cells in each head lobe, with one pair corresponding to the same domain observed for *Pt-Otx1* at this stage, and the other in the presumptive region of the median eyes (Fig. 3.8 G). Similarly to *Pt-Otx1*, ventral midline expression is first detected at stage 9.1 (Fig. 3.8 G'). Expression in the head lobes and ventral midline is maintained throughout embryogenesis at high levels, as well as the expression domain in the median eyes (Fig. 3.8 H-J').

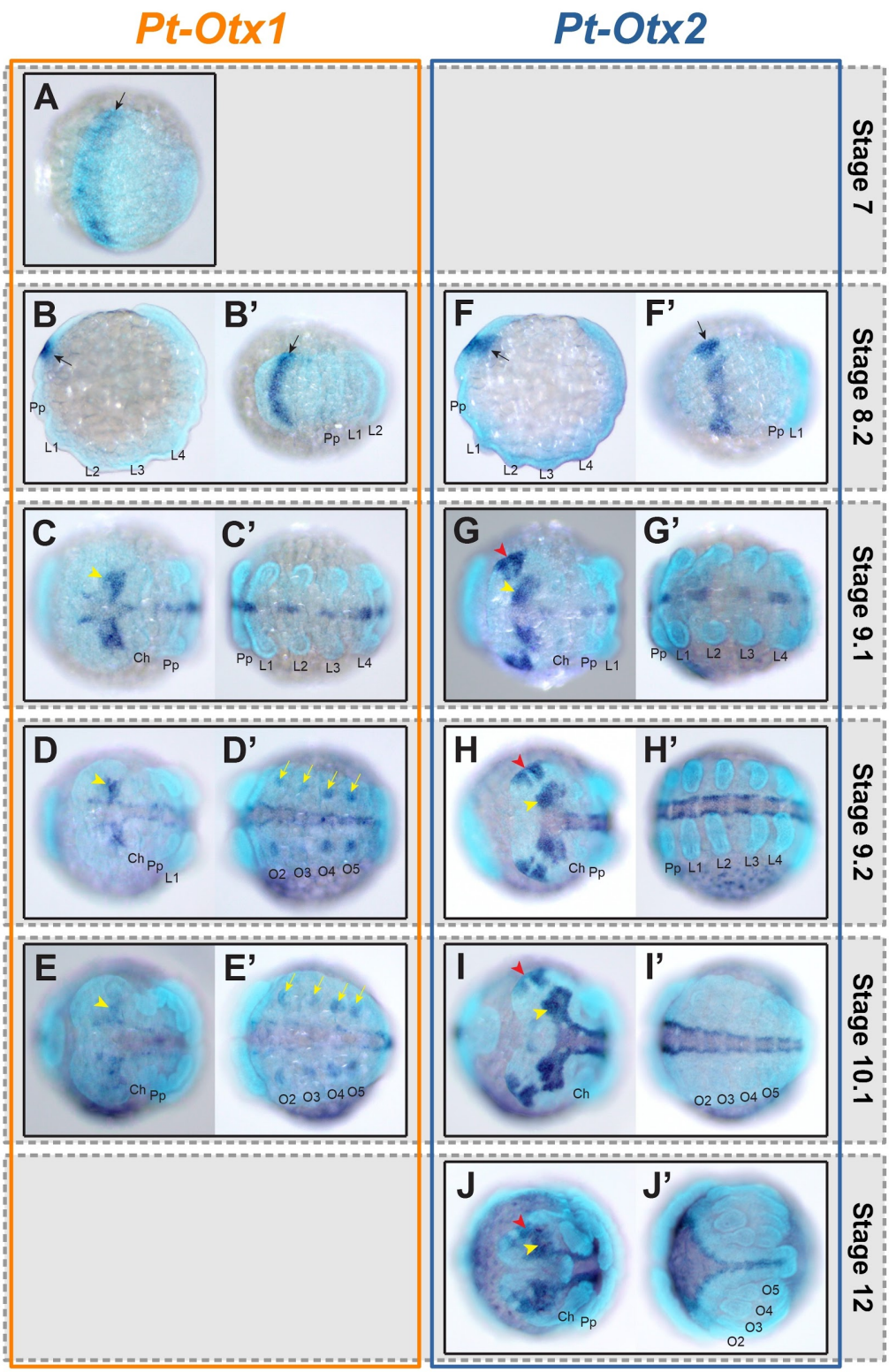

Figure 3.8: Expression patterns of *P. tepidariorum Otx* genes. (A-E') Expression pattern of *Pt-Otx1* in stage 7 (A), 8.2 (B, B'), 9.1 (C, C'), 9.2 (D, D') and 10.1 (E, E') embryos. (F-J') Expression pattern of *Pt-Otx2* in stage 8.2 (F, F'), 9.1 (G, G'), 9.2 (H, H'), 10.1 (I, I') and 12 (J, J') embryos. Anterior is to the left in all images. Arrows mark early anterior band of expression (black) and expression in the opisthosomal organs (yellow). Arrowheads mark expression in the head lobes (yellow) and *Pt-Otx2* expression in the presumptive median eye primordia (red). Ch, chelicerae; Pp, pedipalps; L1-L4, walking legs 1-4; O2-O5, opisthosomal segments 2-5.

In summary, the expression pattern observed in *P. opilio* appears to have been subdivided between the two copies of *P. tepidariorum*, with a possible new domain in the opisthosomal organs in *Pt-Otx1*. Taken together with previously published data on this gene family, subfunctionalization appears to be a plausible scenario of the evolutionary trajectory after duplication, with *Pt-Otx1* maintaining the role in anterior segmentation and *Pt-Otx2* retaining the role in eye development.

3.2.2.9. Expression patterns of *Pitx* genes

The *Pitx* genes of vertebrates are functionally diverse, with roles in nervous system, cardiovascular, muscle and eye development[142,143]. Much less in known about the single ortholog found in *D. melanogaster* (*Ptx1*), which is thought to be involved in midgut and nervous system development[144].

Po-Pitx expression was first detected at stage 10, in two clusters of cells in the pre-cheliceral region (Fig. 3.9 A). At stage 14, the head lobe expression becomes restricted to the mid-anterior edge of the pre-cheliceral region (Fig. 3.9 B). Additional expression was detected at this stage, in small groups of cells along the neuroectoderm and as stripes in opisthosomal segments (Fig. 3.9 B'). This expression pattern is maintained during later embryonic stages (Fig. 3.9 C and C').

The two copies of *Pitx* present in *P. tepidariorum* were previously characterised by Daniel Leite and exhibit similar patterns to those of *Po-Pitx*[72]. *Pt-Pitx1* expression was first detected at stage 11, in cell clusters along the neuroectoderm (Fig. 3.9 D-E'). *Pt-Pitx2* starts being expressed at stage 10.1, in two groups of cells in the head lobes and in segmental clusters of cells along the germ band (Fig. 3.9 F-F"). The head lobe expression becomes restricted to the mid-anterior edge of the pre-cheliceral region at later stages (Fig. 3.9 G and H). The segmental pattern remains unchanged in the opisthosoma, although, in the prosoma it is divided into a domain at the base of the appendages and another as cell clusters along the neuroectoderm (Fig. 3.9 G' and H').

The patterns observed for the *Pitx* orthologs of *P. opilio* and *P. tepidariorum* are similar to those described in *D. melanogaster*, suggesting conservation of ancestral functions[144]. In the spider, these patterns appear to be subdivided between *Pt-Pitx1* and *Pt-Pitx2*, which is suggestive of a scenario of subfunctionalization after WGD.

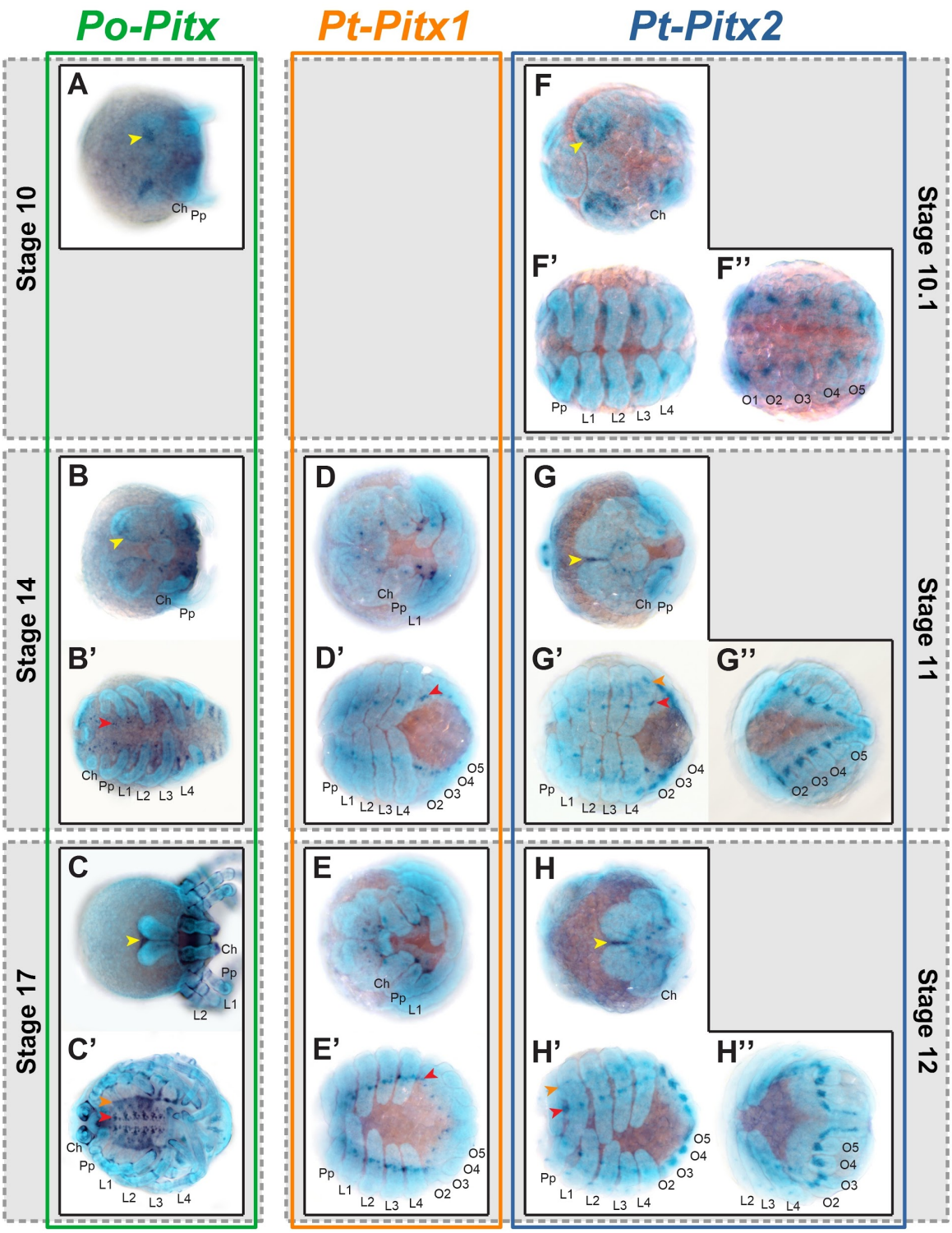

Figure 3.9: Expression patterns of *P. opilio* and *P. tepidariorum Pitx* genes. (A-C') Expression pattern of *Po-Pitx* in stage 10 (A), 14 (B, B') and 17 (C, C') embryos. (D-E') Expression pattern of *Pt-Pitx1* in stage 11 (D) and 12 (E, E') embryos. (F-H'') Expression pattern of *Pt-Pitx2* in stage 10.1 (F-F''), 11 (G-G") and 12 (H, H") embryos. Anterior is to the left in all images. Arrowheads mark expression in the mid-anterior edge of the pre-cheliceral region (yellow), in groups of cells along the neuroectoderm (red) and at the base of the prosomal appendages (orange). Ch, chelicerae; Pp, pedipalps; L1-L4, walking legs 1-4; O1-O5, opisthosomal segments 1-5.

3.2.2.10. Expression patterns of *Lhx2/9* (*Apterous*) genes

Members of the Lhx2/9 homeobox gene family have a broad range of functions in both vertebrates and insects, from dorsoventral patterning to nervous system and appendage development[145,146]. In *D. melanogaster*, the Lhx2/9 gene *apterous* is well known for its role in wing development, but it is also involved in other processes such as muscle and leg development[147–149].

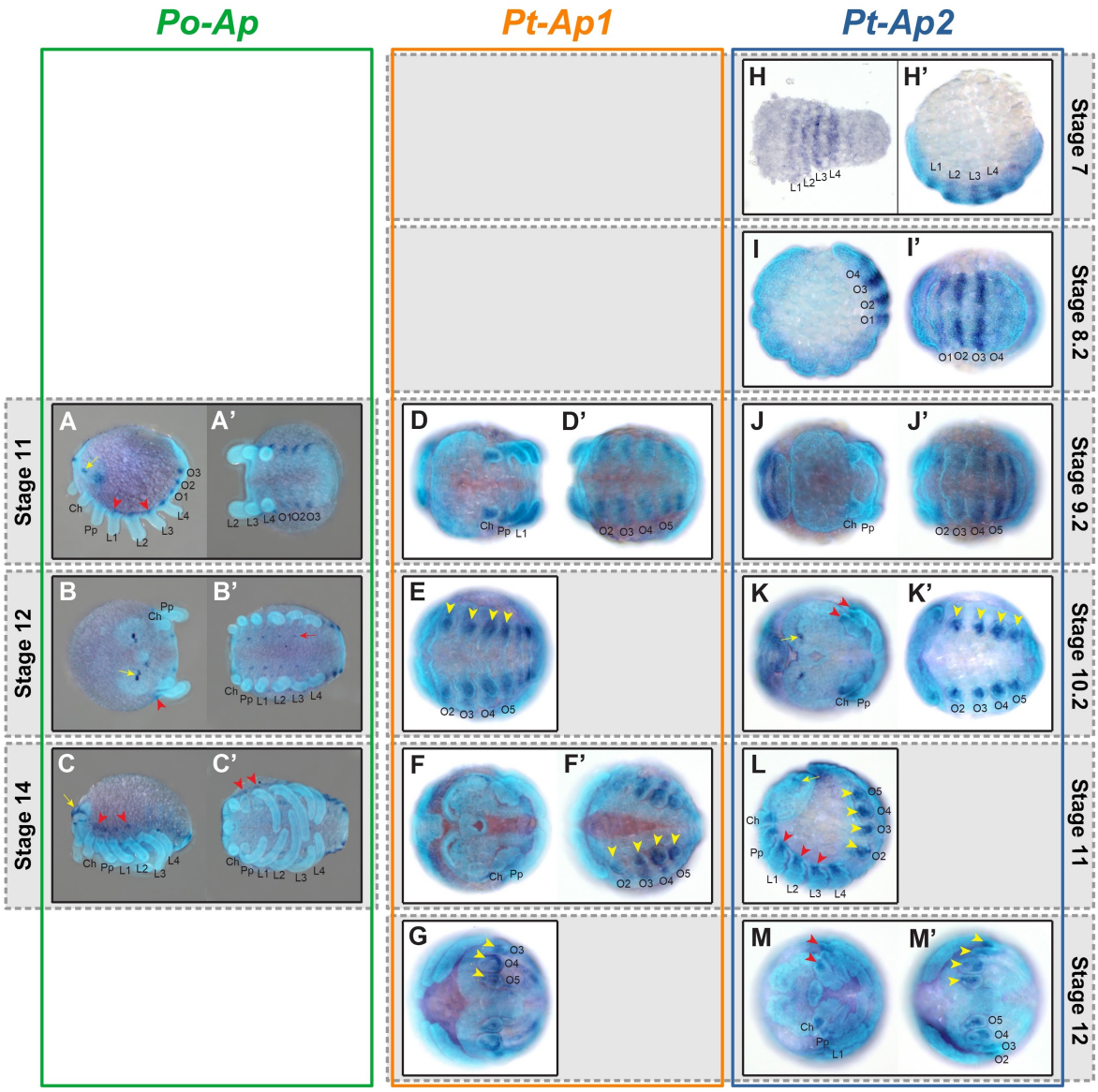

Figure 3.10: Expression patterns of *P. opilio* and *P. tepidariorum Ap* genes. (A-C') Expression pattern of *Po-Ap* in stage 11 (A, A'), 12 (B, B') and 14 (C, C') embryos. (D-G) Expression pattern of *Pt-Ap1* in stage 9.2 (D, D'), 10.2 (E), 11 (F. F') and 12 (G) embryos. (H-M') Expression pattern of *Pt-Ap2* in stage 7 (H, H') (flatmount), 8.2 (I, I'), 9.2 (J-J'), 10.2 (K. K'), 11 (L) and 12 (M-M') embryos. Anterior is to the left in all images. Arrowheads mark expression at the base of the prosomal appendages (red) and in the opisthosomal organs (yellow). Arrows mark expression in the precheliceral region (yellow) and in the neuroectoderm (red). Ch, chelicerae; Pp, pedipalps; L1-L4, walking legs 1-4; O1-O5, opisthosomal segments 1-4.

A total of three *Ap* genes are present in *P. tepidariorum* and a single copy was found in *P. opilio*. *Po-Ap* is expressed segmentally in the opisthosoma and at the base of the prosomal appendages (Fig. 3.10 A-C'). Expression in two groups of cells in the developing head can also be observed (Fig. 3.10 A and B). At stage 12, a dotted expression pattern can be observed in the prosomal region of the neuroectoderm (Fig. 3.10 B'). An additional expression domain can be seen in the pre-cheliceral region, just above the anterior furrow at stage 14 (Fig. 3.10 C).

Pt-Ap1 appears to be first expressed at stage 9.2 in the prosomal appendages and as segmental stripes in the opisthosoma (Fig. 3.10 D and D'). The prosomal expression fades away at later stages and, at stage 11, a new domain can be seen at the lateral edges of the pre-cheliceral region (Fig. 3.10 F). Expression in the opisthosoma is mainly confined to the limb buds and in the region dorsal to these appendages (Fig. 3.10 E, F' and G).

Pt-Ap2 expression is first visible at stage 7 as segmental stripes (Fig. 3.10 H and H'). At stage 8.2, the same segmental pattern becomes restricted to the opisthosoma (Fig. 3.10 I and I'). This pattern is confined to the opisthosomal limb buds at stage 10.2, and prosomal expression appears at the base of the appendages (Fig. 3.10 K and K'). An additional domain can be also observed in the pre-cheliceral region, just above the anterior furrows (Fig. 3.10 K). This pattern remains unchanged during subsequent stages (Fig. 3.10 L-M'). In contrast to its paralogs, ISH for *Pt-Ap3* produced no clear signal, which suggests this copy is no longer expressed during embryogenesis.

In conclusion, *Po-Ap* expression appears to have been subdivided between two of the three copies present in *P. tepidariorum*, with the only exception being the expression domain in the neuroectoderm, which appears to have been lost.

3.2.2.11. Expression patterns of *Irx* genes

The Iroquois homeobox gene family is involved in a broad range of developmental processes in both vertebrates and insects, which include growth regulation, dorsoventral patterning, nervous system development and eye development[150,151]. In vertebrates, *Irx* genes play additional roles in chondrogenesis and cardiovascular development[152,153], and in *D. melanogaster*, the Irx family member *mirror* has a role in segmentation and formation of embryonic dorsal appendages[154,155].

My survey of homeobox genes found three copies of *Irx* in *P. opilio*, which is similar to other non-WGD arthropod species[72]. Four copies of *Irx* genes are present in the genome of *P. tepidariorum*, therefore this homeobox family could still be of interest for this study, as the gain and differential loss of genes could have had an impact on their expression patterns.

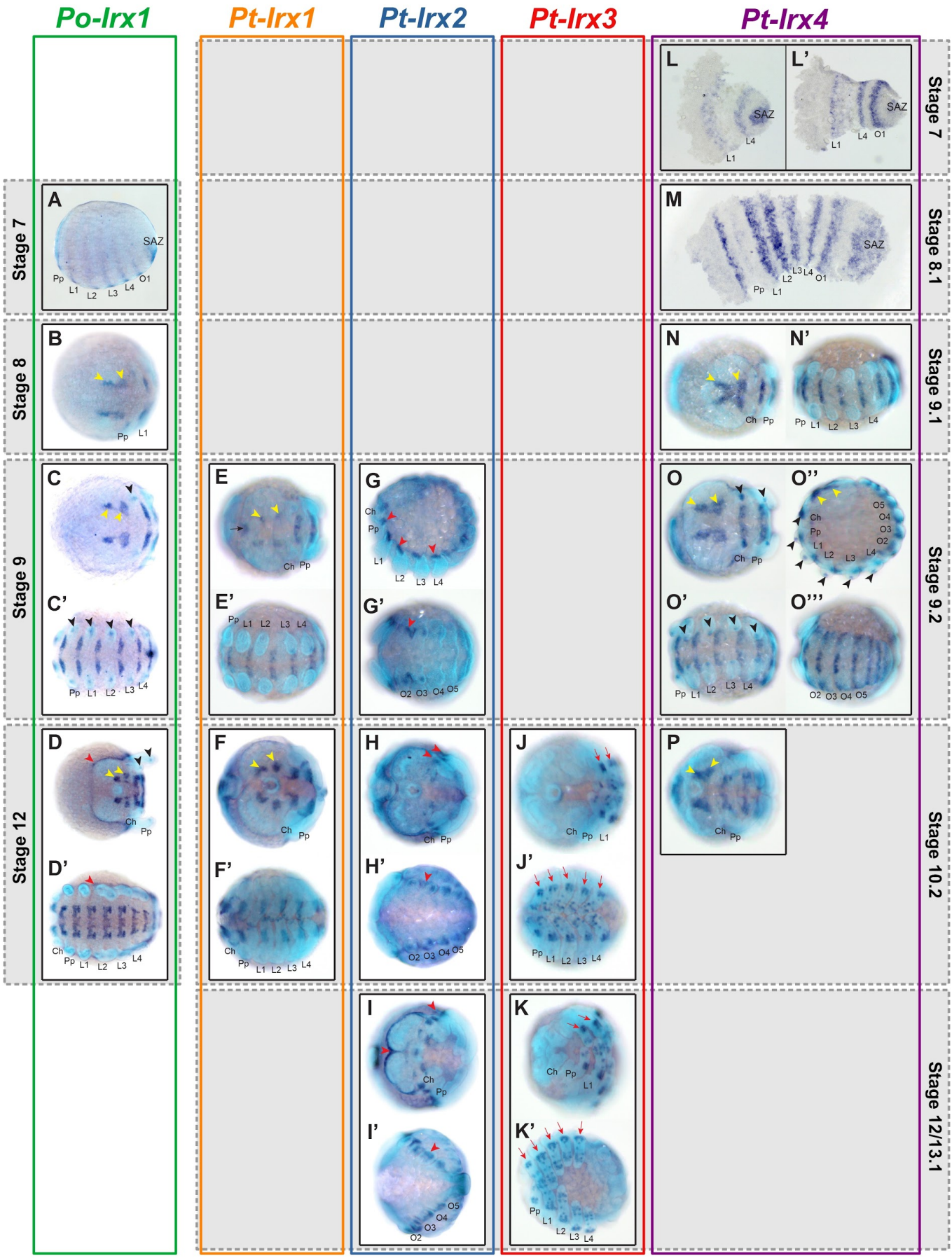
Po-lrx1
Pt-lrx1
Pt-lrx2
Pt-lrx3
Pt-lrx4
Stage 7
Stage 8.1
Stage 8
Stage 9.1
Stage 9
Stage 9.2
Stage 12
Stage 10.2
Stage 12/13.1
A
B
C
C'
D
D'
E
E'
F
F'
G
G'
H
H'
I
I'
J
J'
K
K'
L
L'
M
N
N'
O
O''
O'
O'''
P
SAZ
Pp
Ch
L1
L2
L3
L4
O1
O2
O3
O4
O5

Figure 3.11 (previous page): Expression patterns of *P. opilio* and *P. tepidariorum Irx* genes. (A-D') Expression pattern of *Po-Irx1* in stage 7 (A), 8 (B), 9 (C, C') and 12 (D, D') embryos. (E-F') Expression pattern of *Pt-Irx1* in stage 9.2 (E, E') and 10.2 (F-F') embryos. (G-I') Expression pattern of *Pt-Irx2* in stage 9.2 (G, G'), 10.2 (H-H') and 12 (I-I') embryos. (J-K') Expression pattern of *Pt-Irx3* in stage 11 (J, J'), and 13.1 (K, K') embryos. (L-P) Expression pattern of *Pt-Irx4* in stage 7 (L, L') (flat-mount), 8.1 (M), 8.2 (N-N'), 9.2 (O-O''') and 10.2 (P) embryos. Anterior is to the left in all images. Arrowheads mark expression at the tip of the prosomal appendages (black), in the pre-cheliceral region (yellow) and at the edge of the germ band (red). Arrows mark *Pt-Irx1* expression at the anterior edge of the pre-cheliceral region (black) and *Pt-Irx3* expression in the presumptive peripheral nervous system of pedipalps and walking legs (red). SAZ, segment addition zone; Ch, chelicerae; Pp, pedipalps; L1-L4, walking legs 1-4; O1-O5, opisthosomal segments 1-5.

Po-Irx1 expression was first detected at stage 7 as segmental stripes (Fig. 3.11 A). At stage 8, these stripes become restricted to two clusters of cells in the developing neuroectoderm, and faint expression can be seen in the middle of the developing prosomal appendages (Fig. 3.11 B). At stage 9, the anterior most clusters start dividing into two more clusters, and overall expression becomes stronger (Fig. 3.11 C and C'). By stage 12, expression in the neuroectoderm acquires a more complex pattern and two clusters of expression are present in the pre-cheliceral region (Fig. 3.11 D and D'). A new domain of expression at the edge of the germ band is now visible, and expression in the prosomal appendages is now confined to the tip (Fig. 3.11 D and D'). Unfortunately, PCR amplification was unsuccessful for both *Po-Irx2* and *Po-Irx3*.

In *P. tepidariorum*, the early segmental expression seen for *Po-Irx1* was only observed for *Pt-Irx4* (Fig. 3.11 L-M). The neuroectodermal pattern and the expression domain at the tip of the prosomal appendages appear to be partitioned between *Pt-Irx1* and *Pt-Irx4*, with the latter being expressed earlier and the former at later stages (Fig. 3.11 E-F' and N-O'''). *Pt-Irx2* is mainly expressed at the edge of the germ band, with a few additional domains in the pre-cheliceral region (Fig. 3.11 G-K'). *Pt-Irx3* expression is only visible from stage 11 onwards, and is restricted to a subset of cells in the pedipalps and walking legs (Fig. 3.11 J-K'). Lastly, an additional expression domain can be seen at the anterior edge of the pre-cheliceral region for *Pt-Irx1*, from stage 9.1 to stage 10.1 (Fig. 3.11 E).

In summary, the expression observed for *Po-Irx1* appears to be subdivided between three of the four copies present in *P. tepidariorum*. These patterns are consistent with conserved roles in dorsoventral patterning, nervous system development and appendage development. *Pt-Irx3* could possibly have gained a novel expression domain, although, this pattern could also be present in *Po-Irx2* or *Po-Irx3*. Lastly, the early segmental expression pattern observed for *Po-Irx1* and *Pt-Irx4* suggest a conserved role of *Irx* genes in arthropod segmentation.

3.2.2.12. Expression patterns of *Cux* (*cut*) genes

The *Cux* genes of insects and vertebrates have conserved roles in nervous system development and limb development[156,157]. Other functions include Malpighian tubule, tracheal system and wing development in *D. melanogaster*, and cell proliferation and hair follicle development in vertebrates[157].

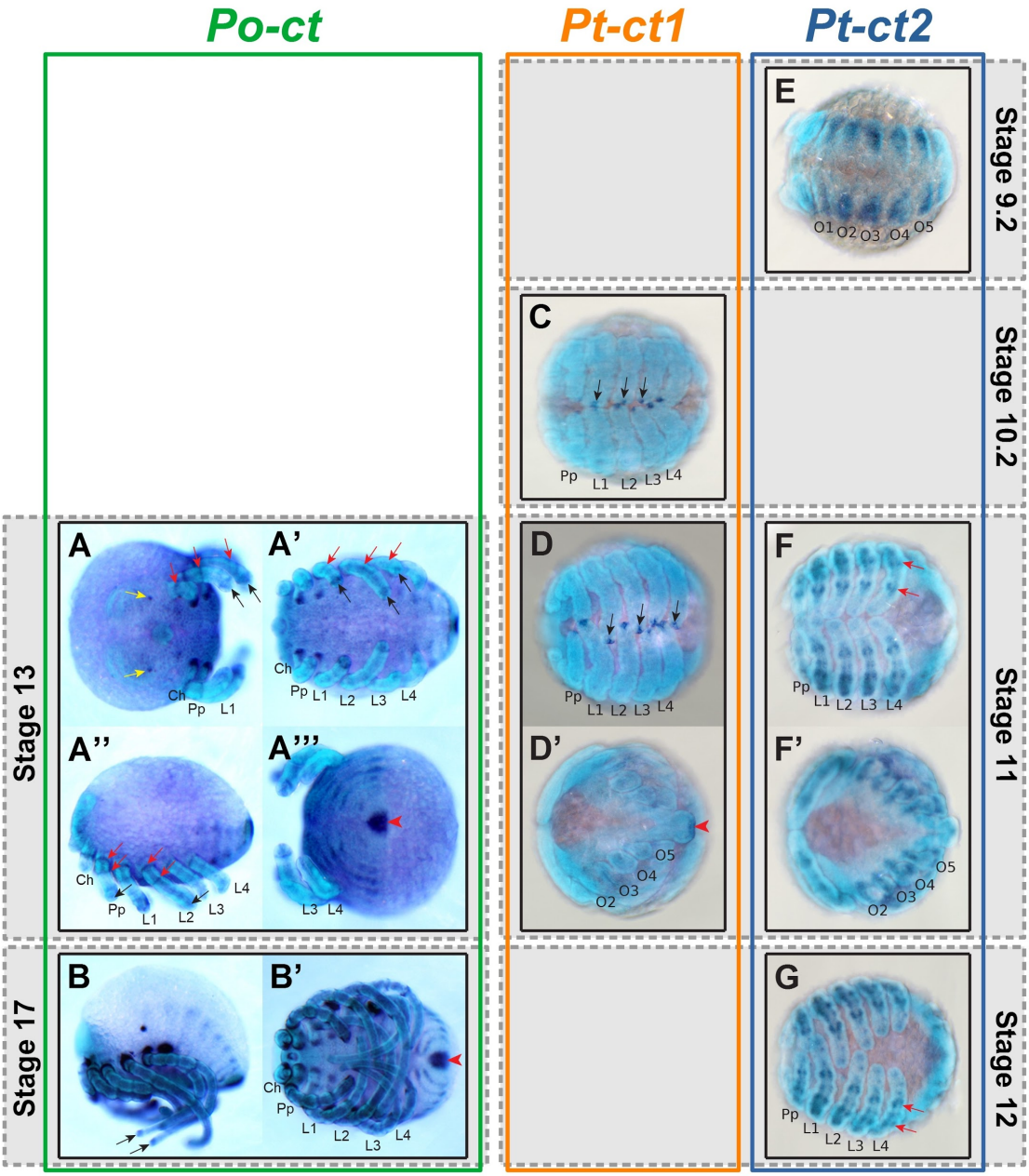

Figure 3.12: Expression patterns of *P. opilio* and *P. tepidariorum cut* genes. (A-B') Expression pattern of *Po-ct* in stage 13 (A-A''') and 17 (B, B') embryos. (C-D') Expression pattern of *Pt-ct1* in stage 10.2 (C) and 11 (D, D') embryos. (E-G) Expression pattern of *Pt-ct2* in stage 9.2 (E), 11 (F, F') and 12 (G) embryos. Anterior is to the left in all images. Arrowheads mark expression at the posterior end of the germ band. Arrows mark expression at the tips (black) and in a ring-like pattern (red) in the prosomal appendages, and in the pre-cheliceral region (yellow). Ch, chelicerae; Pp, pedipalps; L1-L4, walking legs 1-4; O1-O5, opisthosomal segments 1-5.

A single-copy for the Cux family was found in *P. opilio*. *Po-ct* expression was first detected at stage 13 in the prosomal appendages and opisthosomal segments (Fig. 3.12 A-B'). The pattern in the prosomal appendages consists of one (chelicerae), two (pedipalps) or three rings (walking legs), as well as expression at the tip of pedipalps and walking legs (Fig. 3.12 A-A" and B). Strong expression can also be seen in a region ventral to the base of the pedipalps and first pair of legs (Fig. 3.12 A'). *Po-ct* expression in the opisthosoma is mainly segmental, with a stronger domain at the posterior end of the germ band (Fig. 3.12 A'" and B'). Two dots of expression can also be seen in the pre-cheliceral region (Fig. 3.12 A).

The *P. tepidariorum* duplicates, *Pt-ct1* and *Pt-ct2*, were previously characterised by Daniel Leite[72], and appear to have subdivided their expression domains with respect to *Po-ct*. *Pt-ct1* is expressed in the tips of pedipalps and walking legs, as well as at the posterior end of the germ band (Fig. 3.12 C-D'), whilst *Pt-ct2* has retained expression in the posterior segments and prosomal appendages (Fig. 3.12 E-G).

Expression analysis of *Po-ct* suggests that the putative ancestral patterns were subdivided between the two copies present in *P. tepidariorum* after WGD. Moreover, it appears that members of the Cux homeobox gene family may have a conserved role in appendage development.

3.2.2.13. Expression patterns of *Zfh* genes

Zfh genes perform conserved roles in muscle and nervous system development in both insects and vertebrates[158–161]. In *D. melanogaster*, the Zfhx family member *Zfh-1* is necessary for the proper migration of the germ cell line and gonadal development[162].

The two copies of *Zfh* present in *P. tepidariorum* genome were previously analysed by Daniel Leite[72]. However, characterization of the single-copy found in *P. opilio* helped distinguish between potential novel and ancestral expression domains. At stage 8, *Po-Zfh* expression was observed in the developing neuroectoderm and in medial regions of the early prosomal appendages (Fig. 3.13 A). Neuroectodermal expression appears to be maintained throughout embryogenesis (Fig. 3.13 B, C and D). The expression in the prosomal appendages is quite complex. By stage 10, expression in the prosomal appendages is divided into two domains, one at the tip and another at the base (Fig. 3.13 B'). At stage 12, a single ring of expression can be seen in the medial region of the walking legs and pedipalps (Fig. 3.13 C'). Then by stage 14, several additional rings have become visible (Fig. 3.13 D').

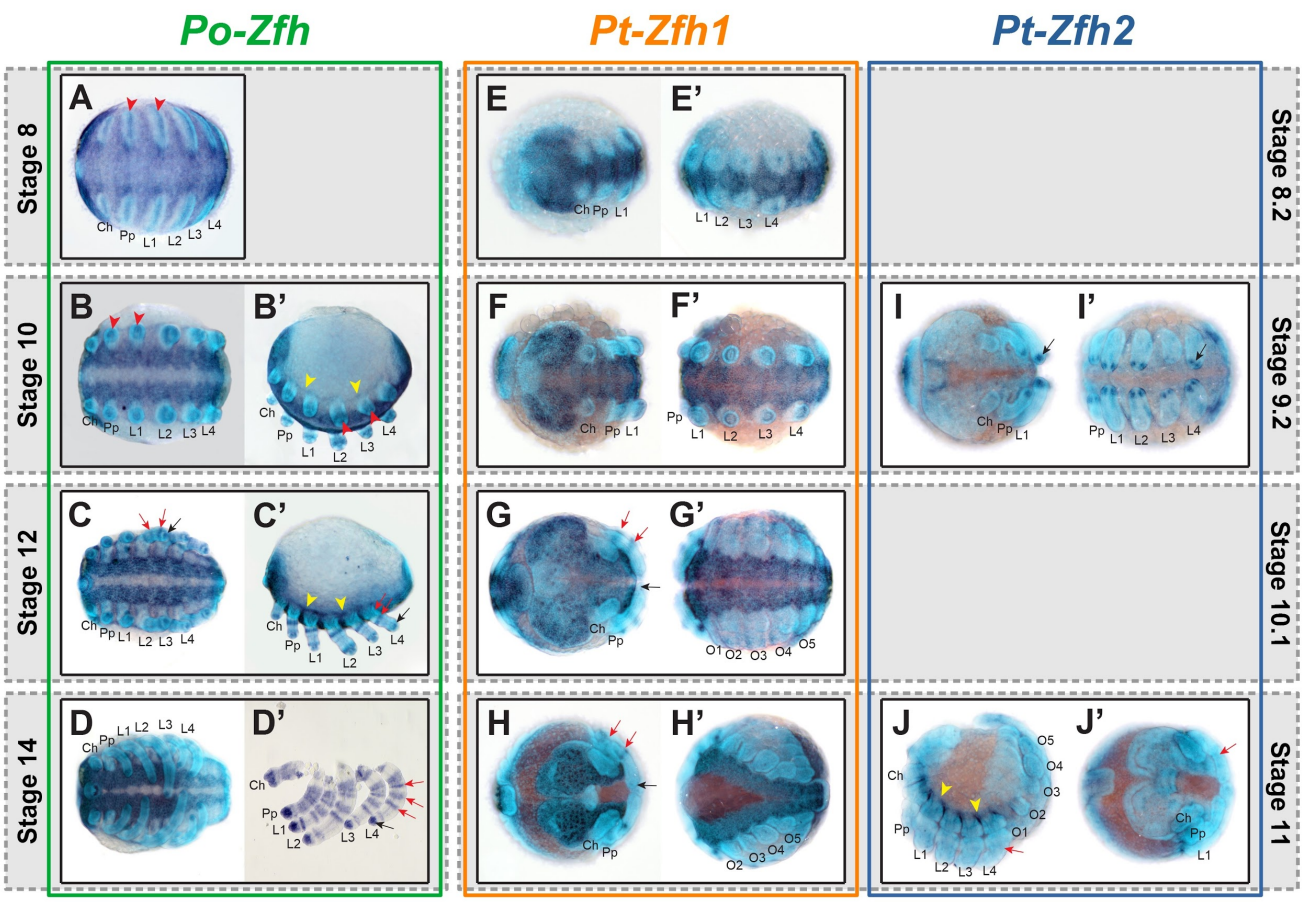

Figure 3.13: Expression patterns of *P. opilio* and *P. tepidariorum Zfh* genes. (A-D') Expression pattern of *Po-Zfh* in stage 8 (A), 10 (B, B'), 12 (C, C') and 14 (D, D') (leg mount) embryos. (E-H') Expression pattern of *Pt-Zfh1* in stage 8.2 (E-E'), 9.2 (F, F'), 10.1 (G, G') and 11 (H, H') embryos. (I-J') Expression pattern of *Pt-Zfh2* in stage 9.2 (I, I') and 11 (J-J') embryos. Anterior is to the left in all images. Arrowheads mark expression in growing prosomal limb buds (red) and at the base of prosomal appendages (yellow). Arrows mark expression at the tips (black) and in a ring-like pattern (red) in the prosomal appendages. Ch, chelicerae; Pp, pedipalps; L1-L4, walking legs 1-4; O1-O5, opisthosomal segments 1-5.

In conclusion, expression patterns previously observed for *Pt-Zfh1* and *Pt-Zfh2* (Fig. 3.13 E-J') appear to represent a subdivision of the *Po-Zfh* expression domains. Neuroectoderm expression appears to be mainly conserved in *Pt-Zfh1* (Fig. 3.13 E-H') and different aspects of the prosomal appendage expression patterns appear to have been partitioned between each duplicate (Fig. 3.13 E-J'). Moreover, these expression patterns are consistent with functional conservation of these genes in nervous system and limb development across Arthropoda, as similar domains of expression have been described for *Zfh* orthologs in *D. melanogaster*[161].

3.3. Discussion

3.3.1. The *P. opilio* homeobox repertoire is consistent with a non-WGD ancestral state

The study of chordate homeobox gene repertoires provided a key piece of evidence for the existence and timing of the two WGD events in the common ancestor of vertebrates[39,75,163–165]. Identification of the homeobox gene repertoires of two putative non-WGD species, the urochordate *Ciona intestinalis* and the cephalochordate *Branchiostoma floridae*, was particularly instrumental to this analysis, providing evidence for the biased increase of homeobox gene number in the vertebrate lineage following the two rounds of WGD[164–166].

Recently, the analysis of arachnid homeobox genes was once again crucial to the identification of the arachnopulmonata WGD[30]. Homeobox gene repertoires have been identified for a total of five chelicerate species, although, only one of these, the tick *I. scapularis*, is a non-WGD representative[72]. Moreover, it has been suggested that the tick lineage (Parasitiformes) exhibit a marked loss of gene content, much like its putative sister group Acariformes[72,167,168]. Therefore, the identification of homeobox genes of other putative non-WGD arachnid orders, such as Opiliones, would provide further evidence for the timing of the WGD event and allow a better inference of the pre-WGD state of arachnid homeobox gene repertoires.

This survey represents the first attempt to identify the complete homeobox gene repertoire in the harvestmen *P. opilio*. The vast majority of the identified homeobox gene families appear to be unduplicated, a result consistent with the scenario of a non-WGD ancestry for this arachnid. However, the set of genes identified is probably incomplete because the total homeobox gene number found is inconsistent with those previously identified for other non-WGD arthropods[72]. Furthermore, I was unable to identify orthologs of several well-conserved families, for which gene loss is improbable[72]. The very low number of duplicated gene families found also supports this conclusion because it is inconsistent with the frequency of duplicated gene families in other non-WGD arthropods[72].

Therefore, higher quality transcriptome data for *P. opilio*, as well as other species from closely related non-WGD arachnid orders (e.g. Solifugae) (Fig. 1.3), is essential to better understand the ancestral repertoire of homeobox genes in Arachnida. Such data would provide a better basis to understand lineage specific gains and losses, as well as patterns of gene retention and functional divergence after the arachnopulmonate WGD. Data from additional arachnopulmonate species (e.g. Amblypygi) would also prove beneficial to better infer the evolutionary fate of arachnopulmonate ohnologs.

3.3.2. Expression divergence of duplicated arachnid homeobox genes

To investigate the impact of WGD on the evolution of homeobox gene function, I compared the expression patterns of duplicated homeobox genes in *P. tepidariorum* to those of their corresponding single-copy orthologs in *P. opilio*. As previously stated, modifications to homeobox gene expression can result in changes to their functional outcome and phenotypic consequences[39,84]. Therefore, if the expression patterns of duplicated homeobox genes in *P. tepidariorum* represent subsets of the single-copy expression pattern observed in *P. opilio*, the most plausible scenario of functional divergence would be that of subfunctionalization. Likewise, if one or more copies present in *P. tepidariorum* acquire an expression domain that is not detected in the single-copy ortholog of *P. opilio*, it is possible that it represents a scenario of neofunctionalization.

All the duplicated spider homeobox gene families analysed showed at least some degree of expression divergence. The most common outcome appears to be the partitioning of the single-copy expression domains of *P. opilio* homeobox genes between the duplicates of *P tepidariorum* (i.e. subfunctionalization). A total of 8 out of the 13 families of homeobox gene analysed appear to have been subfunctionalised. The Emx, Lhx2/9 and Irx homeobox gene families are clear examples of this, where the observed pattern in *P. opilio* has been subdivided between the duplicates in *P. tepidariorum*. Studies of duplicated homeobox genes in vertebrates have often found the same scenario, with ohnologs being expressed in similar tissues and with often overlapping domains, suggesting similar roles that probably represent subfunctionalization of ancestral functions[88,108,126,143,156,169,170]. For example, paralogs of the Emx, Pitx and Cux homeobox gene families of vertebrates are expressed in similar domains in the developing nervous system[130,143,156].

A recent genome wide study on the fate of duplicated genes in the vertebrate lineage have suggested that neofunctionalization, rather than subfunctionalization, appears to be the most common fate[42]. However, these results are in fact demonstrating a higher rate of a duplicate gene fate termed specialization (i.e. maintenance of a specific ancestral domain of expression) rather than neofunctionalization (i.e. gain of a new domain of expression). This is similar to what I have found in some of the homeobox ohnologs here analysed in *P. tepidariorum*, such as the Emx, Irx, Pitx and Cux families, where at least one paralog appears to have maintained a specific aspect of the ancestral expression pattern. Therefore, my results are in line with the findings of this study.

Possible cases of neofunctionalization (i.e. gain of a novel expression domain) were less common. One such case was found in the Msx family, in which *Pt-Msx2* may have acquired expression in a basal region of the chelicerae, a domain that is absent in *Po-Msx*. The prosomal appendage expression of *Pt-Irx3* is another possible example, although, this expression

may also be present in one of the other two copies of *P. opilio*, whose expression was not possible to analyse. Another explanation could be that these expression patterns were lost in *P. opilio*, thus expression analysis of *Irx* orthologs from other non-WGD species would help resolve this question.

A third scenario of expression divergence was observed, where one duplicate retained the pattern of the single-copy ortholog whilst the other did not seem to be expressed during embryogenesis. However, these paralogues might be expressed in post-embryonic stages, and could represent instances of sub or neofunctionalization, depending on whether or not this expression is present in their unduplicated version.

It is also interesting to point out the fact that subfunctionalised expression patterns are usually located in the developing nervous system and prosomal appendages, hinting at possible functional specialization of these genes to subsets of cells in the central and peripheral nervous systems. Interestingly, the biased retention of ohnologs expressed in the nervous system was also reported in the vertebrate lineage, and is thought to have provided the raw material for the evolution of vertebrate novelties in the nervous system[40,41]. Furthermore, the same study that found that the most common fate of vertebrate duplicated genes appears to be specialization, also found that specialized duplicates were often expressed in neural tissues[42].

Additionally, a considerable number of the homeobox duplicates here analysed are also expressed in the opisthosomal limb buds, including the developing book lungs and spinnerets, which are lineage specific traits of the Arachnopulmonata and Araneae respectively. Thus, the duplication of these homeobox gene families could have facilitated the evolution of these novel structures. In conclusion, expression divergence was found for most of the homeobox gene paralogs analysed in this study, suggestive of extensive subfunctionalization and/or neofunctionalization of these genes after WGD that likely had an impact on the evolution and divergence of the arachnopulmonate lineage.

4. Chapter II: Impact of gene duplication in arachnid Sox genes

4.1. Introduction

Sox (Sry-related high-mobility-group box) genes encode a family of transcriptional regulators with many critical roles in development[171,172]. Similarly to homeobox genes, they bind to DNA with partner transcription factors to regulate gene expression[171]. Binding to DNA is mediated by a highly conserved high-mobility-group (HMG) domain, a defining feature of this gene group[171]. In metazoans, Sox genes can be subdivided into 10 groups, from Group A to Group J[173], although, arthropod Sox gene repertoires are restricted to Groups B, C, D, E and F[174,175]. Sox gene number is conserved in most arthropods, with 4 Group B genes (*Dichaete*, *SoxNeuro*, *Sox21a* and *Sox21b*) and a single representative of Groups C-F[176]. Most of our knowledge on Sox gene function in arthropods comes from studies in *D. melanogaster*[176]. Thus, there is a paucity of knowledge regarding Sox gene function and evolution more broadly in the arthropod lineage, with only a few recent studies addressing this issue[65,175].

4.2.1. Sox genes in animal development

Sox genes play critical roles in several key developmental processes such as neurogenesis and chondrogenesis[171,172,177,178]. Thus, any mutations affecting Sox expression can lead to severe developmental abnormalities, as is the case of several human diseases[171,172]. In vertebrates for example, the SoxB genes *Sox1*, *Sox2* and *Sox3* are necessary for the maintenance of neural stem cells, and reduced expression of *Sox2* in mice results in neuronal defects[172]. *Sox9*, a Group E gene, has a critical role in the commitment and differentiation of mesenchymal precursor cells to the chondrocyte cell lineage in humans and mice[172], and *Sox7*, a Group F gene, is necessary for the proper development of the cardiovascular system in mice[171].

Analysis of Sox gene function in arthropods is more scarce and mostly restricted to *D. melanogaster*[176]. *Dichaete* and *SoxNeuro*, two Group B genes, are essential for the correct patterning and development of the CNS, a role evocative of their vertebrate orthologs[176]. In addition to their role in CNS development, these genes are also known for other functions, such as the patterning of larval cuticle and the recently characterised role of *Dichaete* in segmentation[176,179]. The only known function of *Sox21a* is a role in regulation of stem cell proliferation in the adult midgut[180]. Nevertheless, *Sox21a* is expressed during embryogenesis in the anlage of the foregut and hindgut, and in the ventral midline[181]. To date, no functional data is available for *Sox21b* in *D. melanogaster*, although, it is expressed in a segmentally repeated pattern in the ventral epidermis during embryogenesis[181]. Outside Group B, *Sox14* (*SoxC*) is involved

in neurogenesis[182], *Sox102F* (*SoxD*) knockdown leads to several neuronal abnormalities and irregular behaviour[183], *Sox100B* (*SoxE*) is a key factor in testis development[184], and *Sox15* (*SoxF*) regulates cell proliferation during wing development[185].

4.2.2. Sox genes and animal evolution

Despite the functional conservation of Sox genes (e.g. CNS development), this family of transcription factors has also acquired several novel functions during animal evolution[186,187]. In vertebrates, Sox genes have acquired vital roles in the development of tissues and cell types specific to this lineage[171,186,187]. For example, SoxB genes are key regulators of sensory placode development and SoxE genes have a critical role in neural crest cell fate specification, two vertebrate-specific structures[171,186]. Other examples are the roles of *Sox5/6* and *Sox9* in chondrogenesis, an essential component of skeletal development, and the role of SoxF genes in cardiovascular development[171,186].

The involvement of Sox genes in the evolution of arthropod development is less obvious. In *D. melanogaster*, *Sox14*, a Group C gene, is a key regulator of several steroid hormone induced pathways during metamorphosis[188], and *Sox15*, as mentioned above, is necessary for proper wing development[185], both examples of innovations in insects. This is probably only a small window into the impact of Sox genes in arthropod evolution, as studies of Sox gene expression patterns have shown a considerable extent of divergence between different arthropod lineages[65,174,175].

4.2.3. Impact of gene duplication in Sox gene evolution

Sox gene number varies considerably throughout Metazoa, ranging from 3 to 5 copies in sponges and placozoans to over 22 in teleosts[176,189]. Two Sox-like genes have also been identified in a choanoflagellate, hinting at a possible role in the evolution of multicellularity[190]. Sox genes are thought to have evolved from other canonical HMG domain proteins[172,173]. Early expansion of Sox repertoires is still poorly understood, although, lineage-specific duplications appear to be a common trend[173,176,187], for example, arthropod SoxB genes and hymenopteran SoxE genes[174,176]. In addition to this, Group A and G-J genes are restricted to certain lineages and are also thought to be the result of lineage-specific duplication[171,173,176]. For example, the mammal-specific Group A (*Sry*) and G (*Sox15*) genes, are now known to have derived from the Group B genes *Sox3* and *Sox19* respectively[191,192], and subsequently acquired new roles in male sex determination (*Sry*)[171] and in skeletal muscle regeneration (*Sox15*)[193].

Sox gene number and evolution has also been highly influenced by WGD events[187,189]. Like homeobox genes, members of the Sox gene family appear to have been preferentially retained after the WGD events in vertebrates[173,189]. This WGD-mediated expansion and sub-

sequent divergence of vertebrate Sox repertoires facilitated the subdivision of ancestral functions, leading to partial functional redundancy, as well as the acquisition of novel roles in the development of several vertebrate innovations (see above)[186,187]. Furthermore, additional WGD events in teleosts led to further expansion of Sox genes, resulting in lineage-specific divergence of Sox repertoires, which may have contributed to the genetic and phenotypic diversity seen in this lineage[189].

4.2.4. Sox genes in arachnids

The recent discovery of a WGD event at the base of Arachnopulmonates offers a unique opportunity to study the impact of WGD in the evolution of Sox genes independently to the vertebrate WGD events[30]. Within arachnids, identification of Sox gene repertoires has only been carried out for two species of spiders, *P. tepidariorum* and *S. mimosarum*[65]. Similar to the outcome of vertebrate WGDs, substantial retention of duplicated Sox genes was found in both species, with Sox Groups C-F mostly represented by at least two copies each (Fig. 4.1)[65]. In Group B, single copies of *SoxN* and *Dichaete* were found in both species, while *Sox21a* was found to be duplicated in both species and three copies of *Sox21b* were identified in *P. tepidariorum* (Fig. 4.1)[65]. This represents a clear scenario of gene retention upon WGD, as Sox genes are predominantly found as single copies in other non-WGD arthropods[65,175]. Moreover, synteny analysis supports the origin of these duplicates at the arachnopulmonata WGD event, with the only exception being *Pt-SoxB-like*, which is thought to be a tandem duplicate of *Pt-Sox21b-2*[65].

To date our knowledge about arachnid Sox gene expression is limited to a single study in the spider *Parasteatoda tepidariorum*[65]. Of the 15 Sox genes identified in *P. tepidariorum*, the authors only found 6 of them to be expressed during embryogenesis[65]. The expression patterns observed were largely similar to those previously described in other arthropod species, suggesting extensive functional conservation across arthropods[65,175]. A clear example of this is *SoxN*, which is consistently expressed early in the developing nervous system[65,174–176]. No clear instances of sub or neofunctionalization were found, with the expression of only one paralog being detected for each duplicate pair[65]. Additionally, the function of one arachnid Sox gene has been studied during embryogenesis[66]. This study revealed that one of the duplicated SoxB genes, *Pt-Sox21b-1*, is involved in spider segmentation (see Chapter III for a more detailed description)[66].

In this chapter, I provide a broader overview of Sox gene repertoire among arachnids, as well as a more thorough analysis of Sox expression pattern divergence in arachnids with and without an ancestral WGD.

4.2. Results

4.2.1. Identification and comparison of Sox gene repertoires in arachnids and other arthropods

To elucidate how Sox gene repertoires have changed throughout arthropod evolution, especially in the arachnid lineage, I performed a BLAST search against the available genomic resources to identify HMG containing sequences of additional arachnids, namely the scorpion *C. sculpturatus*, the harvestmen *P. opilio* and the tick *I. scapularis*, and a myriapod, the centipede *S. maritima*, and compared them to other previously identified arthropod Sox repertoires[65,175]. Sequences were annotated as the best hits obtained from the SMART BLAST NCBI online tool and by reciprocal BLAST against the proteome of *P. tepidariorum*.

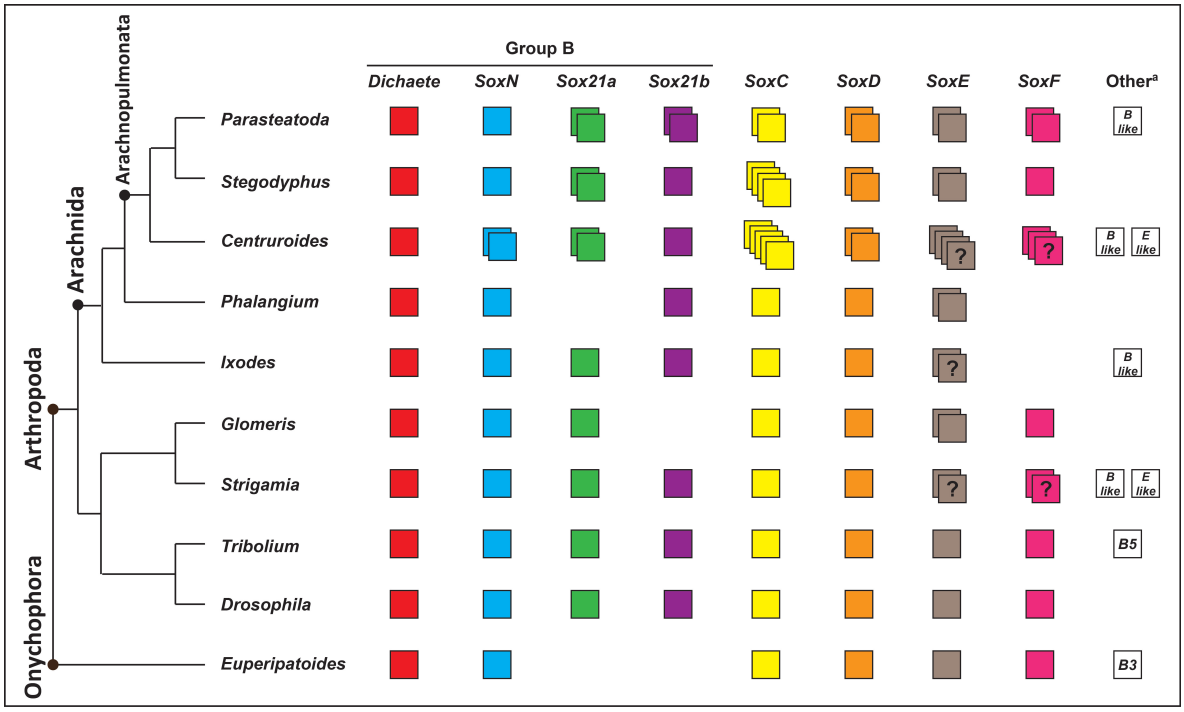

Figure 4.1: Panarthropoda Sox gene repertoires. Diagram showing the number of Sox genes in the surveyed species (*Centruroides sculpturatus*, *Phalangium opilio*, *Ixodes scapularis* and *Strigamia maritima*), as well as other arthropods (*Parateatoda tepidariorum*, *Stegodyphus mimosarum*, *Glomeris marginata*, *Tribolium castaneum* and *Drosophila melanogaster*) and an onychophoran (*Euperipatoides kanangrensis*), for which Sox repertoires were previously identified. Each colour represents a different gene. ? = unclear number of copies due to incomplete sequences. Other[a] = unresolved or highly divergent sequences.

I identified a total of 22 Sox genes in *C. sculpturatus*, 7 in *P. opilio*, 9 in *I. scapularis* and 12 in *S. maritima* (Fig. 4.1). At least one representative of each of Sox group was found for all species surveyed, with the exceptions of Group F genes in *P. opilio* and *I. scapularis*, and *Sox21a* in *P. opilio* (Fig. 4.1). Additionally, a considerable number of sequences retrieved had incomplete HMG domains, probably representing only partial fragments of the complete gene

sequence (Fig. S1 and S2). Therefore, the Sox repertoires here described may be incomplete due to deficient or poor sequence data, because Sox gene loss is very rare especially in the arthropods[176].

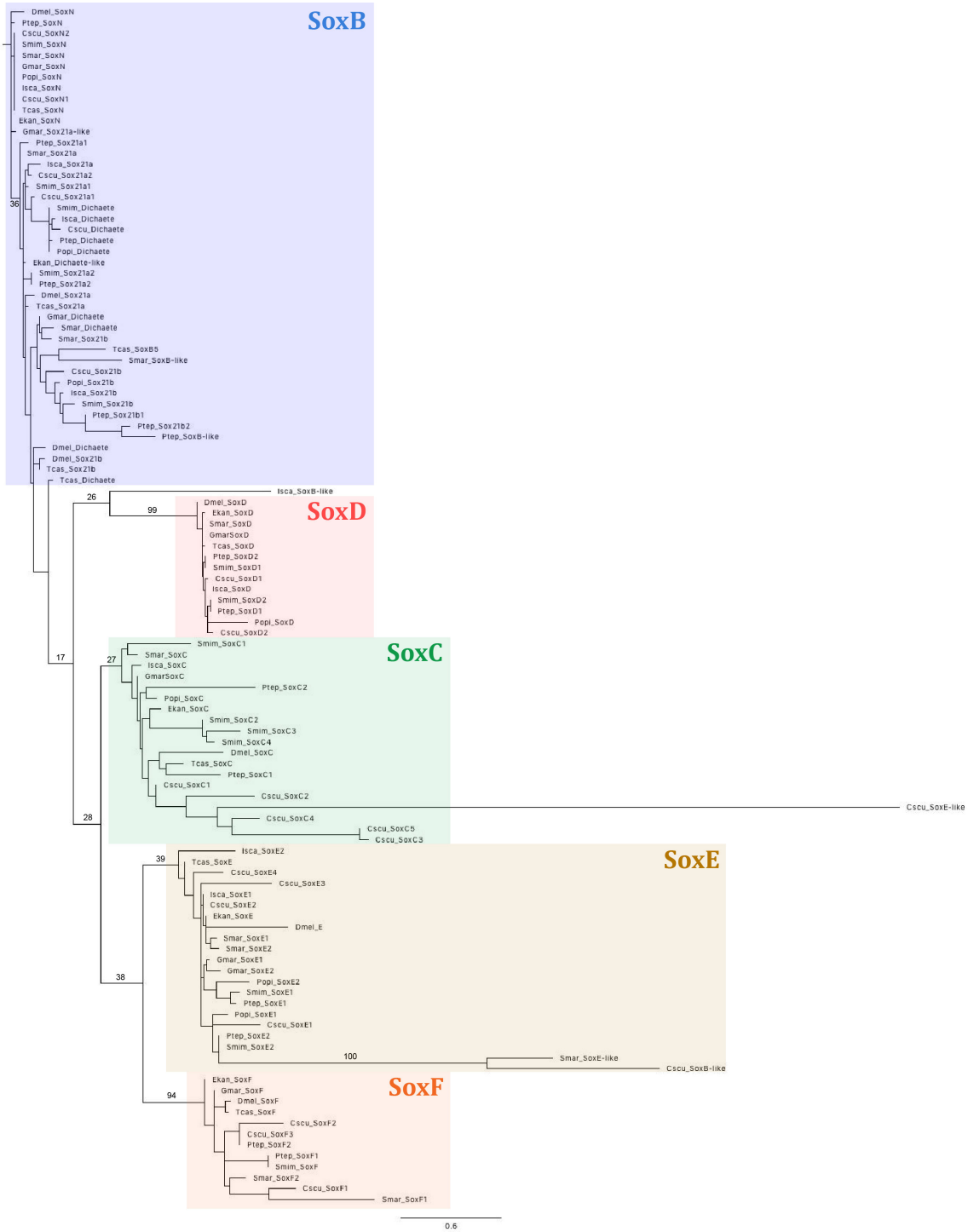

Figure 4.2: Panarthropod Sox HMG domain phylogeny. Maximum likelihood tree of HMG domain sequences from *E. kanangrensis*, *G. marginata*, *S. maritima*, *T. castaneum*, *D. melanogaster*, *I. scapularis*, *P. opilio*, *C. sculpturatus*, *S. mimosarum* and *P. tepidariorum*. Bootstrap values are shown for main branches. RaxML run using PROTGAMMALG model (1000 bootstraps).

Consistent with previous surveys, a single copy of *Dichaete* was found in all species analysed (Fig. 4.1). A single copy of *SoxN* was also found in most species, with the exception of *Centruroides*, and *Sox21b* orthologs follow a similar trend, with only one instance of duplication found in *Parasteatoda* (Fig. 4.1). A single copy of *Sox21a* is present in non-WGD arthropods, whereas two copies were found in all 3 arachnopulmonate species analysed (Fig. 4.1). Additional *SoxB-like* sequences were also found in *Centruroides*, *Ixodes* and *Strigamia*, although, much like other previously identified *SoxB-like* genes (i.e. *Tc-SoxB5* and *Ek-SoxB3*), these sequences diverge significantly from other Group B genes and therefore cannot be reliably classified (Fig. 4.1 and S1).

All non-WGD arthropods have single copies of *SoxC* and *SoxD* (Fig. 4.1). In arachnopulmonates, at least two copies for each of these groups were found, with additional copies of *SoxC* in *Centruroides* (5 in total) and *Stegodyphus* (4 in total) (Fig. 4.1). Single copies of *SoxE* are present in *Euperipatoides*, *Drosophila* and *Tribolium*, and two copies can be found in *Glomeris*, *Phalangium*, *Stegodyphus* and *Parasteatoda* (Fig. 4.1). The survey retrieved two *SoxE* sequences for *Strigamia* and *Ixodes*, although, *Sm-SoxE2* has an incomplete domain, and each copy in *Ixodes* contains non-overlapping fragments of the HMG domain (indicating they might be the same gene) (Fig. 4.1 and S2). I identified a total of 4 *SoxE* sequences in *Centruroides*, although, these also include incomplete HMG domains (Fig. 4.1 and S2). Additional *SoxE-like* sequences were found in *Strigamia* and *Centruroides*, although their sequences are highly divergent from other *SoxE* genes (Fig. 4.1 and S2). Lastly, a single copy of *SoxF* can be found in *Euperipatoides*, *Drosophila*, *Tribolium*, *Glomeris* and *Stegodyphus*, while two copies are present in *Strigamia* and *Parasteatoda*, and a total of 3 sequences were recovered for *Centruroides* (Fig. 4.1). The putative *SoxF* sequences of *Strigamia* and *Centruroides* have incomplete HMG domains, thus, these might represent fragments of a single copy in *Strigamia* and two copies in *Centruroides* (Fig. S2).

To verify the correct assignment of sequences retrieved to each Sox group and possible evolutionary relationships between and within these groups, I constructed a maximum likelihood tree using the HMG domain protein sequences of the species analysed above (Fig. 4.2; see Fig. S1 and S2 for sequence alignment). Overall, the tree supported the classification of most sequences obtained, forming monophyletic clades correspondent to each Sox group (Fig. 4.2), except for a few highly divergent sequences (i.e. *Is-SoxB-like*, *Cs-SoxB-like* and *Cs-SoxE-like*) (Fig. S1 and S2). However, the bootstrap values only supported the SoxD (99) and SoxF (94) Groups, possibly due to the presence of highly divergent Sox-like sequences in the other groups (Fig. 4.2). Low bootstrap values appear to be a common result in phylogenies constructed with HMG sequences[175].

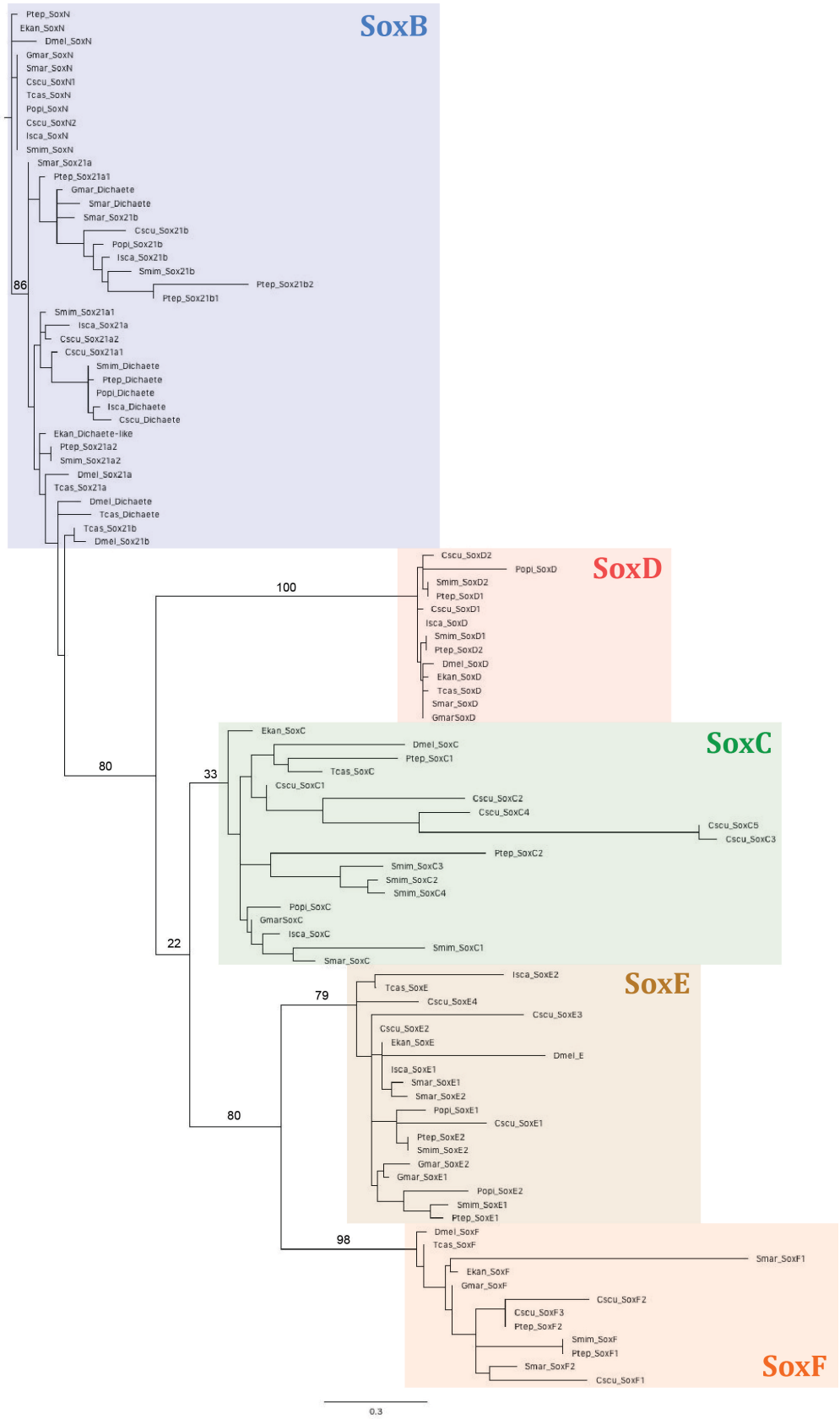

Figure 4.3: Panarthropod Sox HMG domain phylogeny excluding highly divergent sequences. Maximum likelihood tree of HMG domain sequences excluding all SoxB-like and SoxE-like sequences, and *Gmar-Sox21a-like*. Bootstrap values are shown for main branches. RaxML run using PROTGAMMALG model (1000 bootstraps).

To verify if the removal of highly divergent sequences from the analysis would improve the overall support of each Sox Group, I constructed another maximum likelihood tree using the same parameters after removal of highly divergent HMG sequences (Fig. 4.3; see Fig. S3 and S4 for sequence alignment). The sequences removed include all *Sox-like* sequences and *Gmar-Sox21a* (incomplete sequence). The resulting tree shows much higher support values for Sox Group B (80) and Sox Group E (79), with Sox Groups D and F still being very well supported (100 and 98 respectively). The only exception is the branch leading to Sox Group C genes (33), again, possibly due to considerable sequence divergence within this group (Fig. 4.3 and S4). Monophyly of Sox Groups E and F is also well supported by this tree (80) (Fig. 4.3), and within Sox Group B, there is strong support for the monophyly of *SoxN* sequences (86) (Fig. 4.3).

4.2.2. Characterisation of the expression patterns of duplicated *P. tepidariorum* Sox genes and *P. opilio* single-copy orthologs

To investigate the consequences of WGD for the expression of Sox genes, I characterised the expression patterns of duplicated genes in *P. tepidariorum* and their corresponding single-copy orthologs in *P. opilio* by means of ISH. Although *P. tepidariorum* Sox gene expression patterns were previously characterised, expression was only detected for 6 out of 15 Sox genes present in *Parasteatoda*, namely *Pt-SoxN*, *Pt-Sox21b-1*, *Pt-SoxC1*, *Pt-SoxD1*, *Pt-SoxE1* and *Pt-SoxF2*[65]. This is an unexpected result as it contrasts with what is seen in the vertebrate lineage, where most retained duplicates were found to be expressed and have roles during early development[171]. Therefore, I carried out a more rigorous analysis of Sox gene expression patterns in a wider range of embryonic stages of *P. tepidariorum*, by using longer ISH probes that could increase the specificity and signal strength of the staining. I also characterised the expression patterns of identified *P. opilio* single-copy orthologs to distinguish conserved expression from possible novel domains. A result summary is provided in Table S6.

4.2.2.1. Expression patterns of *Dichaete* genes

Both *P. opilio* and *P. tepidariorum* have a single copy of *Dichaete* in their genomes. Unfortunately, due to technical and biological material restrictions, I was unable to analyse *Po-D* expression. *Pt-D* appears to be ubiquitously expressed throughout embryogenesis (Fig. 4.4). However, expression appears to be stronger in specific regions at different stages. At stage 7, *Pt-D* expression appears to be stronger in two anterior stripes, a pattern that is seemingly maintained up to stage 8.2 (Fig. 4.4 C). At stage 9.2, expression appears to be stronger in the prosomal appendages (Fig. 4.4 D). This pattern is maintained at stage 10.1, with an additional expression domain at the edge of the pre-cheliceral region (Fig. 4.4 E).

Though the previous study did not detect any expression for this gene[65], it could be the authors considered the staining represented a background signal. I excluded this conclusion in view of the fact that the embryos stained very quickly, contrary to what is expected of a regular background signal.

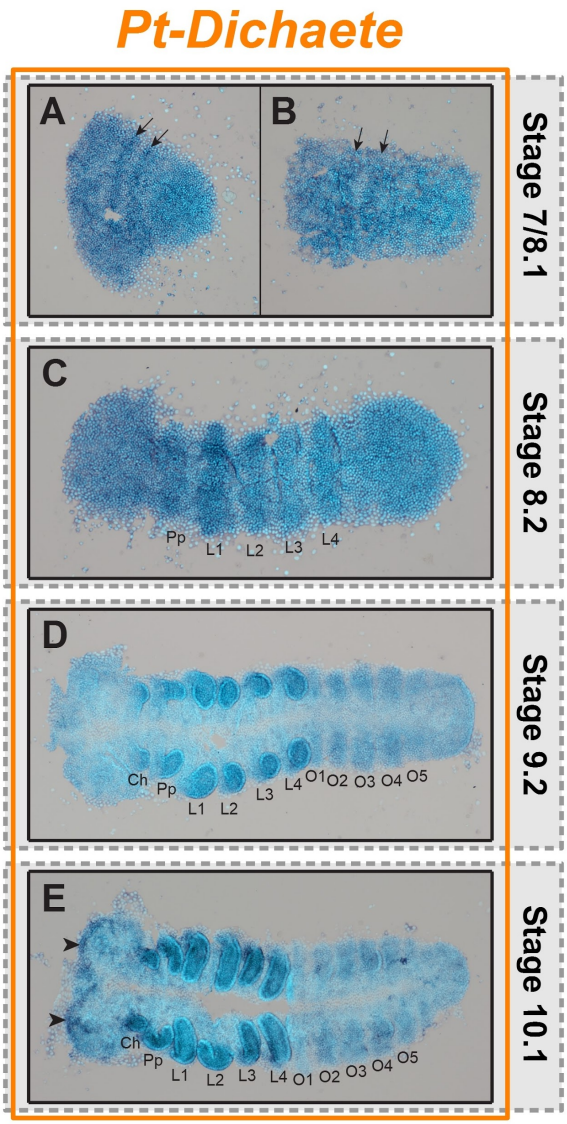

Figure 4.4: Expression patterns of *P. tepidariorum Dichaete*. (A-E) Expression pattern of *Pt-D* in stage 7 (A), 8.1 (B), 8.2 (C), 9.2 (D) and 10.1 (E) flat mounted embryos. Anterior is to the left in all images. Arrowheads mark expression at the edge of the precheliceral region. Arrows mark anterior stripes of expression during early stages of embryogenesis. Ch, chelicerae; Pp, pedipalps; L1-L4, walking legs 1-4; O1-O5, opisthosomal segments 1-5.

4.2.2.2. Expression patterns of *Sox21a* genes

Two copies of *Sox21a* were previously identified in *P. tepidariorum*, although, I was unable to identify any copies of this gene in *P. opilio*. Although the previous study did not detect any expression of these genes during *P. tepidariorum* embryogenesis[65], I was able to detect expression for both copies.

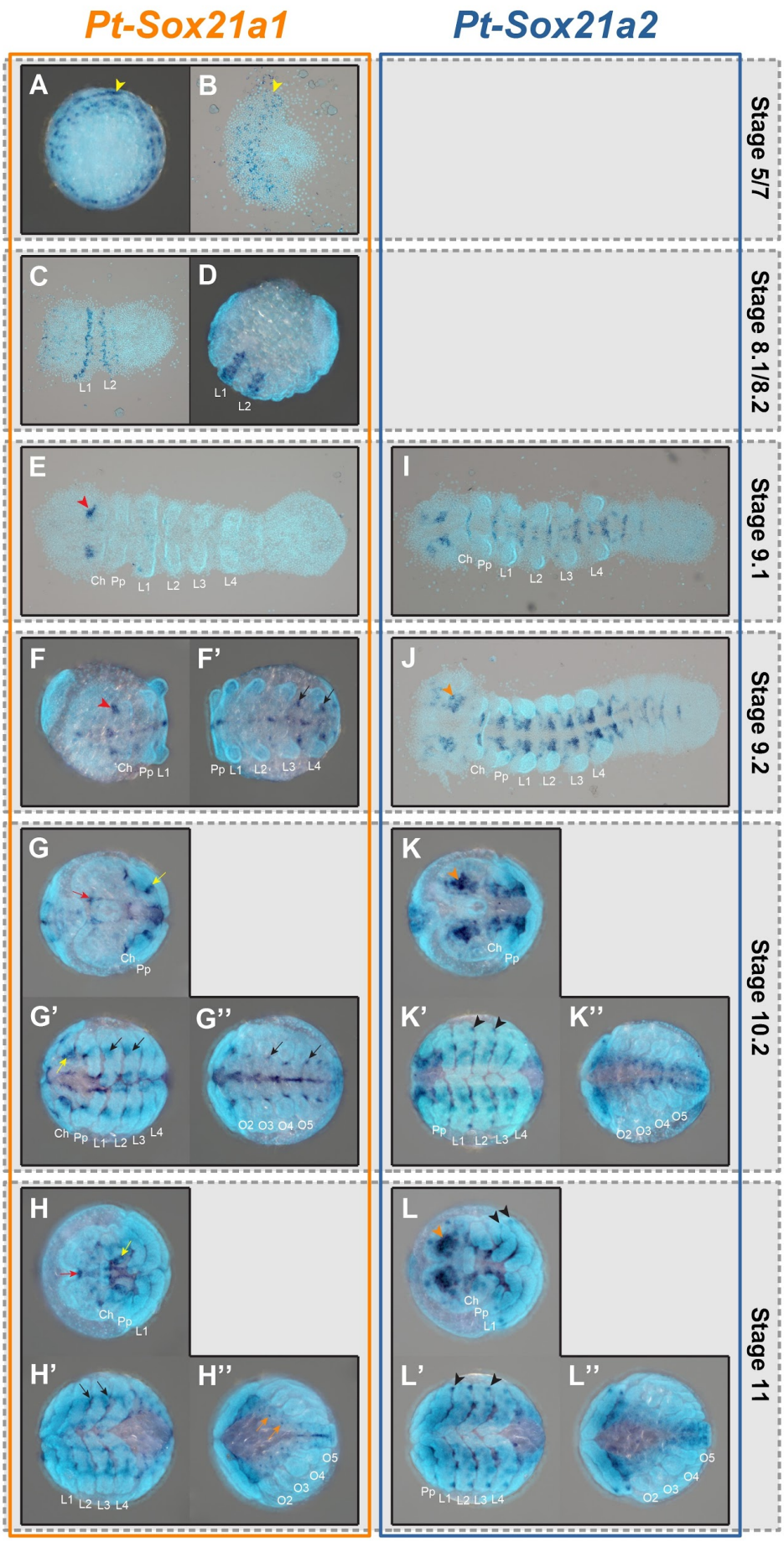
Pt-Sox21a1
Pt-Sox21a2
Stage 5/7
Stage 8.1/8.2
Stage 9.1
Stage 9.2
Stage 10.2
Stage 11
A
B
C
D
E
F
F'
G
G'
G''
H
H'
H''
I
J
K
K'
K''
L
L'
L''
Ch
Pp
L1
L2
L3
L4
O2
O3
O4
O5

Figure 4.5 (previous page): Expression patterns of *P. tepidariorum Sox21a* genes. (A-H") Expression pattern of *Pt-Sox21a-1* in stage 5 (A), 7 (B), 8.1 (C), 8.2 (D), 9.1 (E), 9.2 (F-F'), 10.2 (G-G") and 11 (H-H") embryos. (I-L") Expression pattern of *Pt-Sox21a-2* in stage 9.1 (I), 9.2 (J), 10.2 (K-K") and 11 (L-L") embryos. Anterior is to the left in B-L". Embryos in B, C, E, I and J were flat mounted. Arrowheads mark expression of *Pt-Sox21a1* at the edge of the germ disc (yellow) and in the pre-cheliceral region (red), and *Pt-Sox21a2* expression in the pre-cheliceral region (orange) and at the base of the prosomal appendages (black). Arrows mark *Pt-Sox21a1* expression at the dorsal edge of the neuroectoderm (black), around the stomodaeum (red), at the tip of the chelicerae (yellow) and the dotted-like domain in the opisthosoma (orange). Ch, chelicerae; Pp, pedipalps; L1-L4, walking legs 1-4; O1-O5, opisthosomal segments 1-5.

Pt-Sox21a-1 expression can first be observed at stage 5 as a salt-and-pepper pattern around the edge of the germ disk (Fig. 4.5 A), which is retained after the transition from radial to axial symmetry (Fig. 4.5 B). At stage 8.1 and 8.2, expression is restricted to two bands in the L1 and L2 segments (Fig. 4.5 C and D). This pattern fades away by stage 9.1 and a new domain of expression arises in the pre-cheliceral region (Fig. 4.5 E). At stage 9.2, *Pt-Sox21a-1* starts to be expressed at the midline and small clusters of cells along the dorsal edge of the developing neuroectoderm (Fig. 4.5 F-F'). Expression becomes stronger by stage 10.2, with enlarged domains in the prosomal segments (Fig. 4.5 G-G"). An additional domain can be seen at the tips of the chelicerae and surrounding the stomodaeum (Fig. 4.5 G). By stage 11, expression in the pre-cheliceral and prosomal regions increases in complexity (Fig. 4.5 H and H') and midline expression fades away in most of the opisthosomal segments, replaced by a dotted like expression (Fig. 4.5 H").

Pt-Sox21a-2 is mainly expressed in the neuroectoderm, starting at stage 9.1 in the pre-cheliceral region and in segmental clusters of cells (Fig. 4.5 I). This pattern is maintained throughout embryogenesis, with increasing complexity in the developing brain (Fig. 4.5 J-L"). At later stages, an additional domain of expression is found at the base of the prosomal appendages (Fig. 4.5 K' and L'). Interestingly, *Pt-Sox21a-2* expression pattern appears to be mostly non-overlapping with *Pt-Sox21a-1* expression pattern, especially during earlier stages.

4.2.2.3. Expression patterns of *SoxN* genes

Only one copy of *SoxN* was found in both *P. opilio* and *P. tepidariorum*. *Po-SoxN* and *Pt-SoxN* are strongly expressed in the developing neuroectoderm (Fig. 4.6 A-H'). This is consistent with the results of the previous study of *Pt-SoxN*[65]. An additional domain is present at later stages at the tips and base of the prosomal appendages (Fig. 4.6 B-D and G-H'). Overall, there are no clear differences in expression between the *SoxN* orthologs of *P. opilio* and *P. tepidariorum*, a result consistent with the highly conserved function of this gene in all arthropods studied so far[65,175,176].

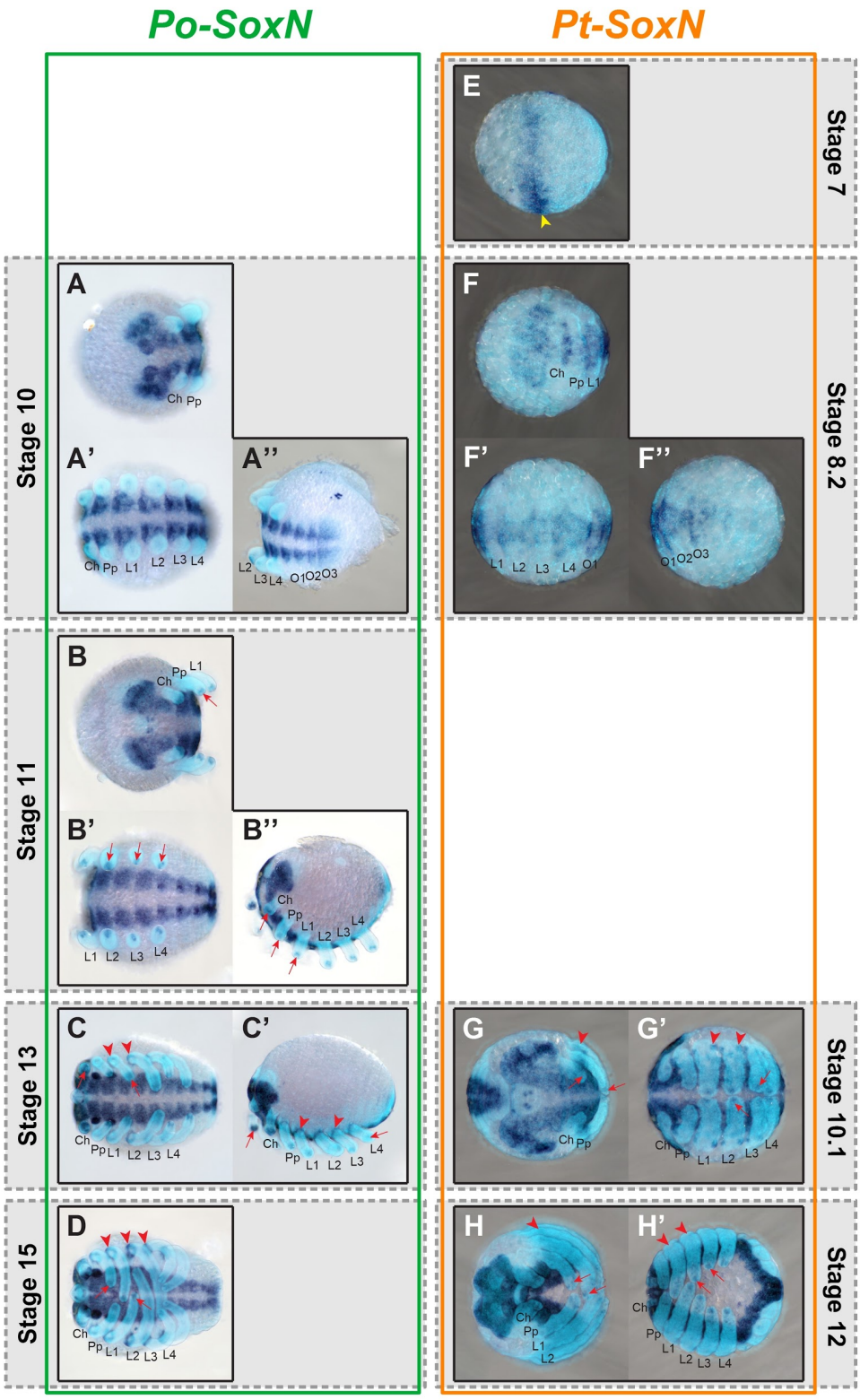

Figure 4.6: Expression patterns of *P. opilio* and *P. tepidariorum SoxN* genes. (A-D) Expression pattern of *Po-SoxN* in stage 10 (A-A"), 11 (B-B"), 13 (C, C') and 15 (D) embryos. (E-H') Expression pattern of *Pt-SoxN* in stage 7 (E), 8.2 (F, F"), 10.2 (G-G') and 12 (H, H') embryos. Anterior is to the left in all images. Arrowheads mark early band of expression in the anterior of the germ band (yellow) and expression at the base of the prosomal appendages (red). Arrows mark expression at the tips of the prosomal appendages. Ch, chelicerae; Pp, pedipalps; L1-L4, walking legs 1-4; O1-O5, opisthosomal segments 1-5.

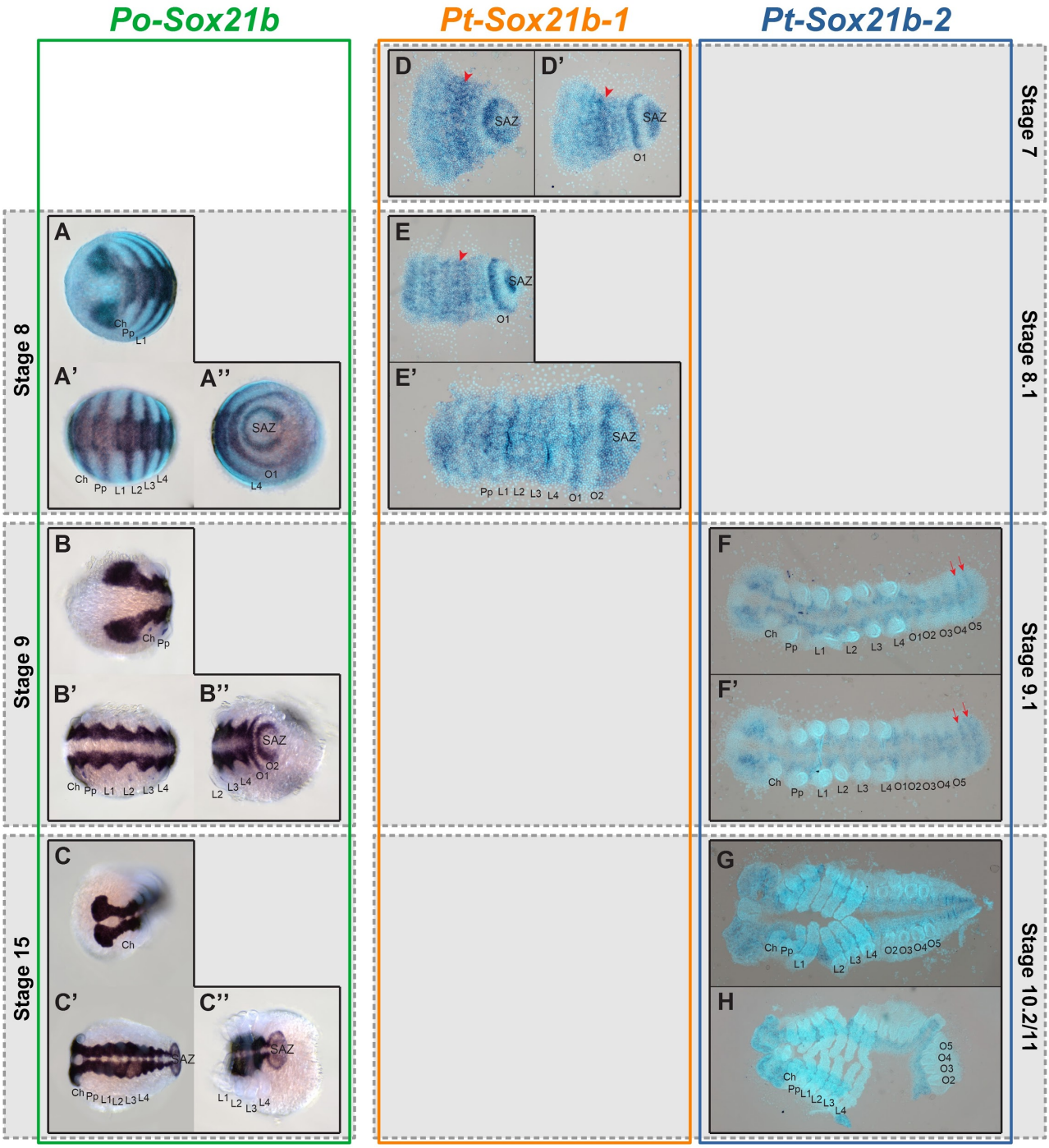

Figure 4.7: Expression patterns of *P. opilio* and *P. tepidariorum Sox21b* genes. (A-C") Expression pattern of *Po-Sox21b* in stage 8 (A-A"), 9 (B-B") and 15 (C-C") embryos. (D-E') Expression pattern of *Pt-Sox21b-1* in stage 7 (D, D') and 8.1 (E, E') flatmounted embryos. (F-H) Expression pattern of *Pt-Sox21b-2* in stage 9.1 (F, F'), 10.2 (G) and 11 (H) flatmounted embryos. Anterior is to the left in all images. Arrowheads mark *Pt-Sox21b-1* expression in the presumptive L2-L4 segments. Arrows mark *Pt-Sox21b-2* segmental expression in the last two opisthosomal segments. SAZ, segment addition zone; Ch, chelicerae; Pp, pedipalps; L1-L4, walking legs 1-4; O1-O5, opisthosomal segment 1-5.

4.2.2.4. Expression patterns of *Sox21b* genes

A single copy of *Sox21b* is present in *P. opilio* and three copies were previously identified in *P. tepidariorum*[65]. *Po-Sox21b* expression is remarkably similar to that of the previously characterised expression pattern of *Pt-Sox21b-1*[65]. Both these genes are expressed in a segmental pattern during early stages and in the neuroectoderm at later stages (Fig. 4.7 A-E'). ISH for *Pt-Sox21b2* produced a faint signal in the developing neuroectoderm, suggesting this paralog is expressed at lower levels in this tissue (Fig. 4.7 F-H). Additionally, at stage 9.1, expression appears to be segmental in the last two segments in the opisthosoma (Fig. 4.7 F-F'). No signal was detected for *Pt-Sox21b-2* during earlier stages of embryogenesis. In summary, it appears that the putative ancestral patterns, here represented by *Po-Sox21b*, have been mostly conserved in *Pt-Sox21b-1*, with *Pt-Sox21b-2* only maintaining the neuroectoderm expression at seemingly lower levels.

4.2.2.5. Expression patterns of *SoxC* genes

The Sox Group C is represented by a single copy in *P. opilio* and two copies in *P. tepidariorum*. As in the previous study[65], *Pt-SoxC2* ISH produced no visible staining and I observed the same neuroectodermal expression pattern for *Pt-SoxC1* (Fig. 4.8 C-E). In addition to the expression domains previously described[65], *Pt-SoxC1* is also expressed in the opisthosomal appendages during later stages (Fig. 4.8 E-F') and in the remaining extraembryonic tissue during dorsal closure (Fig. 4.8 F).

The same expression pattern in the developing neuroectoderm, with more pronounced expression in the clusters of differentiating neuronal progenitors, was observed for *Po-SoxC* (Fig. 4.8 A-B). At later stages, *Po-SoxC* expression is also detected in the prosomal appendages (Fig. 4.8 B), a domain also present at late stages for *Pt-SoxC1* (Fig. 4.8 E and F').

4.2.2.6. Expression patterns of *SoxD* genes

The single Group D gene of *P. opilio*, *Po-SoxD*, is mainly expressed in the prosomal appendages and as dorsally restricted segmental stripes (Fig. 4.9 A-B"). Expression is at first restricted to the centre of the developing appendages, forming a bifurcated pattern at the tips (Fig. 4.9 A-A"). This pattern is maintained in the chelicerae at later stages, developing into 2-4 rings of expression in the pedipalps and walking legs (Fig. 4.9 B-B"). Expression at the edge and above the anterior furrows of the pre-cheliceral region is also observed at later stages (Fig. 4.9 B).

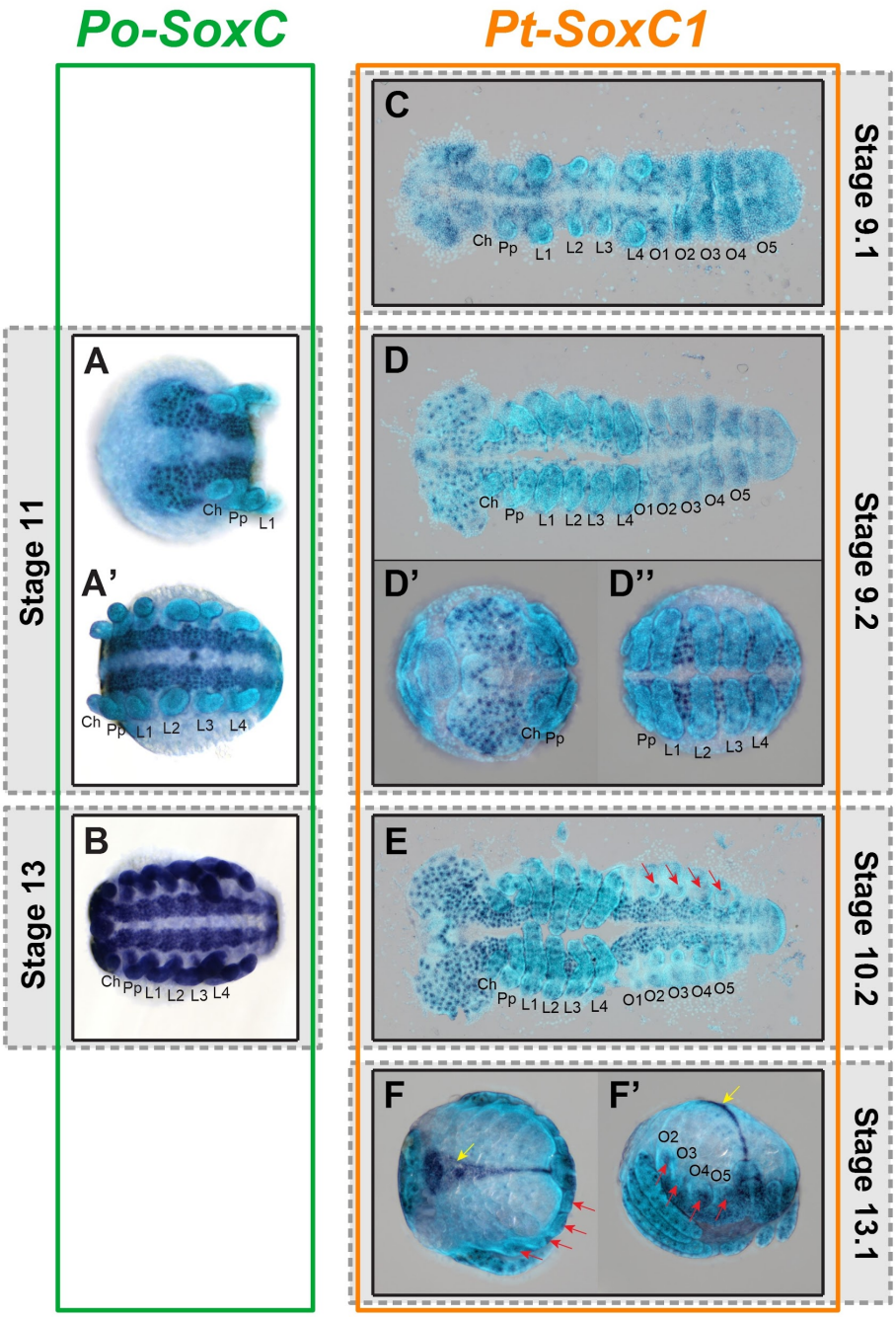

Figure 4.8: Expression patterns of *P. opilio* and *P. tepidariorum SoxC* genes. (A-B) Expression pattern of *Po-SoxC* in stage 11 (A-A') and 13 (B) embryos. (C-F') Expression pattern of *Pt-SoxC1* in stage 9.1 (C), 9.2 (D-D''), 10.2 (E) and 13.1 (F, F') embryos. Anterior is to the left in all images. Embryos in C, D and E were flat mounted. Arrows mark expression in the opisthosomal organs (red) and in extraembryonic tissue (yellow). Ch, chelicerae; Pp, pedipalps; L1-L4, walking legs 1-4; O1-5, opisthosomal segment 1-5.

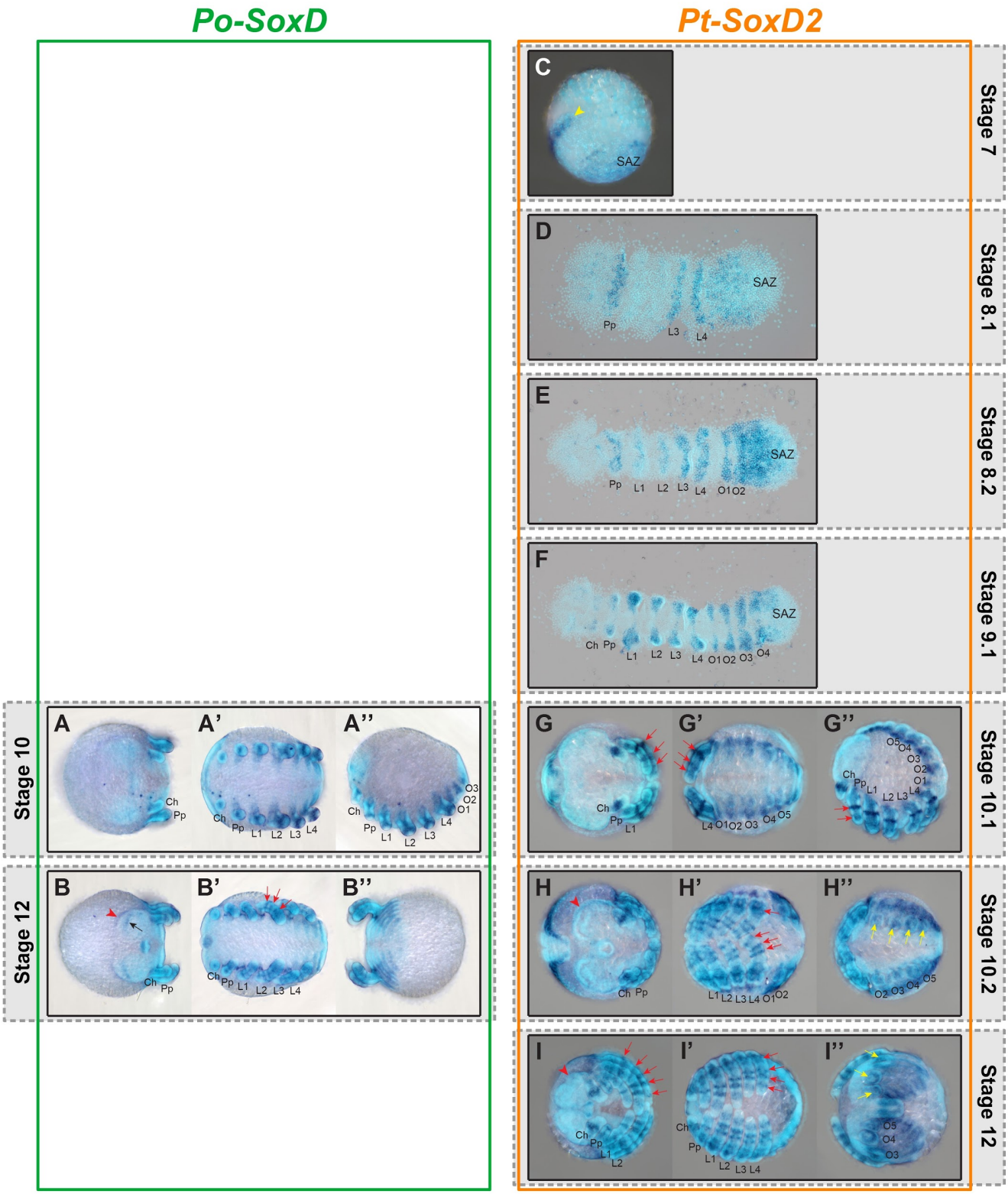

Figure 4.9: Expression patterns of *P. opilio* and *P. tepidariorum SoxD* genes. (A-B") Expression pattern of *Po-SoxD* in stage 10 (A-A") and 12 (B-B") embryos. (C-I") Expression pattern of *Pt-SoxD2* in stage 7 (C), 8.1 (D), 8.2 (E), 9.1 (F), 10.1 (G-G"), 10.2 (H-H") and 12 (I-I") embryos. Anterior is to the left in all images. Embryos in D, E and F were flatmounted. Arrowheads mark expression at the edge of the pre-cheliceral region (red) and the early band of expression in the anterior of the germ band (yellow). Arrows mark expression above the anterior furrows (black), the ring-like expression in the prosomal appendages (red) and expression in the opisthosomal organs (yellow). SAZ, segment addition zone; Ch, chelicerae; Pp, pedipalps; L1-L4, walking legs 1-4; O1-5, opisthosomal segment 1-5.

This pattern of expression is conserved in *Pt-SoxD2* (Fig. 4.9 G-I"). *Pt-SoxD2* expression starts at stage 7 as an anterior stripe and at the posterior of the germ band (Fig. 4.9 C). At stage 8.1, expression is restricted to the pedipalp, L3 and L4 segments, and anteriorly to the SAZ (Fig. 4.9 D). Stripes of expression are continuously added to opisthosomal segments and additional stripes are now detected in the L1 and L2 segments (Fig. 4.9 E). By stage 10.1, expression becomes dorsally restricted (Fig. 4.9 G"). Expression in the prosomal appendages is similar to that of *Po-SoxD*, with expression in several rings during later stages (Fig. 4.9 G-I"). Expression in the opisthosoma becomes restricted to the opisthosomal appendages and the tissue dorsal to these (Fig. 4.9 G', H" and I"). Consistent with *Po-SoxD*, expression can also be found at the edge of the pre-cheliceral region, but not above the anterior furrows (Fig. 4.9 H and I). PCR amplification of *Pt-SoxD1* was unsuccessful despite several attempts.

These results differ with those of the previous study, which only found *Pt-SoxD1* to be expressed in the neuroectoderm, an expression domain not detected for the single-copy of *P. opilio*[65]. This is suggestive of a scenario of neofunctionalization, with *Pt-SoxD2* maintaining the putative ancestral pattern while *Pt-SoxD1* gained new expression in the developing neuroectoderm.

4.2.2.7. Expression patterns of *SoxE* genes

SoxE is the only gene for which more than one sequence was found in *P. opilio*, but unfortunately, I was not able to amplify either sequence by PCR. Nevertheless, I carried out ISH for both copies of *P. tepidariorum SoxE*. In contrast to the previous study, which only detected expression for *Pt-SoxE1*[65], my results show that both duplicates are actually expressed during embryogenesis.

Pt-SoxE1 expression is first visible at stage 9.1 in the developing prosomal appendages of the L1, L2 and L4 segments (Fig. 4.10 A-A"). Expression extends to the other prosomal appendages, and by stage 10.2 it becomes restricted to the mesodermal tissue (Fig. 4.10 B'), with the exception of the chelicerae where it is restricted to a small group of cells at the base (Fig. 4.10 B). This pattern remains unchanged during later stages (Fig. 4.10 C and C'). Two additional expression domains were observed at stage 10.2, in the pre-cheliceral region (Fig. 4.7 B) and in the last pair of opisthosomal organs (Fig. 4.10 B"). Expression in the developing head increases in complexity during later stages (Fig. 4.10 C). At stage 12, faint expression is visible in all opisthosomal appendages and in the expanding dorsal tissue (Fig. 4.10 C").

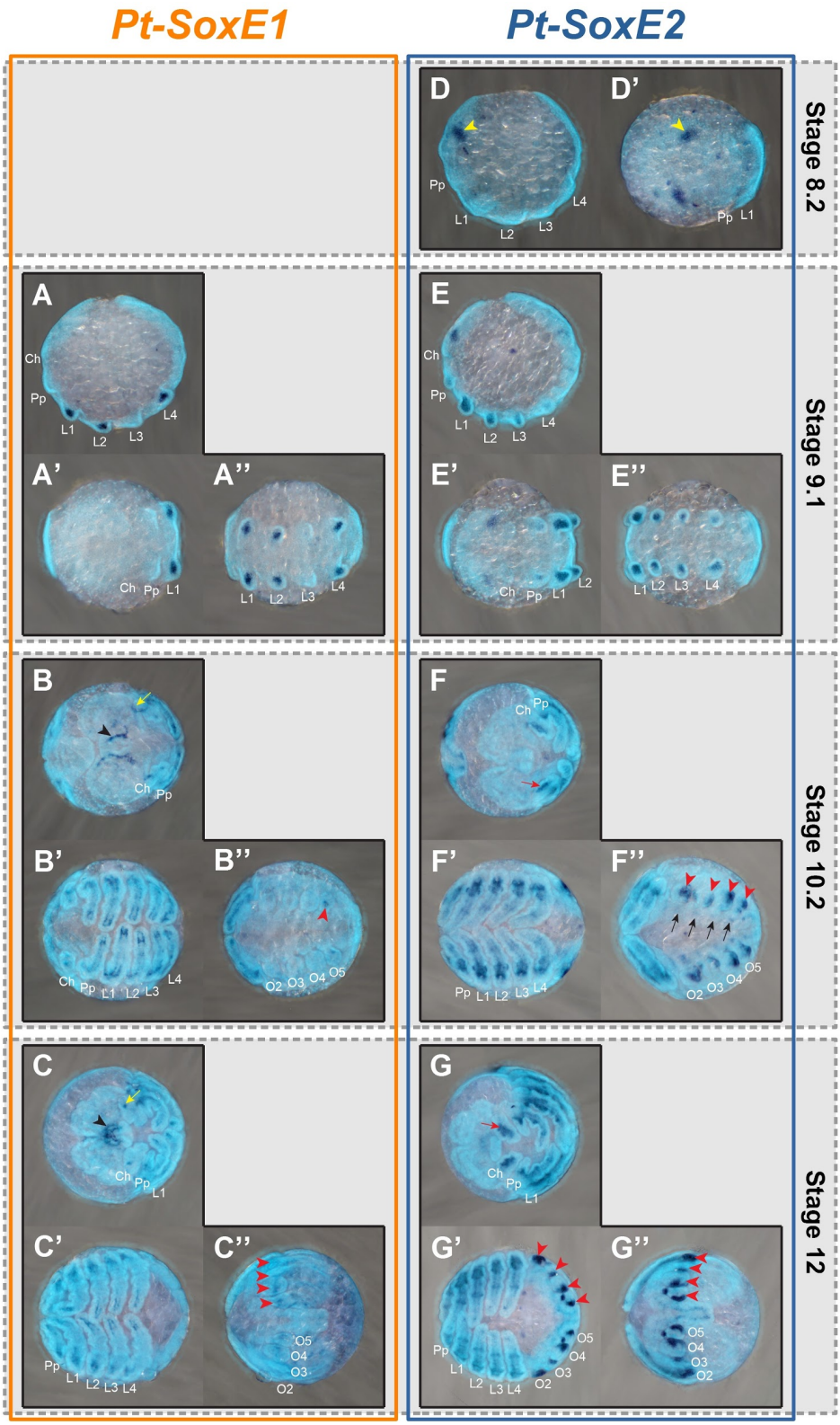

Figure 4.10: Expression patterns of *P. tepidariorum SoxE* genes. (A-C") Expression pattern of *Pt-SoxE1* in stage 9.1 (A-A"), 10.2 (B-B") and 12 (C-C") embryos. (D-G") Expression pattern of *Pt-SoxE2* in stage 8.2 (D, D'), 9.1 (E, E"), 10.2 (F-F") and 12 (G-G") embryos. Anterior is to the left in all images. Arrowheads mark different expression domains of *Pt-SoxE1* (black) and *Pt-SoxE2* (yellow) in the pre-cheliceral region, and expression in the opisthosomal organs (red). Arrows mark different expression domains of *Pt-SoxE1* (yellow) and *Pt-SoxE2* (red) in the chelicerae, and expression of *Pt-SoxE2* in the presumptive germline progenitors (black). Ch, chelicerae; Pp, pedipalps; L1-L4, walking legs 1-4; O1-5, opisthosomal segment 1-5.

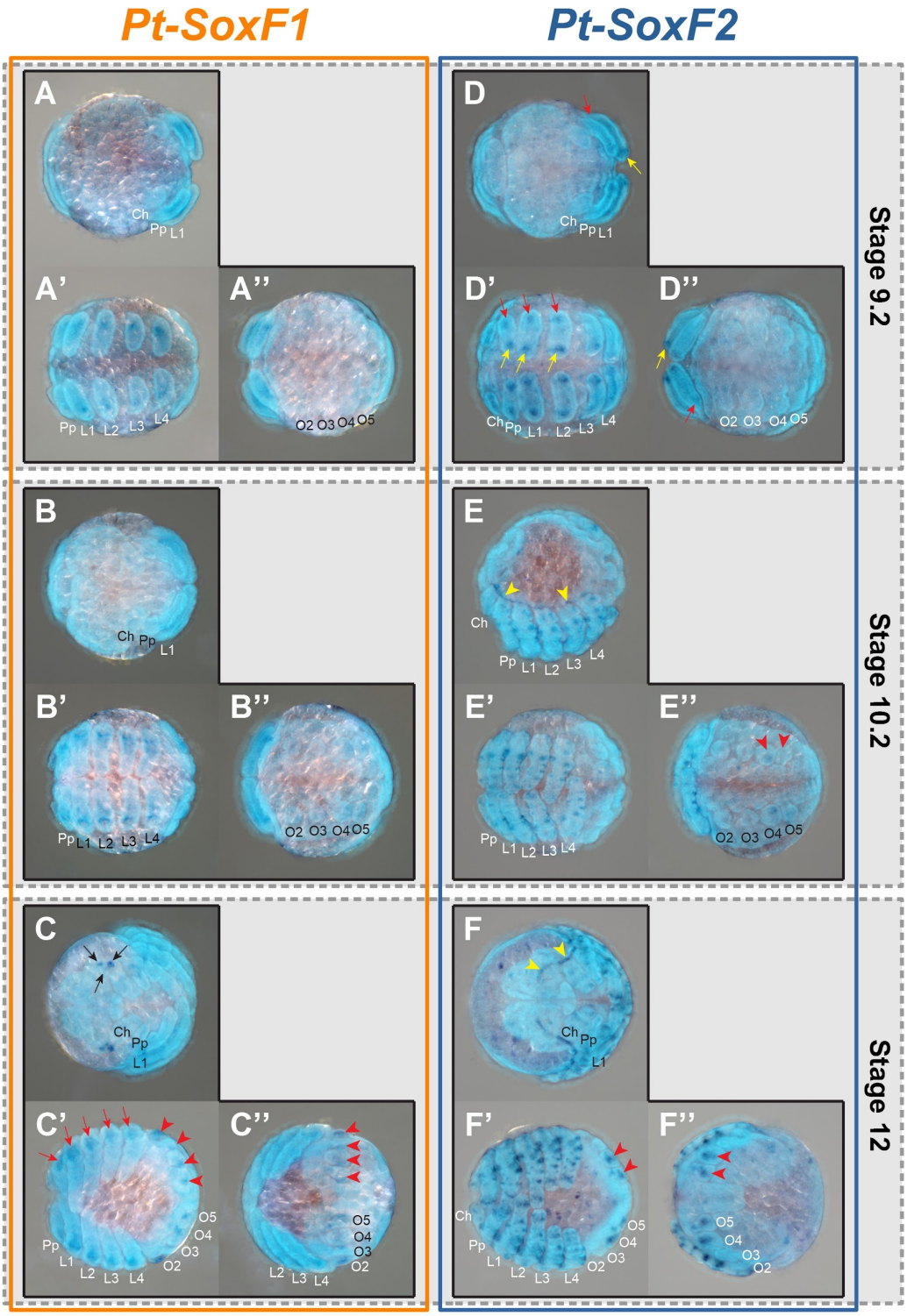

Figure 4.11: Expression patterns of *P. tepidariorum SoxF* genes. (A-C") Expression pattern of *Pt-SoxF1* in stage 9.2 (A-A"), 10.2 (B-B") and 12 (C-C") embryos. (D-F") Expression pattern of *Pt-SoxE2* in stage 9.2 (D-D"), 10.2 (E-E") and 12 (F-F") embryos. Anterior is to the left in all images. Arrowheads mark expression in the opisthosomal organs (red) and at the edge of the non-neurogenic ectoderm (yellow). Arrows mark expression in the lateral eyes (black), at the base of the prosomal appendages (red) and at the tips of the prosomal appendages (yellow). Ch, chelicerae; Pp, pedipalps; L1-L4, walking legs 1-4; O2-5, opisthosomal segment 2-5.

Pt-SoxE2 ISH staining revealed expression in the prosomal appendages similar to that of *Pt-SoxE1*, with the exception of the chelicerae, where expression extends to the tip (Fig. 4.10 D-G"). At stage 8.2, *Pt-SoxE2* is expressed in two clusters of cells in the pre-cheliceral region (Fig. 4.10 D and D'), which starts fading at stage 9.1 (Fig. 4.10 E) and is no longer visible during later stages (Fig. 4.10 F). Highly specific and strong expression can be seen in all opisthosomal appendages (Fig. 4.10 F", G' and G"). Faint expression is also present in small clusters of ventral cells in the opisthosoma (Fig. 4.10 F"), possibly in the germline progenitors[194].

4.2.2.8. Expression patterns of *SoxF* genes

No copies of *SoxF* were found in the *P. opilio* transcriptome. Nonetheless, I characterised the expression pattern of both previously identified copies of *P. tepidariorum*[65]. As with Group E genes, my results were different than those of the previous study[65]. Faint expression of *Pt-SoxF1* is first visible at stage 9.2 in the prosomal appendages (Fig. 4.11 A-A"). This pattern remains unchanged up to stage 12 (Fig. 4.11 B-C"), when it becomes restricted to the base of the legs (Fig. 4.11 C'). At stage 12, additional expression domains arise in the lateral eye primordia (Fig. 4.11 C) and in the opisthosomal appendages (Fig. 4.11 C").

Pt-SoxF2 is also expressed in the prosomal appendages, starting at stage 10.1 in the tips and base of each appendage (Fig. 4.11 D-D"). This pattern is more complex at later stages and becomes restricted to several small groups of cells that possibly belong to the developing PNS (Fig. 4.11 E' and F'). At stage 10.2, *Pt-SoxF2* starts being expressed along the edge of the prosomal region of the germ band, from the cheliceral segment to the L4 segment (Fig. 4.11 E). This domain extends to the pre-cheliceral region at later stages, in a pattern that seems to follow the edge of the growing non-neurogenic ectoderm (Fig. 4.11 F). Lastly, *Pt-SoxF2* is also expressed in the last two pairs of opisthosomal appendages – the presumptive spinnerets (Fig. 4.11 E" and F").

4.3. Discussion

4.3.1. Evolution of Sox gene repertoires in arachnids

Sox genes encode a family of transcription factors with critical roles in animal development[171,176–178]. As such, their evolution and divergence throughout Metazoa, but particularly in vertebrates, has been the focus of several studies[173,174,176,186,187]. Sox gene number vary considerably among animal lineages, with a substantial increase in the vertebrates due to WGD events[173,176,189]. Systematic analysis of Sox gene repertoires has only been carried out in a few arthropod lineages[175,176], for example, in arachnids Sox genes have only been identified in two species of spiders[65]. Thus, the sampling of Sox genes in additional arthropod species is necessary to better understand their evolution in this group of animals, particularly in the arachnids because a WGD event has been identified at the base of the arachnopulmonate lineage. For this purpose, I surveyed available data for the Sox gene repertoires of a scorpion, *C. sculpturatus*, a harvestmen, *P. opilio*, a tick, *I. scapularis* and a centipede, *S. maritima*.

Overall, non-WGD species (i.e. *P. opilio*, *I. scapularis* and *S. maritima*) appear to have a similar composition of Sox genes to that of other non-WGD arthropods (i.e. *G. marginata*, *T. castaneum* and *D. melanogaster*), and the Sox repertoire of *C. sculpturatus* is consistent with duplicate retention after WGD, similar to the previously identified repertoires of *P. tepidariorum* and *S. mimosarum* (Fig. 4.1). These results, in addition to other previously described arthropod Sox repertoires[65,175], are suggestive of a scenario of duplicated Sox gene retention after WGD similar to what has happened in the vertebrate lineage[187,189].

Nevertheless, this survey also suggests the existence of lineage-specific variation in Sox gene number. For Group B, a reasonably well-supported monophyletic clade of Group B genes was obtained after removal of highly divergent sequences (Fig. 4.3). Although, within Group B, only *SoxN* sequences form a well supported clade, a result previously observed in other studies[175,195]. This result further supports the previously proposed separation of *SoxN* from *Dichaete*/*Sox21a*/*Sox21b* into the SoxB1 and SoxB2 subgroups respectively[195]. *SoxN* and *Sox21b* appear to have lineage-specific retention of ohnologs in *C. sculpturatus* and *P. tepidariorum* respectively. Additionally, I was able to identify a copy of *Sox21b* in *S. maritima*, suggesting that an ortholog of this Sox gene may in fact be present in *G. marginata*, contrary to what was previously described[175]. I was unable to identify a *Sox21a* ortholog in *P. opilio*, however, this is probably due to incomplete transcriptomic data, as at least one ortholog is present in all other arthropods surveyed so far[174–176]. The *SoxB-like* sequence found in *S. maritima* appears to be a real Group B ortholog, as it groups well with other *SoxB* sequences. The same cannot be said for *SoxB-like* sequences found in *C. sculpturatus* and *I. scapularis*. A more rigorous analysis of these *SoxB-like* sequences is needed to determine whether these represent bona fide Sox Group B genes.

More than two copies of *SoxC* are present in both *C. sculpturatus* and *S. mimosarum*, possibly representing independent lineage-specific duplications (Fig. 4.1). The phylogenetic analysis supports this scenario because sequences from each species group together independently of other sequences, with only the exception of *Sm-SoxC1* (Fig. 4.2). Group C proteins in general exhibit a higher level of sequence divergence than other Sox proteins, which is reflected in the phylogeny by their long branches and lower bootstrap values (Fig. 4.3). This hints at a greater extent of functional divergence in this group, supported by the highly variable expression of arthropod *SoxC* genes observed in previous studies[174–176], as well as the novel function in *D. melanogaster* in the process of metamorphosis[188]. A similar scenario is seen in vertebrates, where Group C genes likely acquired new roles during development[171].

Interestingly, at least two copies of *SoxE* are present in most species analysed, with the only exceptions being *D. melanogaster*, *T. castaneum* and *E. kanangrensis* (Fig. 4.1). Despite the fact that the identified *I. scapularis* and *S. maritima* sequences may represent fragments of the same transcript, these results are consistent with data from *P. opilio* and *G. marginata*[175]. Curiously, *SoxE* was also found to be duplicated in two species of hymenopterans[174]. However, whether these represent lineage-specific duplications or a possible ancestral duplication requires further investigation. Similarly to *I. scapularis* sequences, the four *SoxE* sequences found in *C. sculpturatus* may be in fact different fragments of two transcripts. Lastly, the two *SoxE-like* sequences found in *C. sculpturatus* and *S. maritima* do not group properly with any other Sox sequences and so these need to be studied further to discern if these are genuine Sox genes.

I was unable to find *SoxF* orthologs in *P. opilio* and *I. scapularis*, though this is possibly the consequence of incomplete datasets because Sox gene loss was rarely detected in most non-WGD lineages analysed so far[176]. A possible lineage-specific duplication of Group F is found in *S. maritima*, however, I cannot discard the possibility that the identified sequences are actually fragments of the same *SoxF* transcript. Loss of a *SoxF* duplicate appears to be specific to *S. mimosarum*, as at least two *SoxF* orthologs are present in *C. sculpturatus* genome.

4.3.2. Evolution and divergence of Sox gene expression in arachnids

To ascertain the potential impact of the arachnopulmonate WGD on Sox gene expression, I characterised the expression pattern of Sox genes in the spider *P. tepidariorum* and in the harvestmen *P. opilio*. As with the homeobox genes analysed above, any differences in expression observed between *P. opilio* single-copy genes and *P. tepidariorum* duplicates could possibly indicate functional divergence during embryogenesis.

In contrast to the previous study, I was able to detect expression of at least one paralog of each duplicated pair of Sox genes during *P. tepidariorum* embryogenesis[65]. Moreover, I was

also able to detect several undescribed expression domains for some of the previously characterized genes[65]. Overall, duplicated Sox genes in *P. tepidariorum* appear to have experienced extensive expression pattern divergence. Once again, the most common outcome seems to be the partitioning of the ancestral expression domains between each paralog (e.g. Group E and F). Although there are no clear cases of neofunctionalisation, several of these paralogs are expressed in the developing book lungs and spinnerets (*Pt-SoxC1*, *Pt-SoxD2*, *Pt-SoxE1*, *Pt-SoxE2*, *Pt-SoxF1* and *Pt-SoxF2*), both innovations of the arachnopulmonate and spider lineages respectively. Interestingly, several of the subfunctionalised patterns appear to be associated with nervous system development, with each duplicate possibly having acquired specialised roles in the differentiation of specific cell subsets, an outcome similar to that of duplicated Sox genes in vertebrates[171,186].

A conserved role of SoxB2 genes during panarthopod segmentation has been previously proposed, with *Dichaete* performing this role in mandibulates and *Sox21b-1* in spiders[65,66,175]. My results are consistent with this hypothesis, with additional evidence from *P. opilio* that suggests a conserved role for *Sox21b* in arachnid segmentation. In addition to this, *Pt-Sox21b2* expression is suggestive of an overlapping role in CNS development. However, it is still unclear whether *Dichaete* has a role in segmentation or not, because *Pt-D* appears to be segmentally expressed in a few anterior segments during early stages of embryogenesis. Characterisation of additional arachnid orthologs of *Dichaete*, such as *Po-D* in *P. opilio*, and further analysis of *Pt-D* function would likely help answering this question.

Both *Sox21a* paralogues have largely non-overlapping expression patterns, hinting at a possible subfunctionalisation. Identification and characterization of *Po-Sox21a* will allow us to determine whether these patterns represent separate instances of the ancestral domain or novel expression domains. Previous data on *Sox21a* expression in other arthropods is suggestive of an ancestral expression pattern at the ventral midline[175,176]. *SoxN* expression appears to be conserved in the CNS[175] of both *P. opilio* and *P. tepidariorum* but both species have expression at the tips of the prosomal appendages, implying a possible lineage-specific function in arachnids.

Arthropod Sox Group C orthologs appear to have conserved roles in central and peripheral nervous system development[175], a role apparently retained by *Pt-SoxC1*. Even though *Pt-SoxC2* expression was not detected in both studies, this paralog could still be expressed in earlier stages of embryogenesis, or during juvenile and/or adult stages.

SoxD genes appear to have conserved roles in mesodermal and appendage development, reflected in their conserved expression patterns in these tissues observed in several arthropods[175]. *Po-SoxD* and *Pt-SoxD2* have remarkably similar patterns, indicating these roles might be conserved throughout the arthropods.

The Group E and F duplicates appear to be clear examples of subfunctionalisation. Despite the lack of data on ortholog expression in *P. opilio*, the combined expression of *P. tepidariorum* paralogs strikingly mirror previously described expression patterns of the corresponding orthologs in other arthropods, such as *G. marginata* and *T. castaneum*[175]. Furthermore, *Pt-SoxE2* is strongly expressed in the developing opisthosomal appendages, alluding to a possible role in the development of these novel structures. Curiously, *Pt-SoxF1* is expressed in the developing lateral eyes of *P. tepidariorum*, an expression domain also detected for the ortholog of *G. marginata*[175].

To conclude, my results demonstrate that the duplicated Sox genes identified in *P. tepidariorum* genome experienced extensive expression pattern divergence after the arachnopulmonate WGD. Furthermore, it appears that the putative ancestral expression patterns of arthropod Sox genes (i.e. the expression patterns of Sox genes in *P. opilio* and previously characterised orthologs in other arthropod species[174–176]) was partitioned between the ohnologs present in *P. tepidariorum*, suggesting that subfunctionalization rather than neofunctionalization was the most common fate of duplicated Sox genes after WGD. Nevertheless, some of these genes may be involved in the development of arachnopulmonate-specific traits, such as the book lungs and the spinnerets of spiders.

5. Chapter III: Role of *Pt-Sox21b-1* in spider segmentation

5.1. Introduction

The subdivision of the anterior-posterior body axis into metameric segments, a process termed segmentation, is a distinct characteristic of three major animal phyla: arthropods, annelids and vertebrates[196,197]. It is currently believed that the remarkable morphological diversity found in these taxa owes much to the modular organization of their body plan, which allowed the flexible adaptation of different segments to fulfil particular functions[196–198]. Thus, the origin and evolution of segmentation has long been one of the main focal subjects of Evo-Devo studies[196–198].

Whether segmentation has a common origin at the base of the Bilateria or arose independently twice or in each phylum is still a much-debated topic[179,199–201]. In order to address this problem, much research has been focused on characterising the gene regulatory networks (GRNs) controlling this crucial process in a wide range of segmented species[179,199–203]. The discovery of apparently conserved signalling pathways regulating vertebrate and arthropod segmentation, such as Wnt and Delta-Notch signalling, has been argued by some to be indicative of a common ancestral origin[201,204]. Although, others have suggested this could be achieved by the independent co-option of conserved pathways in each phylum[197,204].

The study and comparison of segmentation among arthropods has been crucial in the characterisation of segmentation mechanisms and their evolutionary origins[196–198]. Arthropod segmentation can be subdivided into two main modes; long germ, where all segments are specified simultaneously, and short germ, where a variable number (depending on the species) of anterior segments are first specified simultaneously, followed by the sequential formation of posterior segments from a growth zone or segment addition zone (SAZ)[203].

Unsurprisingly, segmentation is best understood in *D. melanogaster*, which exhibits long germ segmentation[198,203]. In this higher dipteran, segmentation is regulated by a hierarchical gene cascade, which ultimately subdivides the embryo into individual segments[198,203]. Maternally deposited A-P axis determinants, such as *bicoid* (*bcd*) and *caudal* (*cad*), activate the first tier of zygotically expressed segmentation genes, the gap genes (e.g. *hunchback* (*hb*) and *giant*), which subdivide the embryo into large regions containing several adjacent segments[198,203,205]. Gap genes and maternal genes in turn activate the next tier of the cascade, the pair-rule genes such as *hairy* (*h*) and *even skipped* (*eve*), which subdivide the embryo into periodic series of seven stripes in every second parasegment[179,205]. Lastly, pair-rule genes help to activate the last tier of the hierarchy, the segment-polarity genes *engrailed* (*en*) and

wingless (*wg*), which mark the boundaries of the parasegments and compartments of the final 14 segments[205].

Long germ segmentation is confined to holometabolous insects, while short germ segmentation is extensively represented in other insects and all other major arthropod subphyla (i.e. crustaceans, myriapods and chelicerates)[198,203]. Therefore, long germ segmentation is probably a derived state, while short germ segmentation appears to be closest to the ancestral state of arthropod segmentation[198,203]. It is then essential to investigate the genetic mechanisms controlling short germ segmentation in carefully selected arthropod taxa, to help infer the ancestral mechanisms regulating this process. Over the last two decades, work on short germ insects, myriapods and chelicerates have provided further insights into the evolution of arthropod segmentation, by identifying both novel and conserved developmental genetic mechanisms[52,179,198,206]. Work on spiders in particular has contributed tremendously towards our knowledge on chelicerate segmentation and the evolution of arthropod segmentation[52,201].

5.1.1. Segmentation in the spider *P. tepidariorum*

Over the last two decades, the spider *P. tepidariorum* has been used as a model to study segmentation in chelicerates[52]. Spider segmentation falls within the short germ category[203], with some anterior prosomal segments being formed early in development, more or less simultaneously, while posterior opisthosomal segments are sequentially added from a SAZ[52].

It is thought that the anterior and posterior modes of segmentation are regulated by largely independent GRNs. Formation of the most anterior segments (i.e. pre-cheliceral, cheliceral and pedipalpal segments) is controlled by an autoregulatory GRN composed of *hedgehog* (*hh*), *otd1/Otx1*, *cubitus interruptus* (*ci*), *hairy* (*h*) and *odd-paired* (*opa*)[58,207]. Expression of *otd1*, *ci* and *opa* is necessary for the splitting of anterior stripes of *hh*, *opa* and *h* corresponding to the pre-cheliceral, cheliceral and pedipalpal segments and knockdown of *otd1* expression results in the loss of these segments[58,207].

To date, only three genes have been shown to play a role in the formation of leg-bearing segments, *hb*, *Distal-less* (*Dll*) and *Sox21b-1*[56,208]. *Pt-hb* and *Pt-Dll* are expressed in the early presumptive regions of specific segments, and knockdown of their expression leads to the loss of leg segments corresponding to these early patterns[56,208]. A more detailed description of the role of *Sox21b-1* in anterior segmentation is provided below.

Much more progress has been made towards our understanding of the GRN regulating posterior segmentation in *P. tepidariorum*[52,80,201]. Both *Pt-Wnt8* and *Pt-Delta* (*Pt-Dl*) are required for the establishment and maintenance of the SAZ[57,209]. Knockdown of either of these genes results in truncated embryos, with the absence of all opisthosomal segments[57,209]. *Pt-*

Wnt8 and *Pt-Dl* also interact with each other, with *Pt-Dl* restricting *Pt-Wnt8* expression to the posterior-most region of SAZ, while *Pt-Wnt8* appears to regulate dynamic expression of *Pt-Dl*[57,80]. Both *Pt-Dl* and *Pt-Wnt8* are also necessary to activate *Pt-cad* expression, as demonstrated by the absence of *Pt-cad* expression in embryos lacking *Pt-Dl* or *Pt-Wnt8* expression[57,209]. Like its role in *T. castaneum*, *Pt-cad* is necessary to activate the pair-rule gene orthologues *Pt-eve* and *Pt-run-1*, which are periodically expressed in the SAZ and nascent opisthosomal segments[80].

Most of what we know about other putative segmentation genes in spiders is exclusively based on expression pattern analysis[198]. The repertoire of segment-polarity genes appears to be conserved in spiders (e.g. *en* and *ci*), and their expression is suggestive of a conserved role in defining parasegmental boundaries[111,117], although *Wnt8* may play the segment polarity role of *wg*/*Wnt1* in *P. tepidariorum*[80]. Spider pair-rule orthologs, such as *eve*, *odd-skipped* and *paired*, are amongst the first genes to be expressed in a segmental pattern, suggestive of a conserved role in segmentation[109,206].

However, a recent study discovered that a Sox gene is involved in both anterior and posterior segmentation in *P. tepidariorum*[66]. This appears to be similar to the role of the related SoxB group gene *Dichaete* during segmentation in both *D. melanogaster* and *T. castaneum*[179,210].

5.1.2. The role of *Pt-Sox21b-1* in spider segmentation

A previous study of Sox gene expression in *P. tepidariorum* found *Pt-Sox21b-1* to be expressed early in embryogenesis, in a pattern consistent with a role in segmentation[65]. Functional analysis with parental RNAi revealed a role of this SoxB duplicate in both anterior and posterior segmentation[66]. Knockdown of *Pt-Sox21b-1* results in the loss of all leg segments and disrupts the formation of the SAZ, resulting in truncated embryos missing all segments posterior to the pedipalps[66]. *Pt-Sox21b-1* knockdown embryos lack expression of both *Pt-Dl* and *Pt-Wnt8*, key regulators of SAZ formation, which explains the loss of opisthosomal segments (Fig. 5.1)[66]. *Pt-cad* expression is also absent, although, this could be an indirect effect of the loss of *Pt-Dl* and *Pt-Wnt8*[66]. Conversely, *Pt-Sox21b-1* expression is still present in *Pt-Dl* and *Pt-Wnt8* RNAi knockdown embryos[66]. Therefore, *Pt-Sox21b-1* appears to act upstream of the posterior GRN by regulating *Wnt8* and *Delta-Notch* signalling during SAZ formation (Fig. 5.1)[66].

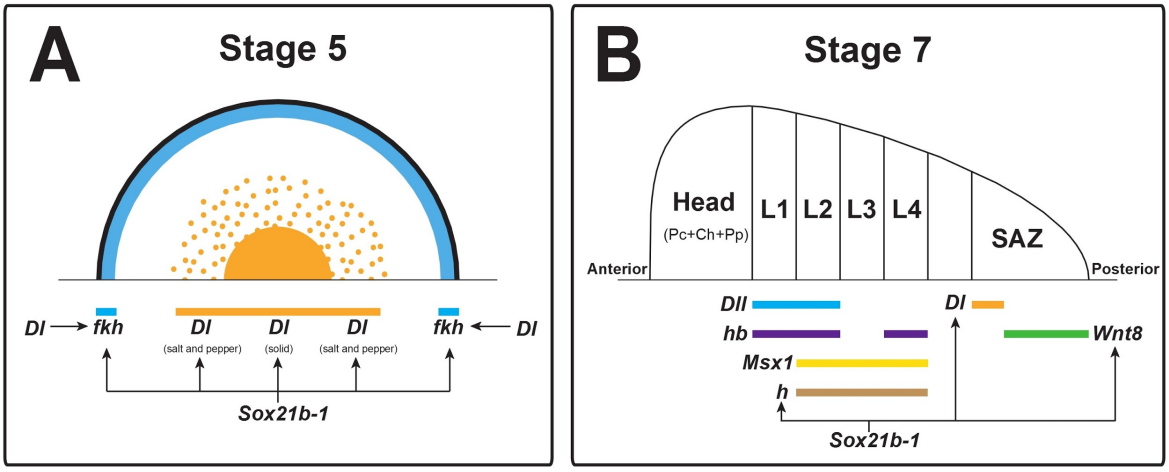

Figure 5.1: Role of *Pt-Sox21b-1* during *P. tepidariorum* segmentation. (A) At stage 5, *Sox21b-1* appears to regulate both *Dl* and *fkh* expression[66]. The *fkh* expression domain at the rim of the germ disc is also regulated by *Dl*[210]. (B) At stage 7, *Sox21b-1* appears to regulate both *Dl* and *Wnt8* expression in the segments addition zone[66]. Additionally, *Pt-Sox21b-1* regulates the expression of *h* in prosomal segments[66]. Pc, Pre-cheliceral segment; Ch, Cheliceral segment; Pp, pedipalpal segment; L1-L4, leg segments 1-4; O1-2, opisthosmal segments 1 and 2; SAZ, segment addition zone.

Even though this study provided a great insight into the roles of *Pt-Sox21b-1*, further investigation is required to better understand the mechanisms by which it acts. It is still uncertain how RNAi induced knockdown of *Pt-Sox21b-1* interferes with the proper formation of leg segments and if its effects are direct or indirect. For example, the only piece of evidence that *Pt-Sox21b-1* directly regulates genes in the GRN controlling leg segment formation, is the absence of *Pt-h* expression in *Pt-Sox21b-1* knockdown embryos[66]. *Pt-h* is expressed at stage 5 in the presumptive regions of the head, the L2-L4 segments, and dynamically in the SAZ[58]. This pattern has been shown to be regulated by other segmentation genes, hinting at a possible role of *Pt-h* in both anterior and posterior segmentation[57,58,208]. It is also unclear how *Pt-Sox21b-1* acts to regulate the posterior segmentation GRN, because, as the authors have suggested, loss of *Pt-Dl* and *Pt-Wnt8* expression in *Pt-Sox21b-1* knockdown embryos could be an indirect consequence of the loss or incorrect specification of cells during early embryogenesis due to the pleiotropic function of *Pt-Sox21b-1* in mesendodermal cell specification[66].

In this chapter, I aimed to further investigate the role of *Pt-Sox21b-1* in the regulation of segmentation by examining the effect of *Pt-Sox21b-1* pRNAi and eRNAi knockdown on the expression of segmentation genes.

5.2. Results

5.2.1. Effect of *Pt-Sox21b-1* RNAi knockdown on anterior segmentation genes

Knockdown of *Pt*-Sox21b-1 expression inhibits the formation of all leg-bearing segments[66]. To further evaluate the effect of *Pt-Sox21b-1* knockdown in the regulation of anterior segmentation I characterised the expression patterns of *Pt-Dll*, *Pt-hb* and *Pt-Msx1*, which all have known roles in this process[56,208], in *Pt-Sox21b-1* knockdown embryos. For these experiments *Pt-Sox21b-1* knockdown was carried out by means of pRNAi, using the same 549 bp dsRNA (fragment 1) used in the previous study[66]. The same range and frequencies of phenotypic classes were observed for all cocoons from injected females[66]. ISH was performed in embryos ranging from stage 5 to 8.2 obtained from the third and fourth cocoons. Because phenotypic class III embryos do not transition from radial to axial symmetry and display major developmental defects, expression analysis was focused on embryos of phenotypic classes I and II[66].

5.2.1.1. Effect of *Pt-Sox21b-1* knockdown on *Pt-Dll* expression

The early expression pattern of *Pt-Dll* and its role in anterior segmentation have been previously described[56]. *Pt-Dll* is expressed in the presumptive region of the L1 segment from as early as stage 5, first as a ring-like pattern which transforms into a stripe at the germ disc to germ band transition (Fig. 5.2 A-D)[56]. Additional expression domains in the SAZ and pre-cheliceral region arise at stages 6 and 7 respectively, as well as a faint stripe in the L2 segment (Fig. 5.2 C)[56]. Subsequently, by stage 8.2 expression is present in all prosomal appendages[56]. pRNAi knockdown of *Pt-Dll* results in the loss of the L1 and L2 segments, revealing the important role of this gene in the patterning of anterior segments[56].

Overall, *Pt-Dll* expression appears to be relatively unaffected by the knockdown of *Pt-Sox21b-1* expression using pRNAi (Fig. 5.2 E-J'). The early ring-like expression domain in L1 is still present in stage 5 knockdown embryos (n=5) and maintained as a stripe up to stage 7 (n=3) (Fig. 5.2 E-G). However, this domain appears to be fainter from stage 8.1 onwards and the stripe of expression in L2 is missing (n=5) (Fig. 5.2 H-I'). The expression domain in the SAZ is reduced at stage 7 (n=3) and completely disappears at later stages (n=8), probably due to the effect of *Pt-Sox21b-1* knockdown in posterior segmentation (Fig. 5.2 G-J'). Ectopic expression of *Pt-Dll* was detected near the posterior of some *Pt-Sox21b-1* knockdown embryos at stage 8.1 (n=2) (Fig. 5.2 H). The pre-cheliceral domain, as well as the expression in the developing appendages at late stages, was still present in all the knockdown embryos analysed (n=8) (Fig. 5.2 H-J').

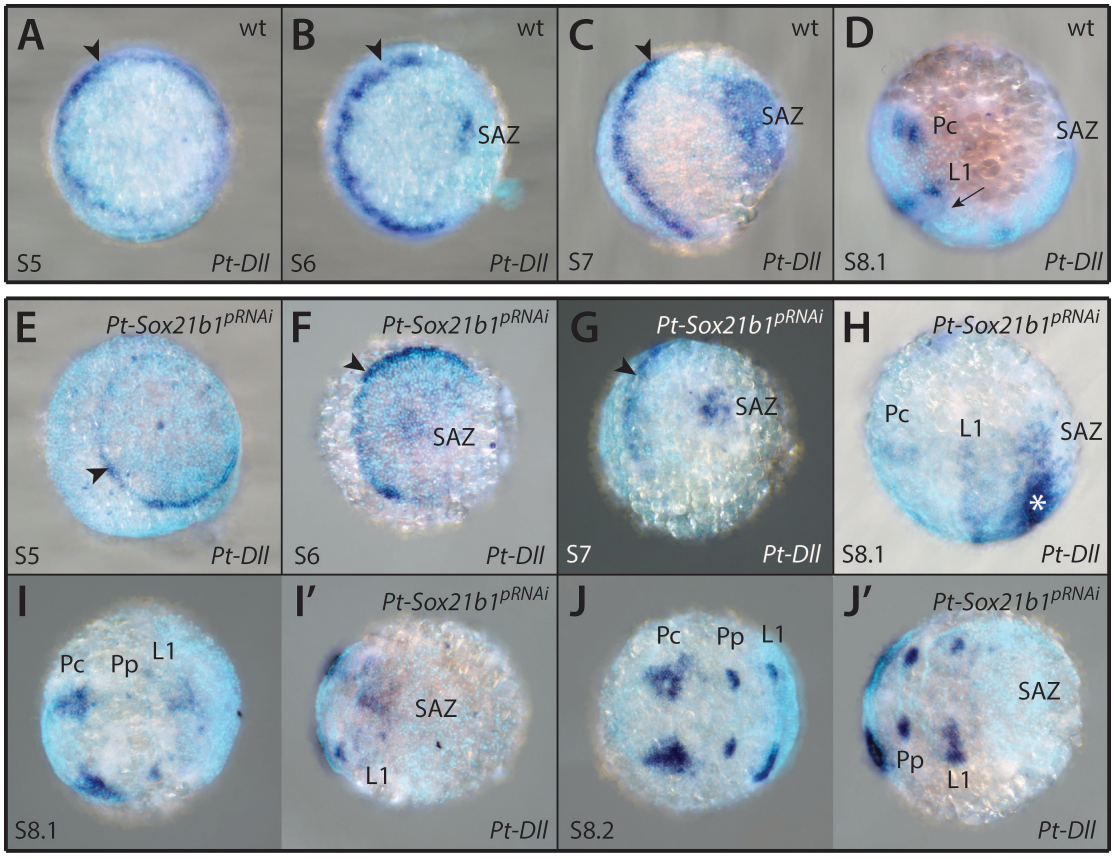

Figure 5.2: Effect of *Pt-Sox21b-1* pRNAi knockdown on *Pt-Dll* expression. (A-D) Expression pattern of *Pt-Dll* in stage 5 (A), 6 (B), 7 (C) and 8.1 (D) wildtype embryos. (E-J') Expression pattern of *Pt-Dll* in stage 5 (E), 6 (F), 7 (G), 8.1 (H, I and I') and 8.2 (J and J') *Pt-Sox21b-1* pRNAi embryos. Anterior is to the left in all images. Arrowheads: ring-like early expression that is later restricted to the presumptive L1 segment. Arrow: faint stripe in the presumptive L2 segment. Asterisk: ectopic expression near the SAZ. Pc, pre-cheliceral region; Pp, presumptive pedipalpal segment; L1, presumptive L1 segment; SAZ, segment addition zone.

In conclusion, it appears that *Pt-Sox21b-1* is not necessary for the activation of *Pt-Dll* expression in pRNAi knockdown embryos, although, some aspects of *Pt-Dll* patterning appear to be affected. However, it is difficult to conclude whether this is a direct or indirect effect because of the pleiotropic effects of *Pt-Sox21b-1* parental knockdown.

5.2.1.2. Effect of *Pt-Sox21b-1* knockdown on *Pt-hb* expression

The embryonic expression and role of *Pt-hb* in anterior segmentation was previously reported[208]. At stage 5, *Pt-hb* is strongly expressed at the periphery of the germ disc and in a faint band corresponding to the presumptive head, L1 and L2 segments[208]. A second band of expression arises at stage 7 in the presumptive L4 region, while the anterior band splits into two bands (head and L1+L2 segments) (Fig. 5.3 A and B)[208]. Subsequently, *Pt-hb* expression becomes segmentally restricted, with a domain covering the pre-cheliceral and cheliceral segments, and as stripes in all other anterior segments, with the strongest expression in the

L1, L2 and L4 segments (Fig. 5.3 C)[208]. Knockdown of *Pt-hb* leads to the loss of the L1 and L2 segments, and disruption of the L4 segment[208].

Pt-hb expression is still present in *Pt-Sox21b-1* pRNAi knockdown embryos, although, the dynamics of *Pt-hb* expression pattern appear to be perturbed (Fig. 5.3 D-J). At stage 6, the strong expression in the periphery of the germ disc and the band of expression in the presumptive region of the head, L1 and L2 segments appears to be unaffected (n=6) (Fig. 5.3 D). At stage 7, splitting of the anterior expression domain into two bands of expression (head and L1+L2 segments) also appears to be unaffected (n=7) (Fig. 5.3 E-H). However, the posterior L4 domain seems to be absent in stage 7 knockdown embryos (n=5) (Fig. 5.3 E and F), or is at least perturbed (n=2) (Fig. 5.2 G and H). During later stages, it appears that the L1+L2 domain does not split, although, the pre-cheliceral and pedipalpal domains appear to be unaffected (n=4) (Fig. 5.3 I and J).

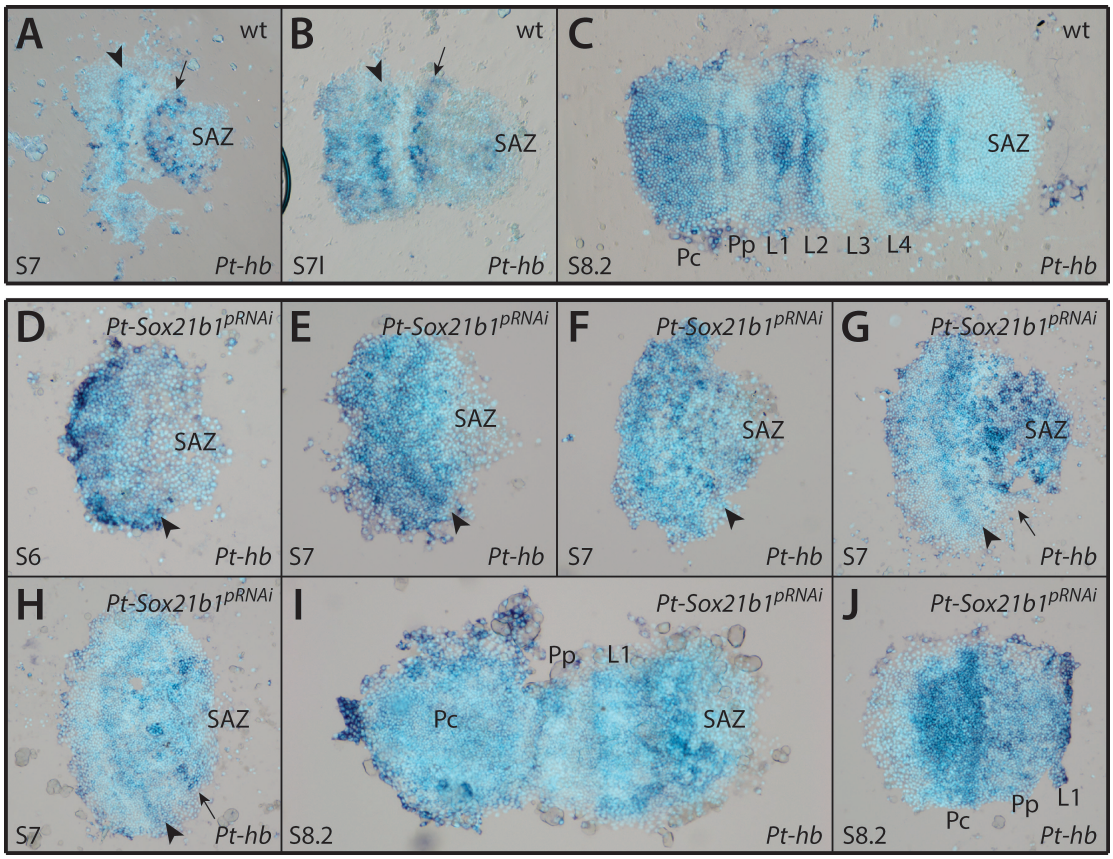

Figure 5.3: Effect of *Pt-Sox21b-1* pRNAi knockdown on *Pt-hb* expression. (A-C) Expression pattern of *Pt-hb* in stage 7 (A), 8.1 (B) and 8.2 (C) wildtype embryos. (D-J) Expression pattern of *Pt-hb* in stage 6 (D), 7 (E-H), 8.1 (I) and 8.2 (J) *Pt-Sox21b-1* pRNAi embryos. Anterior is to the left in all images. Arrowheads: expression in the presumptive L1+L2 segments. Arrows: expression in the presumptive L4 segment. Pc, pre-cheliceral region; Pp, presumptive pedipalpal segment; L1, presumptive L1 segment; L2, presumptive L2 segment; L3, presumptive L3 segment; L4, presumptive L4 segment; SAZ, segment addition zone.

These results suggest that activation of *Pt-hb* expression does not seem to require *Pt-Sox21b-1*, but certain aspects of the *Pt-hb* expression pattern appear to be affected in *Pt-Sox21b-1* knockdown embryos. However, again, it is difficult to ascertain if this is due to a direct regulation of *Pt-hb* by *Pt-Sox21b-1* or an indirect result of cell/segment loss.

5.2.1.3. Effect of *Pt-Sox21b-1* knockdown on *Pt-Msx1* expression

As described in Chapter I, *Pt-Msx1* is expressed in a segmental pattern during early embryogenesis (Fig. 5.4 A-D). *Pt-Msx1* is first expressed at stage 5 in the presumptive region of the L2-L4 segments, forming a band of expression at stage 6 (Fig. 5.4 A and B). This band of expression is still present in early stage 7, when two additional domains of expression arise in the SAZ and in a faint stripe corresponding to the presumptive head segments (Fig. 5.4 C). This anterior stripe of expression becomes stronger by late stage 7 and the L2-L4 domain starts splitting into 3 stripes corresponding to each leg segment (Fig. 5.4 D). By late stage 7, faint expression is still present in the SAZ and a faint stripe can be detected in the first opisthosomal segment (Fig. 5.4 D).

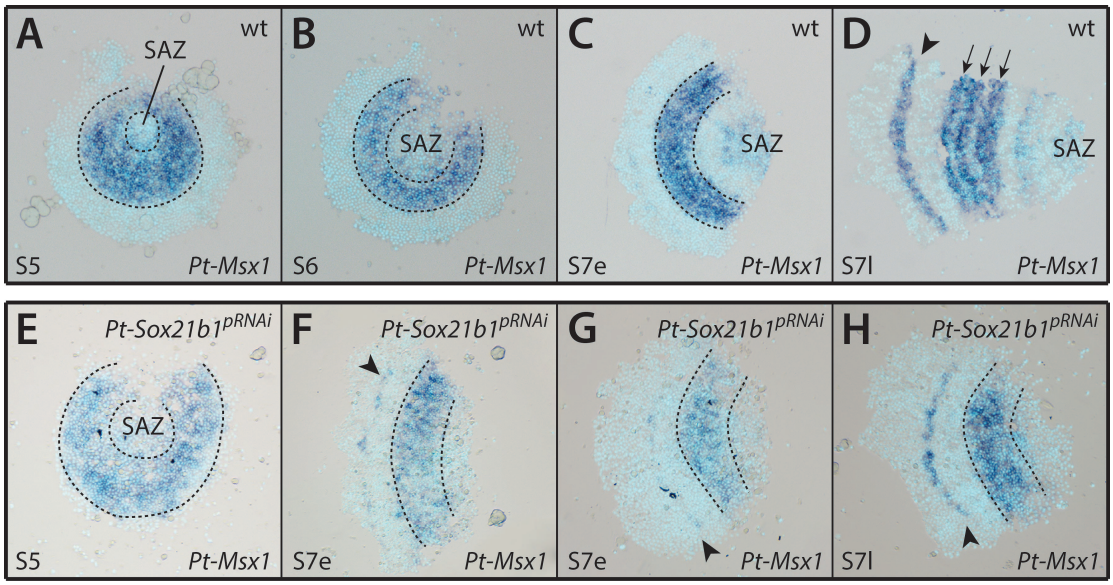

Figure 5.4: Effect of *Pt-Sox21b-1* pRNAi knockdown on *Pt-Msx1* expression. (A-D) Expression pattern of *Pt-Msx1* in stage 5 (A), 6 (B) and 7 (C, D) wildtype embryos. (E-H) Expression pattern of *Pt-Msx1* in stage 5 (E) and 7 (F-H) *Pt-Sox21b-1* pRNAi embryos. Anterior is to the left in all images except A and B. Dashed lines: expression in the presumptive L2-L4 segments. Arrows: splitting stripes of expression in the L2, L3 and L4 presumptive segments. Arrowheads: stripe of expression in the presumptive head segments. SAZ, segment addition zone.

Analysis of *Pt-Msx1* expression in *Pt-Sox21b-1* knockdown embryos revealed a scenario similar to that observed for *Pt-hb* (Fig. 5.4 E-H). *Pt-Msx1* is still expressed at stage 5, although, due to the highly deformed state of the embryos analysed, it is difficult to tell whether or not

the pattern is restricted to the presumptive L2-L4 segments or if it extends anteriorly to the presumptive head and L1 segments (n=5) (Fig. 5.4 E). Later at stage 7, the band of expression in the presumptive region of L2-L4 segments can still be detected, as well as the anterior stripe in the presumptive head segments, which is faint in early stage 7 knockdown embryos (n=6) (Fig. 5.4 F-H). It is impossible to determine whether or not the SAZ expression depends on *Pt-Sox21b-1* because the tissue posterior to the presumptive prosomal segments is lost in these embryos (Fig. 5.4 F-H). At late stage 7, while the anterior stripe of *Pt-Msx1* expression appears to be unaffected, the splitting of the L2-L4 band of expression is perturbed – the band appears to never split into the stripes in the presumptive L2-L4 segments (n=2) (Fig. 5.4 H).

Similarly to the other two genes analysed above, activation of *Pt-Msx1* expression does not seem to require the expression of *Pt-Sox21b-1*. However, splitting of the L2-L4 band of *Pt-Msx1* expression into distinct segmental stripes appears to be dependant on *Pt-Sox21b-1* expression.

5.2.2. Clonal analysis of *Pt-Sox21b-1* knockdown in early stages of segmentation

Knockdown of *Pt-Sox21b-1* expression using pRNAi results in severe pleiotropic effects during early embryogenesis[66]. This loss or incorrect specification of cells poses an obstacle to studying the role of this gene in the regulation of other segmentation genes because any observed effects on their expression could be the indirect result of the effect of *Pt-Sox21b-1* knockdown on these cells. To help address this problem, I used eRNAi to induce *Pt-Sox21b-1* knockdown in small subsets of cells. This excludes any maternal effects of *Pt-Sox21b-1* knockdown, and reduces the number of cells affected, allowing a more precise analysis of the effect of *Pt-Sox21b-1* knockdown on specific expression domains of segmentation genes. Embryos were injected with the same 549 bp dsRNA fragment used for the pRNAi assay[66].

To verify if knockdown of *Pt-Sox21b-1* expression was successful in the obtained eRNAi clones, I performed ISH for *Pt-Sox21b-1* on injected embryos (Fig. 5.5 D-I). Note that injected embryos for which the production of a clone was unsuccessful were used as controls (Fig. 5.5 A-C).

Expression of *Pt-Sox21b-1* was significantly reduced in the cells of all clones obtained (n=12) (Fig. 5.5 D-I). In some embryos, *Pt-Sox21b-1* knockdown appeared to leak outside of the clone area as previously reported for eRNAi with other genes (n=4) (Fig. 5.5 E-G)[80,207].

At stage 5, the area affected by the clone appears to have a lower cell density when compared to the remaining area of the germ disc (n=1) (Fig. S4 A-A"). Interestingly, clones in the presumptive leg segments of stage 7 embryos appeared to cause deformation of the germ band (n=3) (Fig. 5.5 E). This effect was clearer from stage 8.1 onwards (n=7) (Fig. 5.5 F-H)

and appears to be caused by a decrease of cell number in the affected area resulting in a constriction of the germ band in the AP axis, rather than the DV axis (Fig. S4 B-E"). This also seems to affect the expression pattern in more posterior segments, which are outside the affected area, as seen by the distortion of the segmental stripe of expression corresponding to the first opisthosomal segment (Fig. 5.5 F-H). However, this type of deformation of the germ band was not observed when clones where obtained in regions posterior to the presumptive leg segments (n=1) (Fig. 5.5 I).

Overall, knockdown of *Pt-Sox21b-1* by eRNAi was successful but cell loss in the affected areas still needs to be considered when analysing segmentation gene expression in clones obtained through this method.

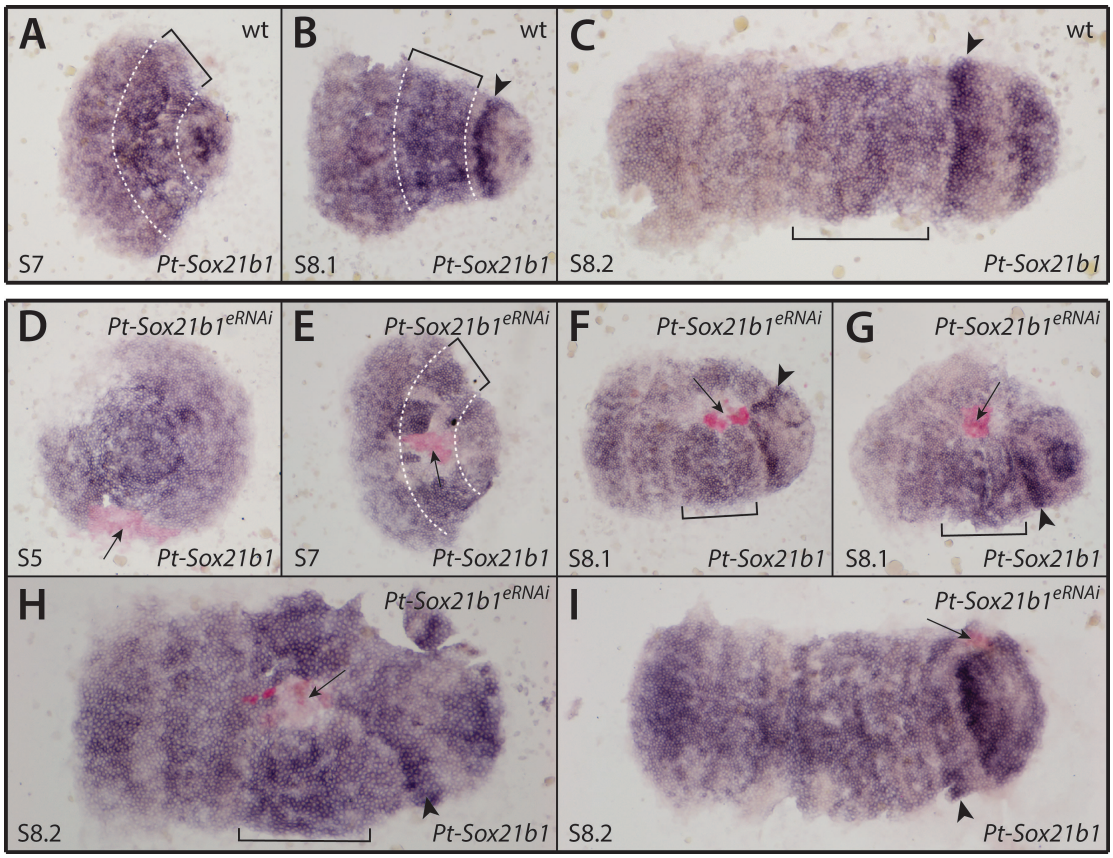

Figure 5.5: Effect of *Pt-Sox21b-1* eRNAi knockdown in early segmentation. (A-C) Expression pattern of *Pt-Sox21b-1* in stage 7 (A), 8.1 (B) and 8.2 (C) wt injected embryos. (D-I) Expression pattern of *Pt-Sox21b-1* in stage 5 (D), 7 (E), 8.1 (F, G) and 8.2 (H, I) *Pt-Sox21b-1* eRNAi embryos. Anterior is to the left in all images except D. Dashed white lines and brackets mark the presumptive L2-L4 segments. Arrowheads: presumptive O1 segment. Arrows: knockdown clones marked by red staining.

5.2.3. Clonal analysis of the effect of *Pt-Sox21b-1* knockdown on *Pt-Dl* expression

Pt-Dl is essential for SAZ formation and is a crucial part of the GRN regulating posterior segment addition[209]. At early stage 5, *Pt-Dl* is expressed in a subset of cells at the rim and in

the centre of the germ disc, in a salt and pepper pattern[209]. By the end of stage 5, the expression at the rim fades away and the posterior salt and pepper domain clears from the centre, where a new solid domain of expression arises (Fig. 5.5 A)[209]. At stage 6, the posterior domain also clears from the centre, with *Pt-Dl* expression being restricted to two adjacent bands of expression, a more anterior salt and pepper band and a more posterior solid band[80,209]. By stage 7, expression reappears in the SAZ and an anterior stripe of expression is formed in the presumptive pedipalpal segment (Fig. 5.5 B)[209]. At stage 8.2, expression is restricted to segmental stripes of varying strength and the SAZ, and the salt and pepper domain is no longer present (Fig. 5.5 C). *Pt-Sox21b-1* knockdown embryos appeared to lack expression of *Pt-Dl* during early stages, although, the authors argued this could be due to cell loss[66]. To verify the effect of *Pt-Sox21b-1* on *Pt-Dl* expression, I carried out ISH for *Pt-Dl* in embryos with *Pt-Sox21b-1* knockdown clones.

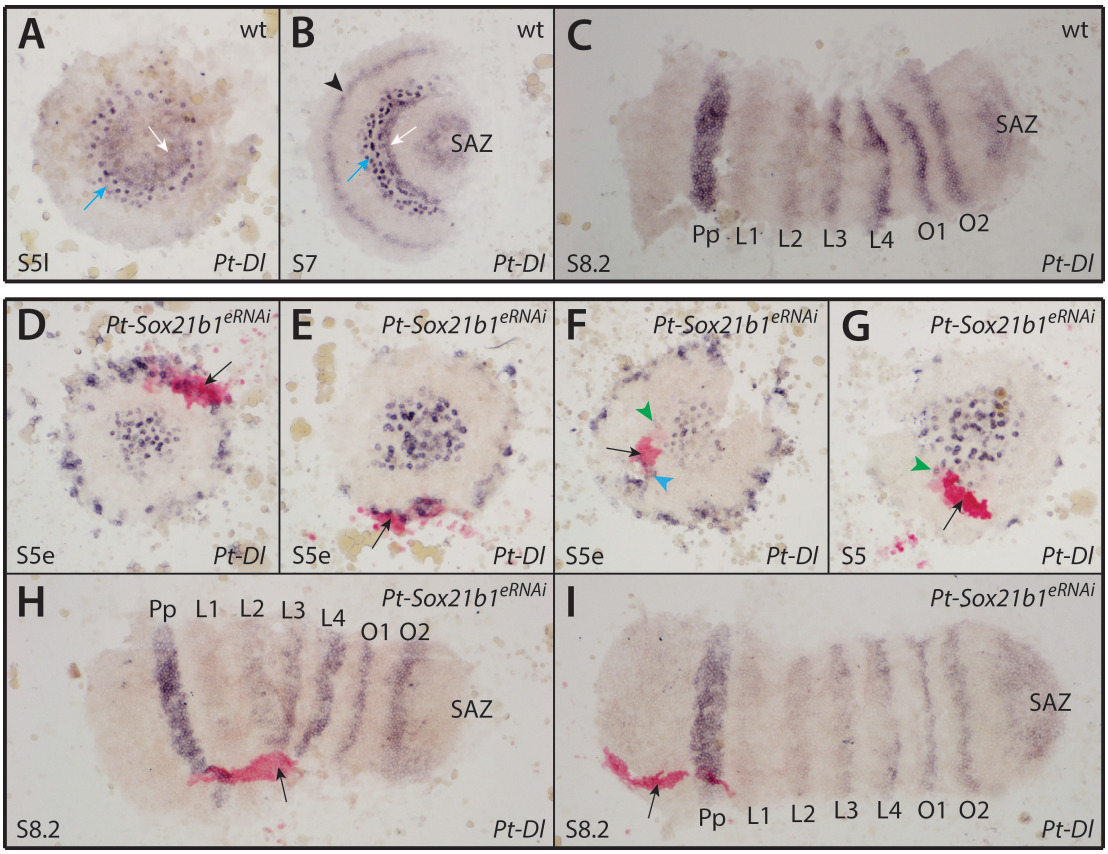

Figure 5.6: Effect of *Pt-Sox21b-1* eRNAi knockdown on *Pt-Dl* expression. (A-C) Expression pattern of *Pt-Dl* in late stage 5 (A), 7 (B) and 8.2 (C, D) wildtype embryos. (D-H) Expression pattern of *Pt-Dl* in early stage 5 (D-F), stage 5 (G) and 8.2 (H, I) *Pt-Sox21b-1* eRNAi embryos. Anterior is to the left in B, C, H and I. Blue arrows: salt and pepper band of expression. White arrows: solid band of expression. Black arrowhead: expression in the presumptive pedipalpal segment. Black arrows: knockdown clones marked by red staining. Green arrowheads: *Pt-Dl* expressing cells in putative knockdown clone. Blue arrowhead: ectopic expression in the presumptive head segments. Pp, presumptive pedipalpal segment; L1, presumptive L1 segment; L2, presumptive L2 segment; L3, presumptive L3 segment; L4, presumptive L4 segment; O1, presumptive O1 segment; O2, presumptive O2 segment; SAZ, segment addition zone.

In stage 5 embryos, the salt and pepper domain in the centre of the germ disc appears to be normal in *Pt-Sox21b-1* knockdown clones (n=2) (Fig. 5.6 F and G), however I was unable to obtain any clones that definitely overlapped with this domain. I was able to obtain several clones that overlapped with the domain at the rim of the germ disc (n=6), which consistently induced an increase of *Pt-Dl* expressing cells in this region (Fig. 5.6 D and E). Interestingly, the presence of a clone in the area between these two domains also caused the ectopic expression of *Pt-Dl* in the presumptive head segments (n=1) (Fig. 5.6 F).

At stage 8.2, *Pt-Sox21b-1* knockdown clones in the head region had no detectable effect on *Pt-Dl* expression (n=2) (Fig. 5.6 I). However, knockdown of *Pt-Sox21b-1* in the region of the leg-bearing segments appears to cause the L2-L4 stripes of *Pt-Dl* expression to coalesce towards the clone region, where expression of *Pt-Dl* seems to be depleted (n=1) (Fig. 5.6 H). The faint stripe of *Pt-Dl* expression in L1 appears to be unaffected (Fig. 5.6 H). In this same embryo, the anterior-most stripe of *Pt-Dl* expression appears to be warped, though this is probably due to the deformation of the germ band (Fig. 5.6 H).

To summarise, my results suggest that the early *Pt-Dl* expression domain in the anterior region of the germ disc expands in the absence of *Pt-Sox21b-1*, and correct formation of L2-L4 stripes of *Pt-Dl* expression at stage 8.2 requires *Pt-Sox21b-1* expression. It remains inconclusive whether or not the posterior domains of *Pt-Dl* expression are regulated by *Pt-Sox21b-1* because I was unable to obtain clones in this region.

5.3. Discussion

5.3.1. Possible role of *Pt-Sox21b-1* in regulating the formation of leg segments

A previous study of *Pt-Sox21b-1* function in *P. tepidariorum* revealed its role in both anterior and posterior segmentation[66]. The loss of leg-bearing segments in *Pt-Sox21b-1* knockdown embryos could be the result of direct regulation of anterior segmentation genes, as hinted at by the observed loss of *Pt-h* expression[66]. However, the effect of *Pt-Sox21b-1* knockdown on the expression of other genes known to regulate anterior segmentation, such as *Pt-Dll*, *Pt-hb* and *Pt-Msx1*, was not investigated[66].

My results suggest that *Pt-Sox21b-1* does not directly activate genes regulating the formation of leg segments, as expression of *Pt-Dll*, *Pt-hb* and *Pt-Msx1* is still present in *Pt-Sox21b-1* knockdown embryos. Therefore, the loss of leg segments observed in Class I and II phenotypes appears not to be the outcome of the loss of anterior segmentation gene expression. However, the expression patterns of these genes seem to be disrupted by the knockdown of *Pt-Sox21b-1*. More specifically, early expression in the presumptive regions of leg segments does not seem to split correctly into distinct segmental stripes. This is best perceived when analysing *Pt-Msx1* expression, where the early band of expression in the L2-L4 region does not separate into stripes at late stage 7 (Fig. 5.4 F-H). A similar effect was described in *Pt-hb* knockdown embryos, where *Pt-h* expression is unable to properly split into segmental stripes in the L2 and L3 segments, which correlates with the loss of the L2 segment in these embryos[208]. It is therefore possible that the loss of the leg segments in *Pt-Sox21b-1* knockdown embryos is the result of the inability of the anterior segmentation genes to split into a segmental stripe pattern. Loss of segments due to mispatterning of the segmentation genes was previously described in *Pt-otd1* knockdown embryos[58], which suggests that, even though the mechanisms behind head and leg segments appear to be different[207,208], the splitting of early broad expression domains of segmentation genes into specific segmental stripes is essential for the formation of both head and leg bearing segments.

The effect of clonal *Pt-Sox21b-1* knockdown on *Pt-Dl* expression in stage 8.2 embryos is consistent with this hypothesis. The normally striped pattern of *Pt-Dl* expression in the L2-L4 segments appears to be fused proximal to the clone (Fig. 5.6 H). Again, this effect is very similar to the previously described results of clonal knockdown of head segmentation genes (e.g. *Pt-Otx1*), which resulted in the fusion of segmental stripes of expression specific to the head segments[207]. Interestingly, the faint L1 stripe of *Pt-Dl* expression appears to be normal in the knockdown clone (Fig. 5.6 H). It is possible that *Pt-Sox21b-1* regulates the expression of other genes or the cell behaviour underlying the splitting of L2-L4 segments, therefore only indirectly affecting the expression of the downstream segmentation genes. This would explain the presence of the L1 segment in Class I *Pt-Sox21b-1* pRNAi knockdown embryos. Loss of the

L1 segment in Class II embryos could be interpreted as the result of excessive cell death in this region extraneous to *Pt-Sox21b-1* function, much like the previously suggested explanation for the loss of the L2 segment in *Pt-Dll* knockdown embryos[56].

Taken together, these results further validate the role of *Pt-Sox21b-1* as a gap-like gene in anterior segmentation, which is evocative of the function of *Dichaete* during segmentation in *D. melanogaster*[210]. In *D. melanogaster*, the SoxB gene *Dichaete*, a closely related gene to *Sox21b* (Fig. 4.1), directly regulates primary pair-rule genes such as *eve* and *h*, acting in parallel with other gap genes[210]. Accordingly, loss of *Dichaete* function disrupts the expression of pair-rule genes but does not affect gap gene expression[210]. My analysis of *Pt-Sox21b-1* function suggests it might be acting in a similar way to *Dichaete* in *D. melanogaster*, with respect to its role in patterning the leg-bearing segments in *P. tepidariorum*. Early expression of the spider gap-like genes *Pt-Dll* and *Pt-hb* appears to be unaffected in *Pt-Sox21b-1* knockdown embryos, while expression of *Pt-h*, which is thought to have a conserved function in spider segmentation[57,206], is lost[66] (Fig. 5.7). Even though the expression patterns of *Pt-Dll* and *Pt-hb* appear to be disturbed at later stages of *Pt-Sox21b-1* knockdown embryos, this could be due to an indirect effect of the loss of leg and posterior segments. A more systematic analysis of *Pt-Sox21b-1* function is necessary to better understand how it interacts with other genes of the anterior segmentation GRN.

5.3.2. *Pt-Sox21b-1* knockdown clones show localized morphological defects

Functional analysis of *Pt-Sox21b-1* revealed it was involved in other developmental processes besides segmentation, such as the transition from radial to axial symmetry and mesendodermal fate specification[66]. Consequently, knockdown of *Pt-Sox21b-1* by pRNAi induces highly pleiotropic effects during early embryogenesis, resulting in cell loss which appears to be caused by an increase in cell death and a decrease in cell proliferation[66]. My results demonstrate that *Pt-Sox21b-1* knockdown by eRNAi leads to similar effects that are more restricted to the region of the clone.

There is an apparent decrease in cell density in the knockdown areas (Fig. S4 A-E"), possibly caused by cell death and/or a decrease in cell proliferation, as previously observed[66]. A possible explanation of the early effects of *Pt-Sox21b-1* knockdown is that it could be required for the maintenance of stem/progenitor cell state of a particular cell lineage, which is a common role of several Sox genes in both arthropods and vertebrates[171,176,178]. The effect of localized knockdown of *Pt-Sox21b-1* is aggravated from stage 7 onwards, sometimes leading to the deformation of the germband (Fig. 5.5 E-H and Fig. 5.6 H). This is probably due to cell loss causing malformation of leg and posterior segments, which start forming at these stages.

Consistent with this is the fact that clones obtained in areas anterior to the L1 segment display a reduced effect, with no obvious deformation of the germband (Fig. 5.6 I).

Overall, my results show that knockdown of *Pt-Sox21b-1* in subsets of cells leads to the disruption of segmentation in the affected area. Nevertheless, these effects are not as severe as those caused by pRNAi knockdown, allowing for the analysis of *Pt-Sox21b-1* function in more specific regions of the embryo.

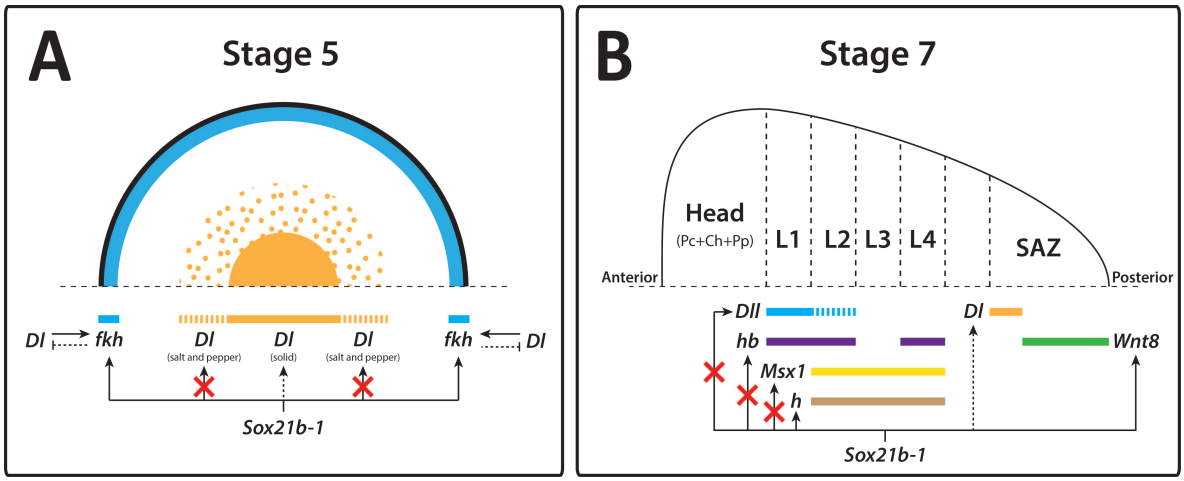

Figure 5.7: Summary of *Pt-Sox21b-1* function in segmentation. (A) At stage 5, *Sox21b-1* and *Dl* regulate *fkh* expression at the rim of the germdisc[66,210]. The presence of a negative feedback loop, with *fkh* inhibiting *Dl* expression at the rim, could explain the increase in *Dl* expressing cells at the rim of the germdisc in *Sox21b-1* knockdown clones. *Sox21b-1* does not appear to regulate the salt and pepper domain of *Dl* expression (red crosses), however, it could still regulate the solid domain at the centre of the germdisc (dashed arrow). (B) At stage 7, *Sox21b-1* regulates *h* and *Wnt8* expression[66]. However, it does not regulate the expression of the spider gap-like genes *Dll* and *hb*[56,209], as well as the putative segmentation gene *Msx1* (red crosses). However, *Pt-Sox21b-1* expression appears to be required for the splitting into segmental stripes of expression of these genes. Additionally, *Sox21b-1* might regulate the *Dl* expression domain in the anterior SAZ, which is thought to be necessary to activate and restrict *Wnt8* expression in the posterior SAZ[80]. Pc, Pre-cheliceral segment; Ch, Cheliceral segment; Pp, pedipalpal segment; L1, L1 segment; L2, L2 segment; L3, L3 segment; L4, L4 segment; O1, O1 segment; O2, O2 segment; SAZ, segment addition zone.

5.3.3. Contrasting effects of *Pt-Sox21b-1* knockdown on distinct domains of *Pt-Dl* expression

One of the effects of *Pt-Sox21b-1* knockdown is the arrest of posterior segmentation through the disruption of SAZ formation, ultimately leading to the loss of all opisthosomal segments[66]. Expression analysis on *Pt-Sox21b-1* knockdown embryos suggests that this is in part due to the loss of *Pt-Dl* and *Pt-Wnt8* expression, two essential genes in the establishment of the SAZ[57,66,209]. However, this could be an indirect result of cell loss or incorrect cell specification due to the earlier functions of *Pt-Sox21b-1*[66]. One such function is the specification of mesendodermal cells, as suggested by the loss of *Pt-forkhead* (*Pt-fkh*) expression in knockdown embryos, an essential gene in the formation of both these layers[66,211]. To try to address

this issue, I analysed the expression pattern of *Pt-Dl* in eRNAi knockdown clones for *Pt-Sox21b-1*.

Contrary to what was previously observed, the salt-and-pepper-like *Pt-Dl* expression domain at the centre of the stage 5 germ disc appears to be unaffected by the knockdown of *Pt-Sox21b-1* expression (Fig. 5.6 G). However, this remains inconclusive because I was unable to obtain a clone that clearly overlapped with this domain. The fact that *Pt-Sox21b-1* appears to be necessary for the proper migration of the cumulus[66], which is an essential step in the transition from germ disc to germband, might explain the lack of knockdown clones in the centre of the germ disc, as any clones obtained in this region could severely disrupt germband formation and kill the embryos earlier. It also remains to be addressed whether *Pt-Sox21b-1* knockdown affects the solid posterior domain of expression that forms at late stage 5, which is thought to be necessary to activate and restrict expression of *Pt-Wnt8* to the posterior SAZ[80] (Fig. 5.7).

Interestingly, knockdown of *Pt-Sox21b-1* at the rim of early stage 5 embryos appears to increase the number of *Pt-Dl* expressing cells in this domain (Fig. 5.6 D-F). Expression of *Pt-Dl* has been previously shown to be necessary for the activation of *Pt-fkh* at the rim of the germ disc[209]. Additionally, it was previously shown that *Pt-fkh* expression is lost in *Pt-Sox21b-1* knockdown embryos[66]. Therefore, it is possible that the observed increase in *Pt-Dl* expressing cells in *Pt-Sox21b-1* knockdown clones at the rim of the germ disc is an indirect effect of the loss of *Pt-fkh* expression, due to a possible feedback loop between these two genes (Fig. 5.7).

Lastly, at stage 8.2, knockdown of *Pt-Sox21b-1* appears to only affect *Pt-Dl* expression domains posterior to the L1 segment (Fig. 5.6 H and I). The stripes of *Pt-Dl* expression in the L2-L4 segments appear to merge together in the region of the knockdown clone, where *Pt-Dl* expression appears lost (Fig. 5.6 H). However, this could still be the indirect result of the loss of these segments, as the germband is clearly deformed in the region of the *Pt-Sox21b-1* knockdown clone (Fig. 5.6 H). Nonetheless, these results are consistent with the hypothesis proposed above that *Pt-Sox21b-1* only regulates the formation of the L2-L4 segments in anterior segmentation, with the loss of the L1 segment in class II embryos being due to excessive cell death.

In conclusion, my results contrast with those previously published[66] because *Pt-Dl* expression generally appears not to be directly affected by *Pt-Sox21b-1* knockdown. Moreover, the effect of *Pt-Sox21b-1* knockdown on *Pt-Dl* expression is dependent on the location of the knockdown clone (e.g. clones at the edge of the germ disc cause an increase of *Pt-Dl* expressing cells in this region). Although, it is still possible that *Pt-Sox21b-1* directly regulates some of the posterior most domains of *Pt-Dl* expression, for which knockdown clones of *Pt-Sox21b-*

1 were not obtained. Despite this, my results demonstrate that the interactions of *Pt-Sox21b-1* with the segmentation GRNs could be more complex than was previously proposed.

6. Chapter IV: Gene duplication, Wnt signalling and spider eye development

6.1. Introduction

The evolution of vision has long fascinated evolutionary and developmental biologists, and theories on its origins can be traced as far back as "The Origin of Species"[212]. In it, Darwin proposed that all eye types had evolved through random variation and selection from a single prototypic eye, composed solely of a light-sensitive cell and a pigment cell[212,213]. It is thought that vision may have its origin in Cyanobacteria, the oldest and simplest known organisms capable of light detection[213]. This allowed for the evolution of light-dark cycles and phototaxis, conferring a selective advantage over other organisms[213]. Subsequent gradual changes in eye complexity allowed for the evolution of increasingly complex behaviours, ultimately leading to the emergence of highly complex visual systems, such as the compound eyes of insects and the camera-type eyes of vertebrates, which are capable of high-resolution vision[213,214].

However, the idea of a single origin of vision is still much debated, with some studies proposing instead multiple origins for the evolution of structures capable of light detection[215,216]. Evidence put forward to support the latter hypothesis include the independent origin of several eye lens crystallins and opsins within the animal lineage, and the independent acquisition of photoreceptors in different eukaryote lineages[215–217].

Studies supporting a single ancestral origin of image forming eyes focused mainly on the genetic basis of eye development[213]. One gene in particular, *Pax6*, was proposed to be the master regulator of eye development, as it was found to be expressed in the eye primordia of most bilaterians studied so far, including vertebrates, insects, molluscs, annelids and planarians[86,218–220]. Furthermore, putative *Pax6* genes of two species of cnidarians were shown to be involved in eye development, and are capable of inducing ectopic eyes when expressed in the imaginal discs of *D. melanogaster*[221,222].

In addition to *Pax6*, other genes were found to have a conserved role in eye development of vertebrates and insects, including members of the SIX, EYA and DAC gene families[213]. Often called the retinal determination gene network[223], this subset of genes is necessary for the establishment of the eye primordia in both vertebrates and insects, and their expression appears to be regulated by conserved signalling pathways, including the Hedgehog, BMP and Wnt signalling pathways[70,224].

However, this could still be the outcome of convergent evolution through co-option of the same regulatory pathways and retinal determination genes[216]. Nonetheless, eye develop-

ment appears to be under control of similar genes (e.g. *Pax6*) and regulatory pathways (e.g. Wnt signalling pathway) across metazoa[213].

6.1.1. Role of retinal determination genes in animal eye development

Studies on animal eye development have consistently found a similar subset of genes to be involved in determining eye fate[213]. The core set these retinal determination genes includes members of the PAX6, SIX, EYA and DAC gene families[223]. In insects and vertebrates, for which eye development has been more widely studied, PAX6 genes appear to be at the top of this regulatory gene network and are usually required for the activation of downstream retinal determination genes[218,223].

The regulation of eye development in insects is best understood in *D. melanogaster*. Two distinct pathways appear to specify each of the two eye types found in *D. melanogaster*: compound eyes and ocelli[213]. During embryogenesis, *twin of eyeless* (*toy*) and *ey*, two PAX6 genes, specify the eye-antennal disc[225]. During larval development, *toy* and *ey* expression becomes restricted to the eye region of the disc[226], activating *so*, *eyes absent* (*eya*) and *dachshund* (*dac*) expression in the eye field (i.e. compound eye primordia) (Fig. 6.1)[227,228]. Expression of *so*, *eya* and *dac* is also activated by *dpp*, which is expressed at the posterior of the eye disc and in the morphogenetic furrow as differentiation proceeds from posterior to anterior (Fig. 6.1)[69]. Expression of *Optix*, a member of the SIX3 gene family, is also activated by *toy*, where, jointly with *hh*, it controls the progression of the morphogenetic furrow by regulating *dpp* expression in this domain (Fig. 6.1)[229]. Two other PAX6 genes, *eyegone* (*eyg*) and *twin of eyegone* (*toe*), are necessary for the correct development of the compound eye[230], although, they appear to act somewhat independently of the other members of the retinal determination network[213].

The ocelli primordia are first visible as two groups of cells expressing *so* and *eya* in the anterior dorsal region of the eye section of the disc (Fig. 6.1)[231]. Activation of these two retinal determination genes is dependent of *toy* and *orthodenticle* (*otd*) expression in this region[231]. Therefore, compound eye and ocelli development appear to be controlled by different pathways albeit requiring a similar subset of genes (Fig. 6.1 B). Loss of function of any of these retinal determination genes leads to the loss of compound eyes, and loss of *toy*, *eya* or *so* leads to the loss of ocelli (Fig. 6.1 B)[85,232]. Together, the compound eye and ocelli pathways activate the expression of other downstream eye development genes, such as the transcription factor *atonal* (*ato*), which has an important role in photoreceptor cell differentiation (Fig. 6.1 B)[233,234].

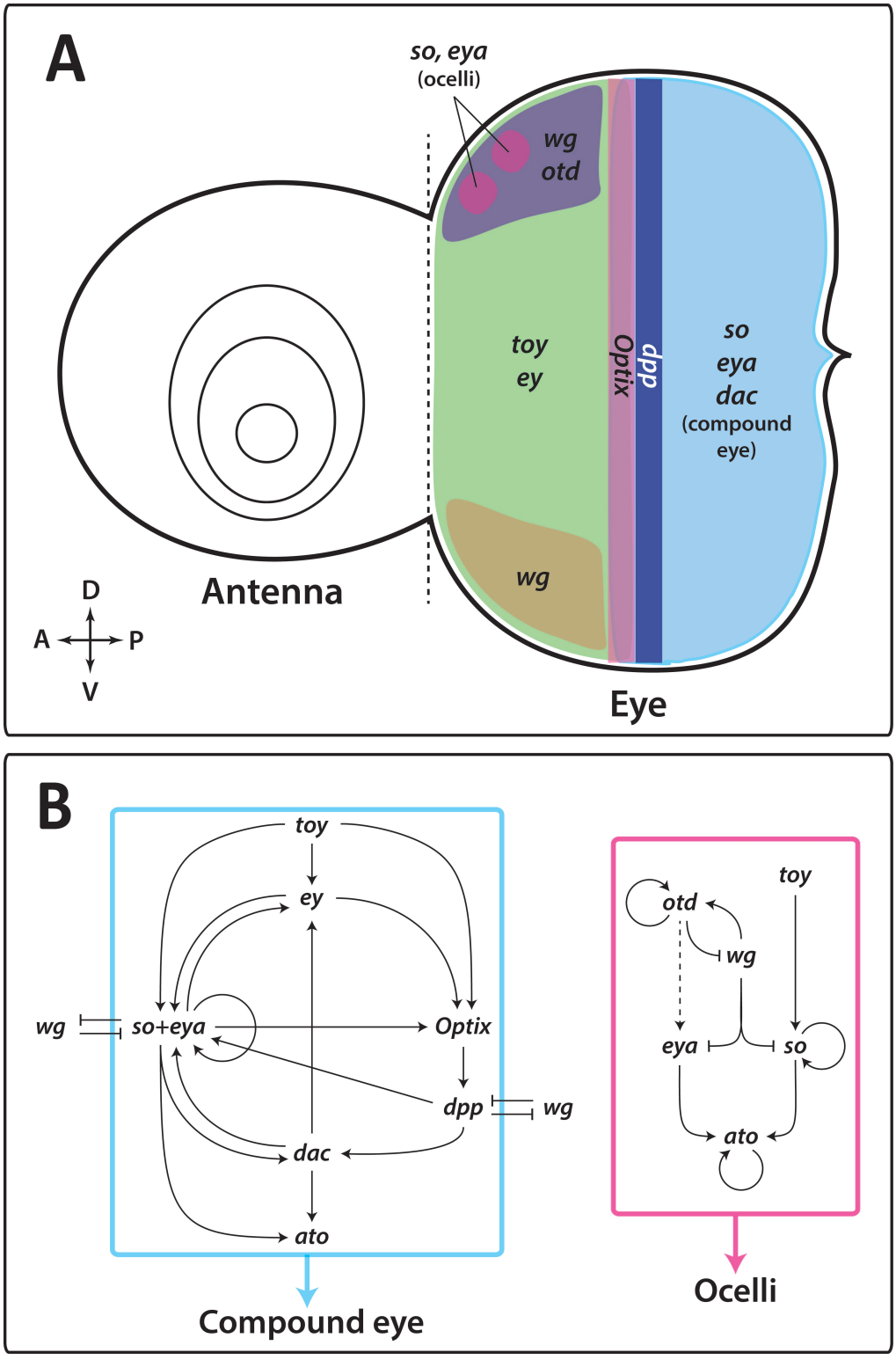

Figure 6.1: Expression and gene regulatory network during eye development in *D. melanogaster*. (A) Expression patterns of *wg*, *otd*, *dpp* and retinal determination genes in the eye-antennal disc during larval development. Dashed line represents the boundary between the antennal and eye regions of the disc. Each unique expression domain is represented by a different colour. The expression domains of all retinal determination genes overlap in the *Optix* domain. Antero-posterior and dorso-ventral axes are represented in the lower left corner. (B) Summary of known interactions between members of the gene regulatory networks regulating compound eye (light blue) and ocelli (pink) development. Dashed arrow represent indirect activation.

Orthologs of retinal determination genes in vertebrates have similar roles in the regulation of eye development. *Otx2* expression in the anterior neural plate is required for the activation of early retinal determination genes, which include *Pax6* and *Six3*[235]. In mice, loss of *Otx2* function results in the loss of anterior structures, including the lens placodes[236], and loss of *Pax6* and *Six3* function leads to the loss of eyes in mice and medaka respectively[237,238]. In *Xenopus*, *Pax6* activates expression of *Six6*, a paralog of *Six3*, in the eye field[235], and overexpression of this gene causes expansion of the retina[239]. The role of vertebrate EYA and DAC genes in eye development are still poorly understood. In mice, *Eya1* and *Eya2* are expressed in the lens placode and require *Pax6* expression to be expressed in this tissue[169]. Lastly, *Dach1* is expressed in the developing eye of both chick and mouse, although, it does not require *Pax6* expression[240] and mice mutant for *Dach1* do not show any obvious malformation in the eye[241].

In summary, eye development in both insects and vertebrates requires a similar set of retinal determination genes in eye development. Incorrect expression of any of these genes often leads to eye defects[83,85], thus, their expression needs to be tightly regulated. Interestingly, the signalling pathways regulating retinal determination genes appear to also be conserved between insects and vertebrates, which include the Hedgehog, BMP and Wnt signalling pathways[70,224].

6.1.2. Role of Wnt signalling in animal eye development

The Wnt genes are a family of signalling molecules with key roles in many developmental processes, from cell proliferation to pattern formation[242,243]. Wnt ligands bind to cell membrane receptors, usually of the Frizzled family, activating the Wnt signalling pathway to regulate the transcription of target genes[242]. Signal transduction upon binding of Wnt ligands to a Fzd receptor is achieved through one of three signalling pathways: the canonical β-catenin dependent pathway, and the non-canonical planar cell polarity (PCP) and Ca2+ dependant signalling pathways243.

In the β-catenin signalling pathway, binding of a Wnt ligand promotes the accumulation of β-catenin by inhibiting the formation of its degradation complex, allowing the translocation of β-catenin to the nucleus and inducing the expression of Wnt-responsive genes243. In addition to the Fzd receptor, induction of PCP signaling requires the presence of other transmembrane receptors (Vangl and Flamingo) to activate cytoplasmic effector molecules such as Rho GTPases (RhoA) and c-Jun N-terminal kinase (JNK), which in turn activate gene expression and promote cytoskeletal rearrangements243. Lastly, in the Wnt/Ca2+ pathway, trimeric G-proteins associated to the Fzd receptor activate phospholipase C, promoting calcium re-

lease through the action of inositol triphosphate (IP3)243. The increase in calcium concentration leads to the activation of downstream transcription factors such as NFAT (nuclear factor associated with T cells) and CREB, which upregulate the expression of several genes243.

Among many other essential roles in animal development, Wnt signalling is crucial in the regulation of eye development of both insects and vertebrates[69,70]. In *D. melanogaster*, the Wnt gene *wingless* has a critical role in confining the expression of retinal determination genes to the eye field in the posterior of the eye-antennal disc, helping to differentiate between dorsal head tissue and compound eye fate[244,245]. Ectopic activation of the *wg* pathway in the eye field represses *eya*, *so* and *dac* expression, resulting in the formation of dorsal head tissue instead of ommatidia[244]. Furthermore, loss of *wg* signalling in the region destined to form the dorsal head capsule leads to the ectopic expression of *eya* and *dac*[244], resulting in the formation of an ectopic compound eye[246]. The Wg signalling pathway is also involved in the regulation of ocelli development in *D. melanogaster*[231]. Wg signalling directly activates *otd* expression in the dorsal region of the eye-antennal disc, which in turn leads to the activation of *eya* expression in the ocelli primordia[231]. Conversely, Wg signalling also inhibits *eya* and *so* expression in the region around the ocelli primordia[231].

A similar scenario was found in vertebrates, with Wnt signalling being essential to the establishment and maintenance of the eye field[70,247]. At the gastrula stage, Wnt signalling restricts the caudal expansion of lens/olfactory placodal cell fate in the rostral neural plate border, specifying neural crest cell fate instead[248]. In mice, activation of canonical Wnt\β-catenin signalling in the eye field inhibits lens formation and represses *Pax6* expression[249], while the loss of Wnt\β-catenin signalling in the extraocular ectoderm leads to ectopic lens formation in small areas that show ectopic expression of *Pax6*[250]. The retinal determination gene *Six3* has been shown to inhibit Wnt\β-catenin signalling in the anterior neural plate, with loss of *Six3* resulting in the ectopic activation of Wnt signalling and loss of rostral forebrain structures, which include the eye field[251].

Several studies have also found evidence for the role of non-canonical Wnt\PCP signalling in promoting the formation of the eye field and as a mediator of morphogenetic movements[70,252]. For example, loss of *Wnt4* function in *Xenopus*, a modulator of non-canonical Wnt signalling, results in the depletion of *Pax6* expression in the eye field and consequent loss of eye structures[253]. Even though the role of Wnt signalling in eye organogenesis appears to be conserved in vertebrates, the individual components involved in this process, such as Wnts and Frizzled receptors, vary in their precise roles among vertebrate species[70].

6.1.3. Spider eye development

Spiders have some of the most diverse visual systems in the animal kingdom[73]. They differ substantially in number, size, arrangement and function of their eyes (Fig. 6.2), resulting in highly specialized vision that is usually tightly related to behavioural necessity[73]. Thus, spider eyes represent a useful model to study the genetic and developmental bases of variation in these important sensory organs.

Most spiders have a total of four pairs of eyes, each being named in line with their relative position on the head: anterior median eyes (AM), anterior lateral eyes (AL), posterior median eyes (PM) and posterior lateral eyes (PL)[73] (Fig. 6.2). According to fundamental differences in their morphology and function, spider eyes can be divided in two types: the principal or median eyes, comprised of the AM eyes, and the secondary or lateral eyes, comprised of the AL, PM and PL eyes[73]. This division into median and lateral eyes is thought to be a conserved trait across Arthropoda, with the AM eyes of spiders being homologous to insect ocelli and the AL, PM and PL eyes of spiders being homologous to insect compound eyes[73].

To date, our knowledge about the genetic basis of spider eye development has been limited to expression analyses of retinal determination gene orthologs in two species, *P. tepidariorum* and *Cupiennius salei*[67,68]. Interestingly, all members of the retinal determination gene network, excluding only *eya*, were found to be duplicated in both species suggesting that the arachnopulmonata WGD event might have had an impact on the evolution of arachnid eye development[67,68]. Most retinal determination gene orthologs analysed are expressed in at least one pair of eyes, but each eye pair appears to express a unique combination of retinal determination genes, with the only exception being the PM and PL eyes of *P. tepidariorum*, which have the same combination of gene expression[67,68]. Interestingly, there are also differences in which retinal determination gene orthologs are expressed in which eyes between *P. tepidariorum* and *C. salei*, possibly reflecting developmental system drift, or differences in the morphology or function of each eye pair between these two species[67,68].

Despite the apparent universal role of *Pax6* in animal eye development, there is still a lack of clear evidence for a conserved role of *Pax6* in chelicerate eye development[73]. While one of the *Pax6* orthologs of *C. salei* is expressed in the AM eyes, neither *P. tepidariorum* paralog is expressed in the developing eyes[67,68], although, additional copies of *Pax6* were recently found in the genome of this spider, which have not been characterized to date[72].

In addition to the retinal determination gene orthologs, expression of *ato* orthologs was also analysed for *C. salei*[68]. However, neither of the two copies of *ato* identified in this species are expressed in the developing eyes[68]. This is inconsistent with the hypothesis that includes *ato* in an ancestral gene network regulating photoreceptor differentiation in both arthropod eye types[73,234].

Analysis of retinal determination gene expression also helped visualise the embryonic origin and development of each eye type in *P. tepidariorum*, for example *Pt-so1* which is expressed in all eye primordia throughout embryogenesis[67] (Fig. 6.2 B-E). The median and lateral eye primordia originate from separate groups of cells in the non-neurogenic ectoderm of stage 10.2 embryos, which is located at the rim of the developing head (Fig. 6.2 B)[67]. The median eye primordia arise just above the anterior furrow, migrating towards the centre of the pre-cheliceral region in subsequent stages, positioned at the shifting edge of the overgrowing non-neurogenic tissue (Fig. 6.2 C)[67]. The lateral eye primordia are first specified as two groups of cells located at the lateral edges of the pre-cheliceral region (Fig. 6.2 B), which then split at stage 12 into three separate groups of cells corresponding to each individual lateral eye (Fig. 6.2 C)[67].

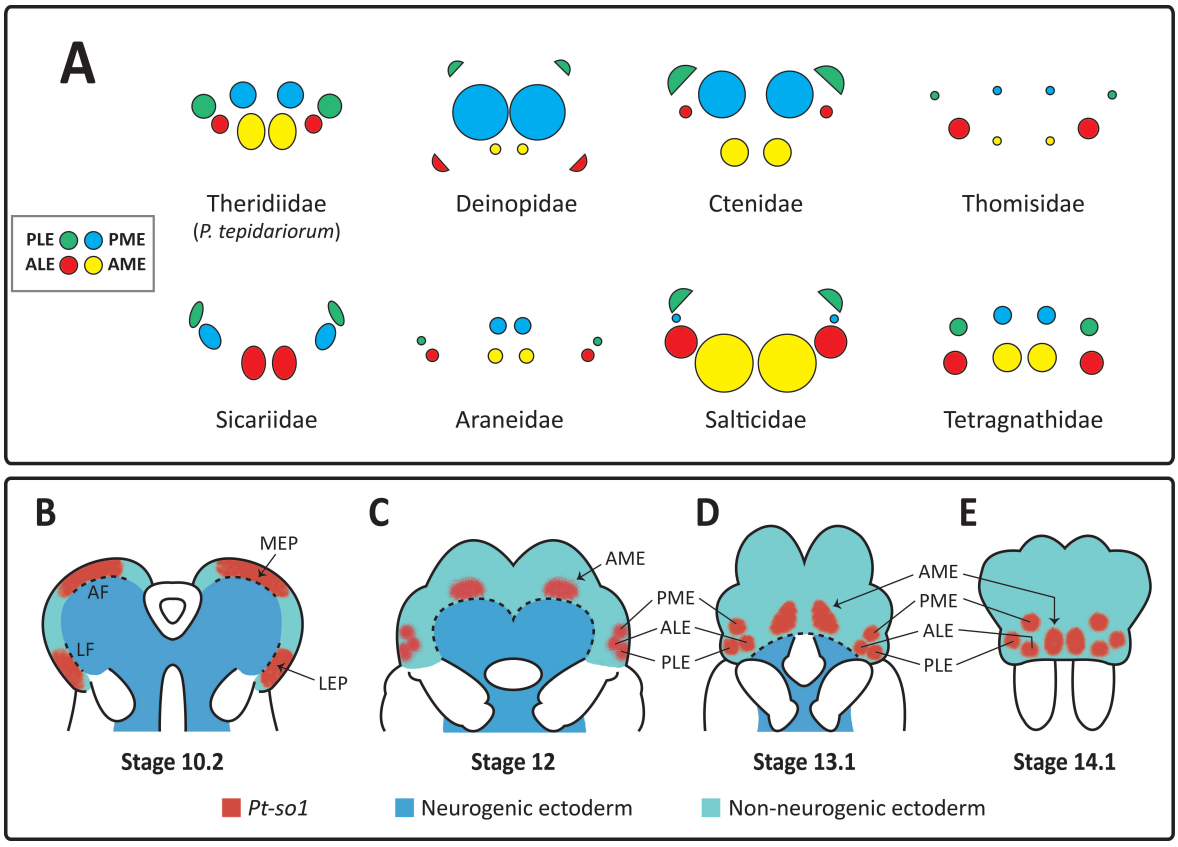

Figure 6.2: Variation in spider eye morphology and dynamics of eye development in *P. tepidariorum*. (A) Diagram displaying the highly diverse eye arrangements of eight different spider families. Adapted from [73]. (B-E) Dynamics of eye development as followed by *Pt-so1* expression throughout embryogenesis in *P. tepidariorum*. At stage 10.2 (B), *Pt-so1* expression is restricted to the regions of non-neurogenic ectoderm adjacent to anterior (AF) and lateral (LF) furrows (dashed lines), establishing the median (MEP) and lateral (LEP) eye primordia. At stage 12 (C), *Pt-so1* in the MEP becomes confined to a smaller group of cells and travels with the overgrowing edge of the non-neurogenic ectoderm (dashed line) towards its final position. *Pt-so1* expression in the LEP splits into three discernible domains corresponding to one of each lateral eye. In the following stages (D), these domains continue migrating towards their final positions in the developing head of *P. tepidariorum* (E). Adapted from [67]. AME, anterior median eye; ALE, anterior lateral eye; PME, posterior median eye; PLE, posterior lateral eye.

As mentioned above, Wnt signalling plays an essential role in the regulation of retinal determination genes of both insects and vertebrates[69,70]. Five Wnt genes were previously shown to be expressed in the developing head of *P. tepidariorum*: *Pt-Wnt2*, *Pt-Wnt5*, *Pt-Wnt7.2*, *Pt-Wnt8* and *Pt-Wnt16*[71], however, their expression relative to the establishment of the eye primordia has not been investigated. As key regulators of animal development, it is plausible that these genes could be directly involved in the patterning of several head structures, possibly playing a role in spider eye development similar to that of *wg* in *D. melanogaster*[69].

Work on the genetic basis of spider eye development is clearly lacking when compared to other well-studied organisms like *D. melanogaster*[73]. To further our knowledge on the gene network regulating eye development in *P. tepidariorum*, I characterized the expression of the two additional copies of *Pax6* and identified and characterized the expression of *ato* orthologs in this species. Furthermore, I analysed the expression of Wnt genes in the developing head of *P. tepidariorum* in relation to *Pt-so1* expression, to investigate the putative role of Wnt signalling in the regulation of spider eye development.

6.2. Results

6.2.1. Expression patterns of *Pt-Toy* and *Pt-Eyg*

Pax6 gene expression has been previously described in two spiders, *P. tepidariorum* and *C. salei*, but only one *Pax6* ortholog of the latter spider was found to be expressed in the developing eyes[67,68]. However, a recent survey of homeobox genes in arachnids identified two additional members of this gene family in *P. tepidariorum*, which were named *Pt-Toy* and *Pt-Eyg* due to their sequence similarity to the respective *D. melanogaster* orthologs[72]. Therefore, considering the important role of this gene family in animal eye development[86], I characterised the expression patterns of these two duplicates during eye development in *P. tepidariorum*.

At stage 10.1, *Pt-Toy* is expressed in paired clusters of cells along the neuroectoderm, a pattern that is maintained throughout embryogenesis (Fig. 6.3 A', B' and C'), as well as in a complex pattern in the developing head (Fig. 6.3 A). The latter appears to be restricted to the developing brain, as cells expressing *Pt-Toy* seem to be covered by the advancing non-neurogenic ectoderm during later stages (Fig. 6.3 B and C). Thus, *Pt-Toy* is not expressed in the developing eyes because the eye primordia are formed from non-neurogenic ectoderm[67].

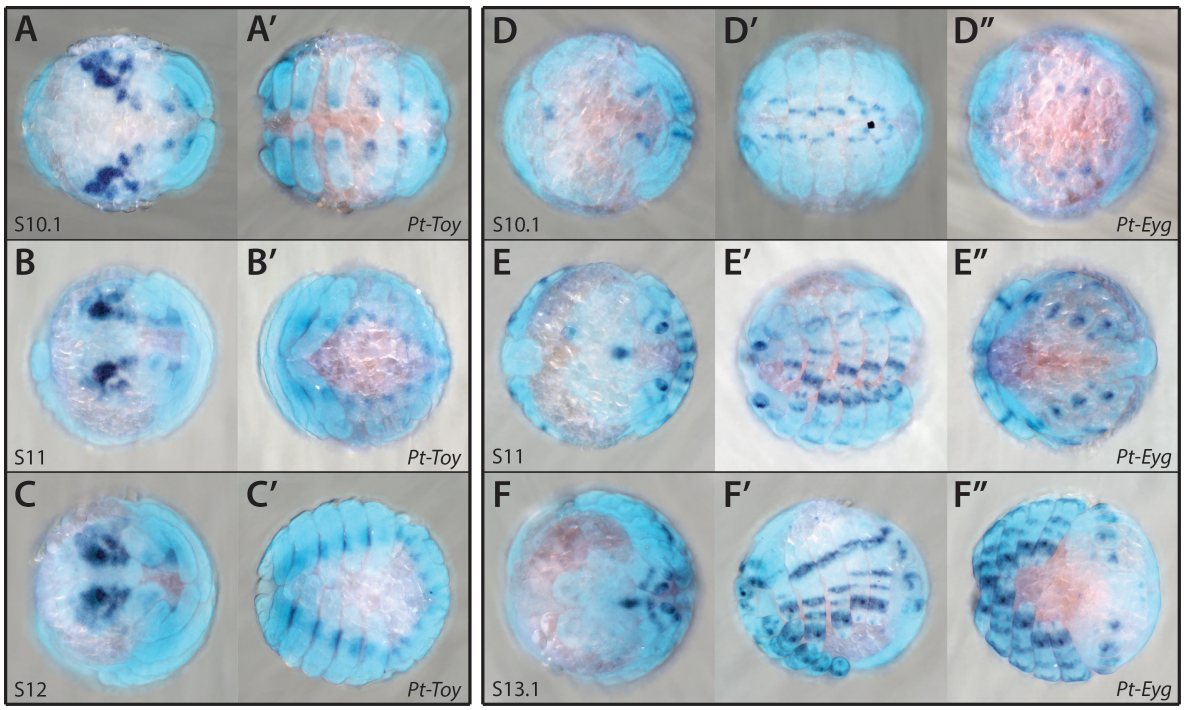

Figure 6.3: Expression pattern of *Pt-Toy* and *Pt-Eyg* genes. (A-C') Expression pattern of *Pt-Toy* in stage 10.1 (A, A'), 11 (B, B') and 12 (C, C') embryos. (D-F") Expression pattern of *Pt-Eyg* in stage 10.1 (D-D"), 11 (E-E") and 13.1 (F-F") embryos. Anterior is to the left in all images.

Pt-Eyg is mainly expressed in the labrum, prosomal appendages and opisthosomal organs (Fig. 6.3 D-F"). Expression is first observed at stage 10.1 in the tips of the chelicerae and as a single ring of expression in the pedipalps and walking legs (Fig. 6.3 D and D'), with faint expression already present in the labrum and opisthosomal organs (Fig. 6.3 D and D"). At stage 11, a ring of expression is visible in the middle of the chelicerae, two additional fainter rings are visible in the pedipalps and three additional rings form in the walking legs (Fig. 6.3 E and E'). A small domain of expression is also present at the tips of the pedipalps and walking legs and expression in the opisthosomal organs becomes stronger (Fig. 6.3 E' and E"). An additional expression domain was detected in the growing dorsal tissue of opisthosomal segments (Fig. 6.3 E"). By stage 13.1, expression appears to be stronger in all the domains (Fig. 6.3 F-F").

In summary, *Pt-Toy* and *Pt-Eyg* have diverse expression patterns, with *Pt-Toy* expression in the nervous system and *Pt-Eyg* being expressed in the developing appendages. However, I did not detect expression of either gene in the eye primordia, suggesting that eye development in *P. tepidariorum* is independent of *Pax6* expression.

6.2.2. Expression pattern of *atonal* genes in *P. tepidariorum*

In *D.* melanogaster, *ato* plays a crucial role in the differentiation of photoreceptors in the compound eye and is necessary for ocelli formation[234]. Expression of *ato* orthologs during spider eye development was previously studied only in *C. salei*, but neither of the two orthologs were found to be expressed in eye primordia[68]. To verify whether this is also true in *P. tepidariorum*, I aimed to analyse the expression patterns of *ato* orthologs in this spider. Through a BLAST search using the sequences of the *atonal* orthologs of *C. salei*, I was able to identify two copies of *ato* in the genome of *P. tepidariorum*, *Pt-ato1* and *Pt-ato2*. I then characterized their expression patterns during eye development by means of ISH.

Expression of *Pt-ato1* is first visible at stage 10.1 in the pre-cheliceral region in four small groups of cells that appear to be similar in position to the developing eye primordia (Fig. 6.4 A). This expression becomes stronger at stage 10.2, and two additional domains in the developing brain appear (Fig. 6.4 C). At stage 12, the expression domain in the median eye primordia moves towards the centre and the expression domain in the lateral eye primordia subdivides into the expected three pairs of lateral eyes (Fig. 6.4 D). This pattern is maintained up to stage 13.2 and appears to be restricted to smaller groups of cells in each developing eye (Fig. 6.4 E and F). *Pt-ato1* is also expressed in several clusters of cells in the developing prosomal appendages at stage 10.1, with an increasingly complex pattern in subsequent stages that is likely associated with the developing peripheral nervous system (Fig. 6.4 A-F).

Pt-ato2 is expressed in two small groups of cells close in the pre-cheliceral region, although these appear to be located in the neuroectoderm (Fig. 6.4 G). Both domains of expression are still present at stage 12 in the middle of the developing brain lobes (Fig. 6.4 H and I). *Pt-ato2* is also expressed in the prosomal appendages, in a similar pattern to that of *Pt-ato1*, though it seems to be expressed in fewer cells (Fig. 6.4 G-I). An additional expression domain can be seen in the forming limb buds of the second opisthosomal segment, perhaps in the booklung primordia (Fig. 6.4 G).

In conclusion, I found one *ato* ortholog to be expressed in the eye primordia of *P. tepidariorum*, suggesting a role of this gene in eye development. This is contrary to what was observed in *C. salei*, where neither of the two *ato* orthologs appears to be involved during eye development[68].

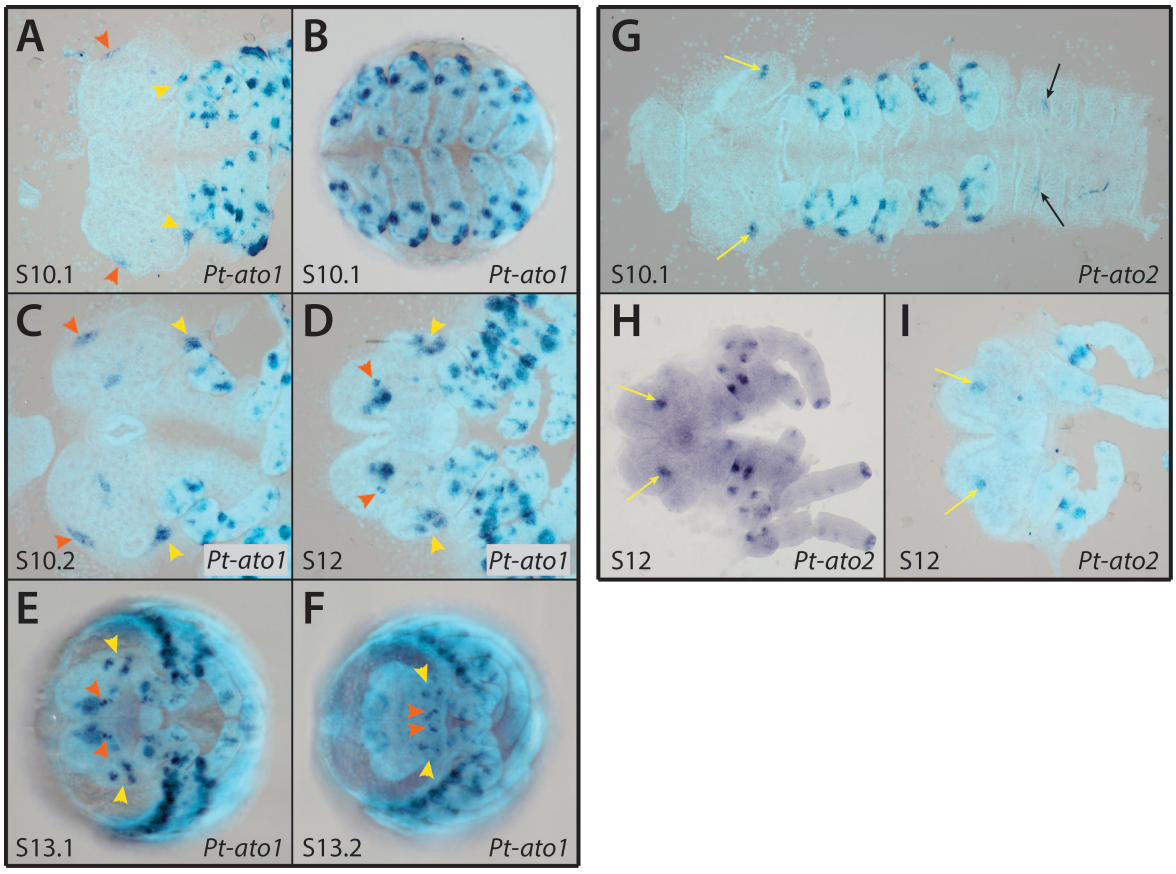

Figure 6.4: Expression patterns of *P. tepidariorum atonal* genes. (A-F) Expression pattern of *Pt-ato1* in stage 10.1 (A and B), 10.2 (C), 11 (D), 12 (E) and 13.2 (F) embryos. (G-I) Expression pattern of *Pt-ato2* in stage 10.1 (G) and 12 (H and I) embryos. Anterior is to the left in all images. In images A, C, D, H and I, only the head region is shown. Embryos in A, C, D and G-I are flatmounted. Arrowheads indicate the median (orange) and lateral (yellow) eye primordia. Arrows indicate the expression domains in the pre-cheliceral region (yellow) and in the second opisthosomal segment (black).

6.2.3. Wnt gene expression in the developing head of *P. tepidariorum*

To verify whether Wnt signalling could be regulating eye development in spiders, I analysed the expression patterns of Wnt genes in the developing head of *P. tepidariorum*, using *Pt-so1* expression as a marker for the developing eye primordia. I focused my analysis on stage 10.2, when *Pt-so1* expression is first visible in the eye primordia, and stage 12, when the expression of *Pt-so1* in the lateral eye primordia separates into 3 distinct pairs of lateral eyes.

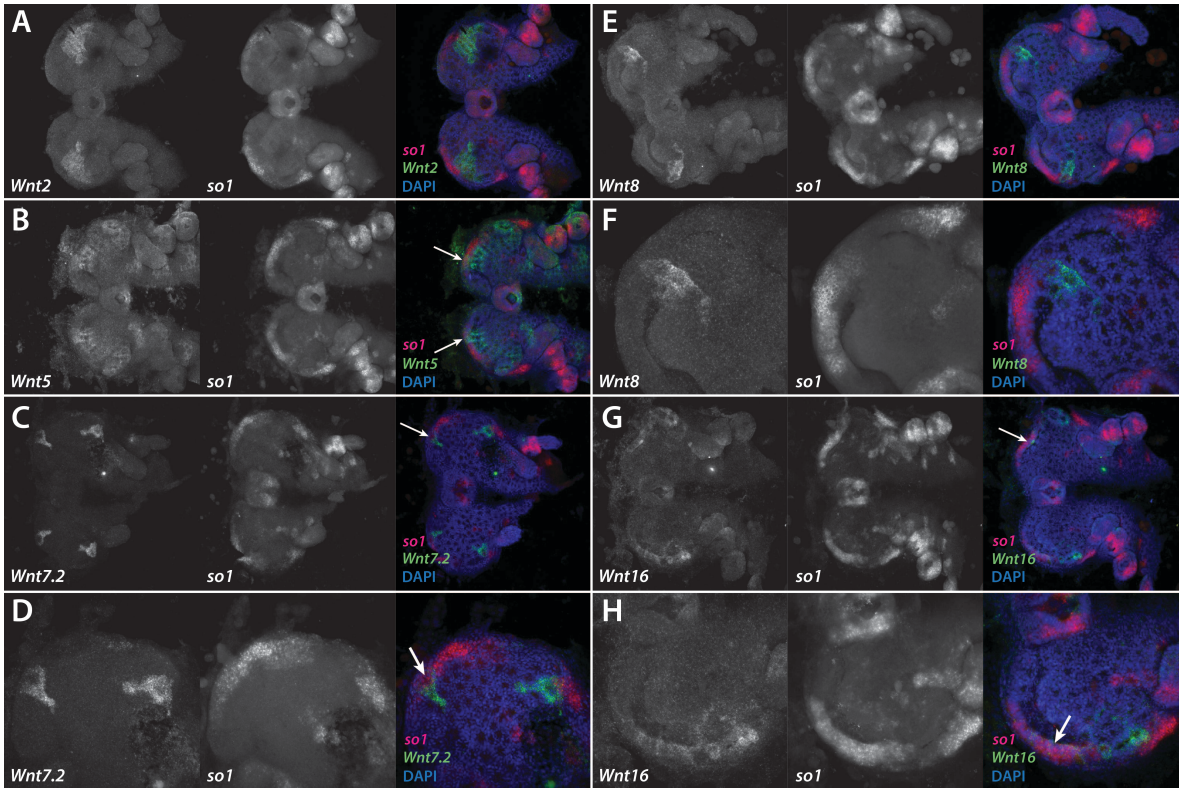

Figure 6.5: Wnt and *Pt-so1* expression in the developing head of stage 10.2 *P. tepidariorum* embryos. Expression of *Pt-Wnt2* (A), *Pt-Wnt5* (B), *Pt-Wnt7.2* (C, D), *Pt-Wnt8* (E, F) and *Pt-Wnt16* (G, H) in the developing head of stage 10.2 embryos. All embryos were double stained for individual Wnts and *Pt-so1*. Panels D, F and H are higher magnification images of panels C, E and G respectively, focusing on one head lobe. Anterior is to the left in all images. White arrows indicate the regions where overlap of Wnt gene expression with *Pt-so1* expression was detected.

At stage 10.2, *Pt-Wnt2* is expressed in two large clusters of cells extending from the centre up to the lateral rim of each head lobe (Fig. 6.5 A). This expression domain does not seem to extend into the non-neurogenic ectoderm and is posteriorly adjacent to *Pt-so1* expression in the median eye primordia (Fig. 6.5 A). At stage 12, *Pt-Wnt2* expression now appears to surround *Pt-so1* expression in the median eye primordia, which by this stage has migrated towards a more central location in the head (Fig. 6.6 A). Furthermore, at this stage *Pt-Wnt2* expression appears to extend to a region just above the PM eye primordia (Fig. 6.6 A).

Pt-Wnt5 is expressed in two large domains in each head lobe of stage 10.2 embryos, which extend into the non-neurogenic region between the two domains of *Pt-so1* expression in both eye primordia (Fig. 6.5 B). *Pt-Wnt5* expression appears to surround both median and lateral eye primordia, but partially overlaps with the anterior-most region of the *Pt-so1* expression domain in the median eye primordia (Fig. 6.5 B). Intriguingly, this particular domain of *Pt-so1* expression does becomes restricted to a smaller group of cells, which suggests that *Pt-so1* expression is probably lost where it overlaps *Pt-Wnt5* expression. At stage 12, *Pt-Wnt5* expression appears to be restricted to two domains, one above the median eye primordia and another in a region below the median and ventral to the lateral eye primordia (Fig. 6.6 B). The anterior-most domain of *Pt-Wnt5* expression appears to partially overlap *Pt-Wnt2* expression in this region (Fig. 6.6 A and B).

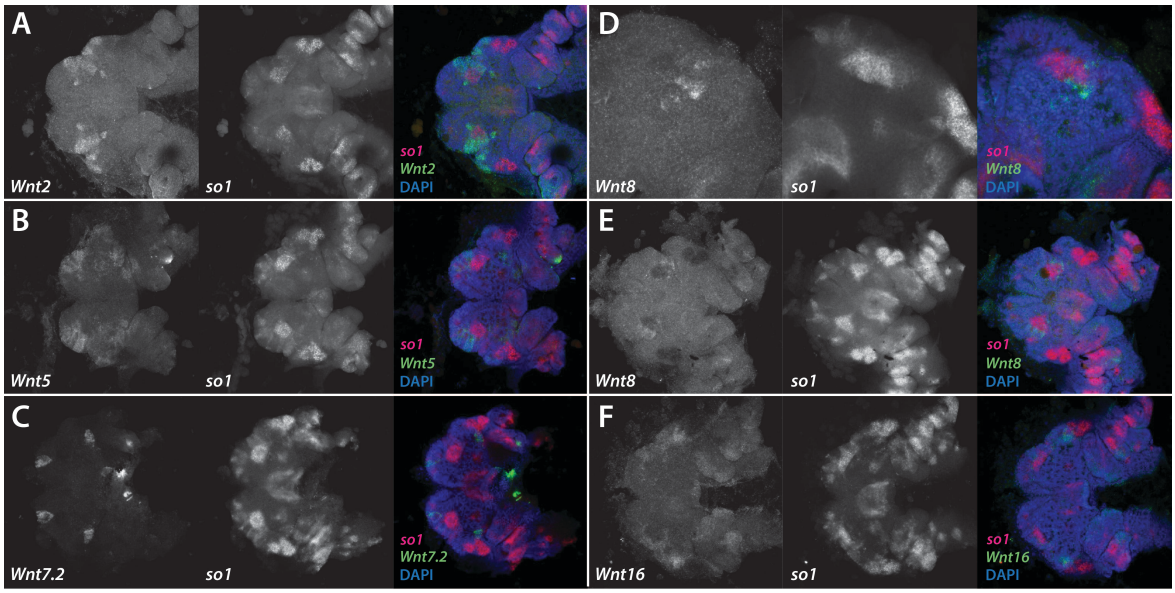

Figure 6.6: Wnt and *Pt-so1* expression in the developing head of stage 12 *P. tepidariorum* embryos. Expression of *Pt-Wnt2* (A), *Pt-Wnt5* (B), *Pt-Wnt7.2* (C), *Pt-Wnt8* (D, E) and *Pt-Wnt16* (F) in the developing head of stage 12 embryos. All embryos were co-stained for *Pt-so1*. Panel D represents a higher magnification image, focusing on one head lobe. Anterior is to the left in all images.

Pt-Wnt7.2 expression is restricted to four small clusters of cells adjacent to the *Pt-so1* expression domains in the eye primordia at stage 10.2 (Fig. 6.5 C). At a higher magnification, it appears that *Pt-Wnt7.2* expression slightly overlaps with *Pt-so1* expression in the same area where *Pt-Wnt5* expression overlaps the median eye primordia (Fig. 6.5 D). At stage 12, *Pt-Wnt7.2* expression is still restricted to two small domains on each head lobe, one at the anterior rim of the head above the median eye primordia and another adjacent to the AL primordia (Fig. 6.6 C). The former appears to be immediately adjacent to the *Pt-Wnt5* expression domain at the rim of the pre-cheliceral region (Fig. 6.6 B and C).

At stage 10.2, *Pt-Wnt8* is expressed in a small group of cells in each head lobe immediately adjacent to the outer-most border of *Pt-so1* expression domain in the median eye primordia (Fig. 6.5 E). This domain does not appear to overlap *Pt-so1* expression (Fig. 6.5 F). Expression of *Pt-Wnt8* becomes fainter at stage 12 and only partially surrounds *Pt-so1* expression in the median eye primordia (Fig. 6.6 D and E).

Lastly, *Pt-Wnt16* expression appears to be restricted to the lateral rim of the developing head at stage 10.2 (Fig. 6.5 G). This domain of expression appears to be immediately adjacent to the lateral eye primordia, though it seems to extend into the *Pt-so1* expression domain in the median eye primordia (Fig. 6.5 G and H). However, *Pt-Wnt16* expression does not completely overlap with *Pt-so1* expression and appears to be expressed at lower levels in the region where their expression overlaps (Fig. 6.5 H). *Pt-Wnt16* expression at stage 12 appears to be restricted to the same area of the developing head, although, it is no longer overlapping *Pt-so1* expression in the median eye primordia (Fig. 6.6 F).

To summarise, Wnt gene expression in the developing head of *P. tepidariorum* appears to be almost entirely extrinsic to *Pt-so1* expression in the eye primordia, although, there is some overlap in a few cells during the early phase of eye primordia establishment in the median eye expression domain (Fig. 6.5). Furthermore, each Wnt gene analysed is generally expressed in different cells around *Pt-so1* expression in the eye primordia, with each eye type being surrounded by different combinations of Wnt genes (Fig. 6.5 and 6.6).

6.3. Discussion

6.3.1. Role of *Pax6* orthologs during spider eye development

I did not detect expression for any of the newly annotated *Pax6* genes in any of the developing eyes of *P. tepidariorum* (Fig. 6.3). Together with the previous study of *Pax6* expression in *P. tepidariorum*, which also did not detect any expression of *Pax6* orthologs in the eye primordia of this spider[67], these results suggest that, in contrast to *C. salei* where one *Pax6* ortholog is expressed in the AM eyes[68], *Pax6* expression is not necessary for the establishment of the eye primordia in *P. tepidariorum*. In the horseshoe crab *Limulus polyphemus*, the only other chelicerate species where *Pax6* expression was analysed in relation to eye development, establishment of the eye primordia also appears to be independent of *Pax6* expression[254].

However, only one copy of *Pax6* was identified and characterised in *L. polyphemus*, even though the horseshoe crab lineage is now believed to have undergone at least two rounds of WGD[29,31]. Therefore, additional unannotated copies of *Pax6* might still be present in the genomes of horseshoe crabs. Furthermore, the authors of the previous study on retinal determination gene expression in *P. tepidariorum* proposed that *Pax6* might be required during early establishment of the eye primordia, as it is expressed in the anterior rim of the germ band in earlier embryonic stages[67]. Thus, I cannot exclude this hypothesis, as I have only characterised the expression of *Pax6* orthologs in *P. tepidariorum* from stage 10.1 onwards.

Despite the putative ancestral role of *Pax6* in metazoan eye development[213], it is still inconclusive whether or not this role is conserved in the chelicerate lineage. Identification and expression analysis of *Pax6* orthologs in other chelicerate species from different lineages, as well as functional analysis through RNAi knockdown in *P. tepidariorum*, could possibly address this question.

6.3.2. Role of *atonal* orthologs during spider eye development

One of the main downstream targets of the retinal determination network in *D. melanogaster* is the proneural gene *ato*, which plays an essential role in the differentiation of photoreceptor neurons[234]. Similar regulation is found in vertebrate eye development, with *Pax6* directly regulating *Ath5*, an ortholog of *ato* required to specify a subset of neurons in the retina[255]. To date, *ato* expression analysis during spider eye development was restricted to a study in *C. salei*, which did not detect *ato* expression in the developing eyes of this spider[68].

I identified two copies of *ato* in the genome of *P. tepidariorum*, one of which is expressed in both eye primordia as early as stage 10.1 (Fig. 6.4 A-F). Interestingly, expression appears to become restricted to a subset of cells in each eye primordia at later stages of eye develop-

ment (Fig. 6.4 E and F). This is evocative of *ato* expression in the eye disc of *D. melanogaster*, where it is at first broadly expressed in a proneural domain and subsequently becomes restricted to a single primary neuronal precursor in each developing ommatidia corresponding to the R8 photoreceptor cell[234]. Thus, it is possible that *Pt-ato1* is performing a similar function during *P. tepidariorum* eye development, at first specifying neuronal fate within the eye primordia before being restricted to a single neuronal cell type.

These results are inconsistent with what was observed in *C. salei*[68]. Furthermore, expression analysis of a horseshoe crab *ato* ortholog found no evidence for its involvement in eye development[254]. This suggests a scenario similar to the one found for *Pax6* in the chelicerate lineage, although, this time a potential role of *ato* in eye development was only observed in *P. tepidariorum*. However, as suggested above, additional copies of *ato* could be present in the genome of *L. polyphemus*, which may show expression during eye development. Nevertheless, why and how *ato* expression was lost in the developing eyes of *C. salei* remains to be addressed, since its position as a downstream target of the retinal determination gene network and associated role in specifying neuronal fate in the eye appears to be conserved in all animals studied so far[73,213,234,255].

To conclude, *Pt-ato1* appears to have conserved its function during eye development of *P. tepidariorum*, although, it is still uncertain whether this role is conserved in the chelicerate lineage. Expanding expression analysis to other chelicerate species, such as ticks, scorpions and harvestmen, might help to address this question in the future.

6.3.3. A conserved role of Wnt signalling in spider eye development

Several studies have found evidence for the regulation of eye development by Wnt signalling in highly divergent phyla, from the simple eyes of a planarian[220] to the highly complex eyes of insects and vertebrates[69,247]. This suggested that this signalling pathway is likely also involved in spider eye development. To investigate the potential role of Wnt signalling in the regulation of spider eye development, I analysed the expression patterns of Wnt genes with respect to *Pt-so1* expression in the developing head of *P. tepidariorum*.

I found that general Wnt expression in the developing head of *P. tepidariorum* is mostly contiguous to the eye primordia. This is especially true in the case of the lateral eye primordia, which is surrounded by the combined expression domains of *Pt-Wnt2*, *Pt-Wnt5*, *Pt-Wnt7.2* and *Pt-Wnt16* (Fig. 6.7). This pattern resembles that of Wnt expression during both insect compound eye and vertebrate camera eye development, where Wnt signalling restricts the expression of the retinal determination genes to their appropriate domains[245,248]. Thus, it is plausible to infer that Wnt signalling plays a role in restricting the expression of retinal determination genes, such as *Pt-so1*, and thus the extent of the retinal field, in the developing

head of *P. tepidariorum*. Moreover, this appears to be achieved by the combined effect of multiple Wnt ligands similar to what is observed in vertebrates[70].

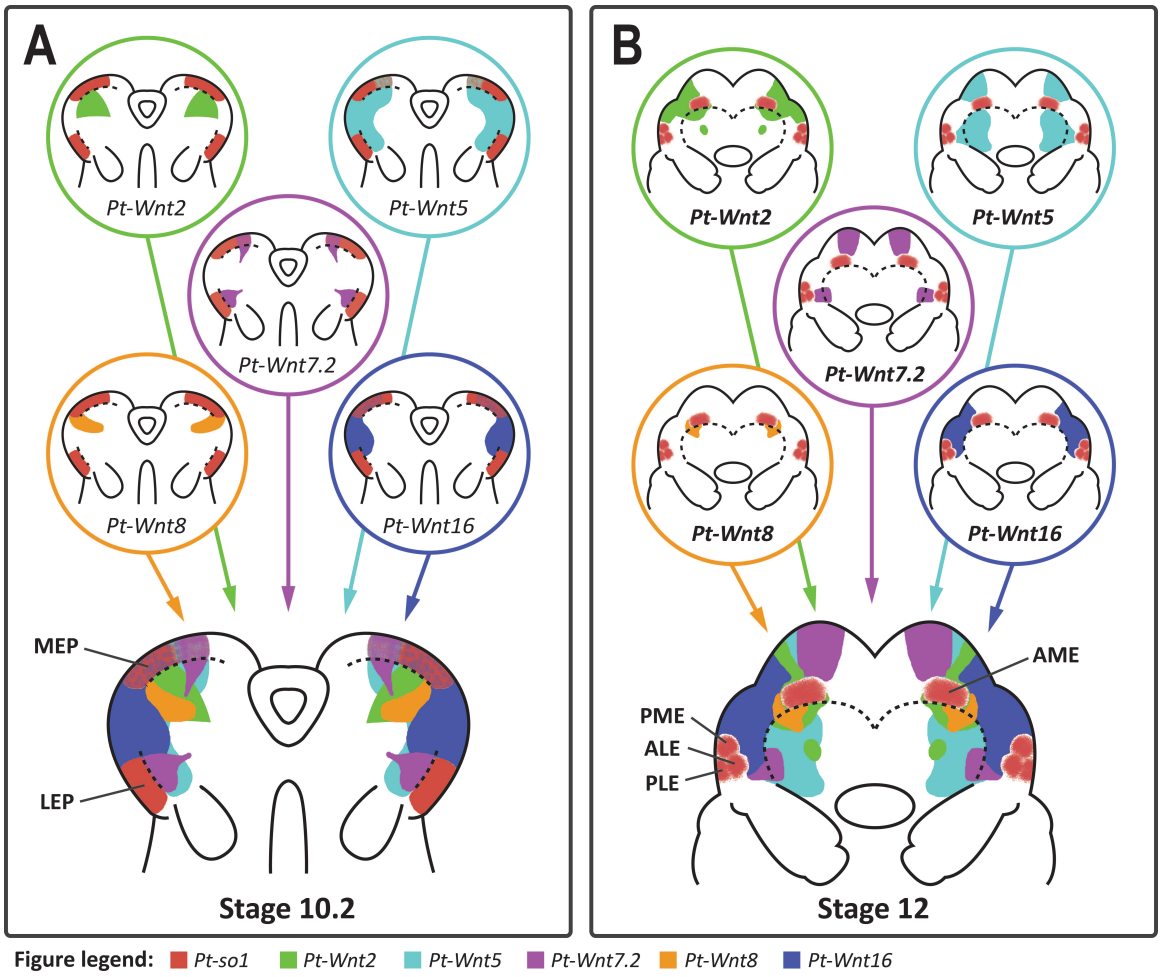

Figure 6.7: Summary of Wnt gene expression in the developing head of *P. tepidariorum*. Summary of the expression patterns of *Pt-Wnt2*, *Pt-Wnt5*, *Pt-Wnt7.2*, *Pt-Wnt8* and *Pt-Wnt16* in the developing head of stage 10.2 (A) and stage 12 (B) *P. tepidariorum* embryos, in relation to *Pt-so1* expression. Expression pattern of each individual Wnt gene (top) as well as the sum of all expression domains (bottom) is shown for both stages. MEP, median eye primordia; LEP, lateral eye primordia; AME, anterior median eye; ALE, anterior lateral eye; PME, posterior median eye; PLE, posterior lateral eye.

I observed a slightly different scenario in the median eye primordia. *Pt-so1* expression in the median eye primordia is contiguous to *Pt-Wnt2* and *Pt-Wnt8* expression, although, it is partially overlapped by the expression domains of *Pt-Wnt5*, *Pt-Wnt7.2* and *Pt-Wnt16* during the early stages of eye development. Interestingly, this is reminiscent of ocelli development in *D. melanogaster* (Fig. 6.1)[231], which are thought to be homologous to the median eyes of spiders[73]. The ocelli primordia are established in a region of the eye disc where *wingless* is expressed, and *otd* and *toy* activate the expression of *eya* and *so* respectively (Fig. 6.1)[231]. The median eye primordia of *P. tepidariorum* also exhibit this distinctive combination of *Pt-Otx2*, *Pt-so1* and *Pt-eya* expression[67]. Therefore, the gene network regulating the establishment of

the median eye primordia in *P. tepidariorum* appears to be conserved with respect to the ocelli primordia in *D. melanogaster*, however, the role played by *wg* in the latter may be carried by the combined effect of multiple Wnt ligands in the former.

In summary, my results suggest Wnt signalling may play a role in regulating eye development of spiders, much like its role in vertebrate and insect eye development. Furthermore, the regulation of each eye type appears to be mechanistically different, reminiscent of what is seen in *D. melanogaster* (Fig. 6.1). Nevertheless, further investigation is necessary to better understand the role of Wnt signalling during spider eye development, and functional analysis of Wnt genes in respect to their function in the developing head of *P. tepidariorum* is crucial for the validation of this hypothesis.

7. General discussion and Future perspectives

7.1. Impact of the Arachnopulmonata WGD on the evolution of developmentally important gene families

In this thesis, I provided further evidence supporting the timing of the WGD at the base of the arachnopulmonata, as well as the high prevalence of ohnolog expression divergence after WGD, by analysing two key developmental gene families: homeobox genes and Sox genes. To date, evidence for a non-WGD ancestry of harvestmen, which are considered to be an outgroup to spiders and scorpions (Fig. 1.3), was limited to the observation of fewer identified duplicated genes when compared to spiders and scorpions, particularly the presence of a single Hox cluster[61,256]. I have added to this by finding that the homeobox and Sox gene repertoires of the harvestmen *P. opilio* are represented by much fewer duplicated genes when compared to the repertoires of scorpions and spiders, consistent with a non-WGD ancestry (Fig. 3.1 A and 4.1).

However, these repertoires are likely to be incomplete, especially that of homeobox genes, which shows a much lower number of genes than the repertoire previously identified in *I. scapularis*[72]. Therefore, data on the putative ancestral state of homeobox and Sox gene repertoires is limited, even more so when taking into account that the tick lineage could have undergone a significant loss of gene content[72,167]. To help address this problem, additional genomic data on other putative non-WGD arachnids, such as camel spiders (Solifugae) and pseudoscorpions (Pseudoscorpiones), as well as higher quality genomic data on harvestmen, would be essential to better infer the state of gene content previous to the arachnopulmonata WGD.

Additionally, sequence data from additional arachnopulmonate species, especially representatives from the Amblypygi, Schizomida and Thelyphonida orders for which genomic data has not yet been produced, would also prove beneficial to study lineage specific gains and losses, as well as patterns of gene retention and functional divergence after the arachnopulmonate WGD. For example, a genome wide study comparing amphioxus to vertebrates detected a higher prevalence of expression specialization in retained ohnologs as well as a significant gain of open chromatin regions, suggestive of an increase in complexity of the gene regulatory landscape[42]. Another study, focused on the evolution of Sox gene repertoires after the teleost-specific WGD, found evidence supporting lineage-specific loss and divergence of retained Sox ohnologs[189]. Similar genome wide studies in the chelicerate lineage would increase our knowledge on possible general trends of genome evolution after WGD.

Expression analysis during embryogenesis of duplicated homeobox and Sox genes of *P. tepidariorum* and single-copy orthologs of *P. opilio* suggests a prevalence of ohnolog subfunc-

tionalization after genome duplication. Although much less prevalent, neofunctionalization through the gain of putatively novel expression domains was also observed (e.g. *Pt-Msx2*; Fig. 3.6). This is consistent with a recent genome wide study that found a high prevalence of subfunctionalization after WGD in the vertebrate lineage when compared to amphioxus, with the only caveat being that the authors found a much higher prevalence of specialization whereas my results suggest an equal proportion of subfunctionalization and specialization[42]. However, this could be explained by the fact that my analysis is limited to the homeobox and Sox gene families, while the analysis between amphioxus and vertebrates takes into account a much wider range of gene families. This could mean that the most common evolutionary fate of ohnologs could be dependent of the gene family analysed. Consistent with this hypothesis, a study comparing expression of duplicated Sox genes in the teleost lineage found most ohnologs were expressed in similar patterns[189], which is suggestive of subfunctionalization and similar to my observations on expression divergence of spider Sox ohnologs.

Another interesting result from my expression analysis was the high prevalence of ohnologs that were expressed in the nervous system. This putative biased retention of duplicated sequences with functions in the nervous system was previously reported in the vertebrate and ctenophore lineages[41,42,257]. It was proposed that this could be the outcome of duplicate retention by strong purifying selection due to the toxic nature of non-functional nervous system gene products[41]. Also interesting is the high prevalence of spider ohnologs with expression in the developing prosomal and opisthosomal appendages, which develop into lineage-specific traits such as the booklungs of arachnopulmonata and spinnerets of spiders. Functional analysis of these genes may reveal a role in the development of these novel structures, which would suggest that WGD could have facilitated the evolution of these novelties.

Lastly, expanding this type of analysis to other key developmental gene families, such as basic Helix-Loop-Helix (bHLH) and Forkhead box (FOX) gene families[258,259], would further improve our knowledge about the impact of the arachnopulmonata WGD on the evolution and divergence of arachnid development.

7.2. Role of *Pt-Sox21b-1* in spider segmentation

My analysis of *Pt-Sox21b-1* function during anterior segmentation in *P. tepidariorum* provides further evidence for its role as a gap-like gene in this spider[66]. My results also suggest that *Pt-Sox21b-1* specifically regulates the formation of L2-L4 segments, possibly acting in parallel with other gap-like genes, such as *Pt-hb* and *Pt-Dll*[56,208], to regulate downstream pair-rule-like genes (e.g. *Pt-h*) (Fig. 5.7)[66]. This is similar to the role of *Dichaete* during segmentation in *D. melanogaster*, where it directly regulates the expression of primary pair-rule genes, such as *eve* and *h*, in parallel to other gap genes[179,210].

However, it is still inconclusive whether or not *Pt-Sox21b-1* directly regulates *Pt-h* because the loss of *Pt-h* expression in *Pt-Sox21b-1* pRNAi knockdown embryos could still be the result of early indirect effects[66]. Analysis of *Pt-h* expression, as well as other pair-rule gene orthologs such as *Pt-eve* and *Pt-runt*[80], in *Pt-Sox21b-1* eRNAi knockdown clones would help investigate the effect *Pt-Sox21b-1* knockdown in the expression of these genes. For similar reasons, it may be reasonable to use the same approach to verify the effects of *Pt-Sox21b-1* knockdown on *Pt-hb*, *Pt-Dll* and *Pt-Msx1* expression.

My results also suggest that *Pt-Sox21b-1* does not directly regulate *Pt-Dl* expression, which is contrary to what was previously proposed[66]. Although, it is still possible that it regulates the early solid domain of *Pt-Dl* expression in the anterior SAZ (Fig. 5.7), which has been proposed to be necessary to activate and restrict *Pt-Wnt8* expression in the posterior SAZ[80]. Further analysis of *Pt-Dl* expression in *Pt-Sox21b-1* knockdown clones is necessary to determine if latter regulates the posterior-most expression domains of the former. Additionally, expression analysis of other posterior segmentation genes, such as *Pt-Wnt8*, *Pt-cad* and *Pt-h*, should also be carried out in *Pt-Sox21b-1* knockdown clones to verify the previously published results[66], since it is still possible these effects are due to the loss or incorrect specification of cells early in development.

Additionally, the role of *Dichaete* in segmentation appears to be conserved in *T. castaneum*, which has a similar mode of segmentation to that of spiders (i.e. short-germ segmentation)[179]. In this study, the authors proposed an ancestral regulatory gene network composed of *cad*, *Dichaete* and *opa* controlling posterior segmentation, downstream of Wnt signalling[179]. Studying possible interactions of *Pt-cad*, *Pt-Sox21b-1* and *Pt-opa* during spider posterior segmentation and how they compare to what is observed in *T. castaneum* may prove beneficial to better understand how segmentation evolved in arthropods.

7.3. Genetic basis of spider visual systems and possible conserved role of Wnt signalling in arthropod eye development

Consistent with previous work on retinal determination gene expression during eye development of *P. tepidariorum*[67], I did not detect expression of *Pax6* orthologs in the developing eyes of this spider. However, *Pax6* expression was previously detected in the AM eyes of another spider, *C. salei*[68]. Whether this represents a lineage-specific co-option of a *Pax6* ortholog to regulate eye development in *C. salei* or a lineage-specific loss of *Pax6* function in eye development of *P. tepidariorum* requires further investigation. Identification and expression analysis of *Pax6* orthologs in other species of spiders would help address this question. Expanding this analysis to other chelicerate orders would also be important to investigate whether or not *Pax6* role in eye development in conserved in this lineage, since *Pax6* expression was previously reported to be absent in the developing eyes of the horseshoe crab *L. polyphemus*[254].

It was also previously proposed that *Pax6* expression might be necessary during earlier stages to specify the future eye anlagen[67]. To further investigate this hypothesis, pRNAi knockdown of *P. tepidariorum Pax6* orthologs expressed at the anterior edge of the germ band in earlier stages of development could be carried out, followed by expression analysis of other retinal determination genes in obtained knockdown embryos, such as *Pt-so1*, to verify if regulation of eye development is affected.

I also found evidence for the involvement of an *ato* ortholog during spider eye development. Furthermore, it appears to play a role similar to the one observed in *D. melanogaster*, possibly specifying neuronal fate within the eye primordia at first and subsequently a specific photoreceptor[232]. However, *ato* expression was not detected in the developing eyes of *C. salei*[68], which is inconsistent with this scenario. Expression of *ato* was also not detected in the developing eyes of *L. polyphemus*[254]. Therefore, it is still uncertain whether or not the role of *ato* in eye development is conserved in the chelicerate lineage. Identification and expression analysis of *ato* orthologs from other species of chelicerates is necessary to address this question.

Additionally, my analysis of Wnt expression in the developing head of *P. tepidariorum* is suggestive of a conserved role of Wnt signalling in regulating the expression of retinal determination genes, similar to what is observed in both insects and vertebrates[70,244]. Moreover, I propose that the regulatory pathways establishing the primordia of each eye type in *P. tepidariorum* are mechanistically different, much like the regulatory pathways controlling ocelli and compound eye development in *D. melanogaster* (Fig. 6.1). To test the role of Wnt signalling in spider eye development, manipulation of Wnt activity could be achieved through the

knockdown of downstream components, such as *dishevelled* (*dsh*), or negative regulators, such as *axin* (*axn*)[242]. In *D. melanogaster*, reduction of Wnt signalling activity through the knockdown of *dsh* leads to the expansion of the compound eyes[260], and increase of Wnt activity through the knockdown of *axn* result in the formation of head cuticle instead of eye tissue[244]. Knockdown of individual Wnt genes may also provide information on possible interactions of specific Wnt ligands in the regulation of particular retinal determination genes.

To conclude, Wnt signalling appears to be regulating eye development by restricting the expressing of retinal determination genes, possibly controlling both eye size and location in the head. Therefore, it would be interesting to study the expression of Wnt and retinal determination genes in the developing eyes of other spiders with divergent eye morphologies, since differences in expression of these genes may reflect the phenotypic differences observed between species[73].

8. References

1. Hurles, M. Gene duplication: The genomic trade in spare parts. *PLoS Biol.* **2**, (2004).
2. Magadum, S., Banerjee, U., Murugan, P., Gangapur, D. & Ravikesavan, R. Gene duplication as a major force in evolution. *J. Genet.* **92**, 155–161 (2013).
3. Zhang, J. Evolution by gene duplication: An update. *Trends Ecol. Evol.* **18**, 292–298 (2003).
4. Ohno, S. *Evolution by Gene Duplication*. (Springer Berlin Heidelberg, 1970). doi:10.1007/978-3-642-86659-3
5. Kaessmann, H. Origins, evolution, and phenotypic impact of new genes. *Genome Res.* **20**, 1313–1326 (2010).
6. Reams, A. B. & Roth, J. R. Mechanisms of Gene Duplication and Amplification. *Cold Spring Harb. Perspect. Biol.* **7**, a016592 (2015).
7. Casola, C. & Betrán, E. The genomic impact of gene retrocopies: What have we learned from comparative genomics, population genomics, and transcriptomic analyses? *Genome Biol. Evol.* **9**, 1351–1373 (2017).
8. Zhou, Q. & Wang, W. On the origin and evolution of new genes-a genomic and experimental perspective. *J. Genet. Genomics* **35**, 639–648 (2008).
9. Hufton, A. L. & Panopoulou, G. Polyploidy and genome restructuring: a variety of outcomes. *Curr. Opin. Genet. Dev.* **19**, 600–606 (2009).
10. Conant, G. C., Birchler, J. A. & Pires, J. C. Dosage, duplication, and diploidization: Clarifying the interplay of multiple models for duplicate gene evolution over time. *Curr. Opin. Plant Biol.* **19**, 91–98 (2014).
11. Kasahara, M. The 2R hypothesis: an update. *Curr. Opin. Immunol.* **19**, 547–552 (2007).
12. Crow, K. D. & Wagner, G. P. What is the role of genome duplication in the evolution of complexity and diversity? *Mol. Biol. Evol.* **23**, 887–892 (2006).
13. Van De Peer, Y., Maere, S. & Meyer, A. The evolutionary significance of ancient genome duplications. *Nat. Rev. Genet.* **10**, 725–732 (2009).
14. Stella, M. K. & Stephan, C. F. Whole-genome duplication in teleost fishes and its evolutionary consequences Whole-Genome Duplication in Teleost Fishes and Its Evolutionary Consequences University of Zurich , Institute of Molecular Life Sciences , Neuroscience Center Zurich and Molecula. **289**, 1045–1060 (2014).
15. Lynch, M. & Conery, J. S. The evolutionary fate and consequences of duplicate genes. *Science (80-.).* **290**, 1151–1155 (2000).
16. Tutar, Y. Pseudogenes. *Comp. Funct. Genomics* **2012**, 1–4 (2012).
17. Zhang, J., Zhang, Y. ping & Rosenberg, H. F. Adaptive evolution of a duplicated pancreatic ribonuclease gene in a leaf-eating monkey. *Nat. Genet.* (2002). doi:10.1038/ng852
18. Sémon, M. & Wolfe, K. H. Preferential subfunctionalization of slow-evolving genes after allopolyploidization in Xenopus laevis. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 8333–8338 (2008).
19. Gout, J. F. & Lynch, M. Maintenance and loss of duplicated genes by dosage subfunctionalization. *Mol. Biol. Evol.* **32**, 2141–2148 (2015).
20. Kowalczuk, M. K., Nanda, S. & Moylan, E. C. Expression of concern: Subfunctionalization reduces the fitness cost of gene duplication in humans by buffering dosage imbalances. *BMC Genomics* **14**, (2013).
21. Force, A. *et al.* Preservation of Duplicate Genes by Complementary,

Degenerative Mutations. *Genetics* **151**, 1531–1545 (1999).
22. Storz, J. F. Gene Duplication and Evolutionary Innovations in Hemoglobin-Oxygen Transport. *Physiology* **31**, 223–232 (2016).
23. Assis, R. Drosophila duplicate genes evolve new functions on the fly. *Fly (Austin).* **8**, 6–10 (2014).
24. Varadharajan, S. *et al.* The Grayling Genome Reveals Selection on Gene Expression Regulation after Whole-Genome Duplication. *Genome Biol. Evol.* **10**, 2785–2800 (2018).
25. Zhang, J., Rosenberg, H. F. & Nei, M. Positive Darwinian selection after gene duplication in primate ribonuclease genes. *Proc. Natl. Acad. Sci.* **95**, 3708–3713 (1998).
26. Moriyama, Y. & Koshiba-Takeuchi, K. Significance of whole-genome duplications on the emergence of evolutionary novelties. *Brief. Funct. Genomics* **17**, (2018).
27. L., H. & C.H., H. 2R and remodeling of vertebrate signal transduction engine. *BMC Biol.* **8**, 146 (2010).
28. Hallinan, N. M. & Lindberg, D. R. Comparative analysis of chromosome counts infers three paleopolyploidies in the mollusca. *Genome Biol. Evol.* **3**, 1150–1163 (2011).
29. Kenny, N. J. *et al.* Ancestral whole-genome duplication in the marine chelicerate horseshoe crabs. *Heredity (Edinb).* **116**, 190–199 (2016).
30. Schwager, E. E. *et al.* The house spider genome reveals an ancient whole-genome duplication during arachnid evolution. *BMC Biol.* (2017). doi:10.1186/s12915-017-0399-x
31. Nossa, C. W. *et al.* Joint assembly and genetic mapping of the Atlantic horseshoe crab genome reveals ancient whole genome duplication. *Gigascience* **3**, 1–21 (2014).
32. Li, Z. *et al.* Multiple large-scale gene and genome duplications during the evolution of hexapods. *Proc. Natl. Acad. Sci.* 201710791 (2018). doi:10.1073/pnas.1710791115
33. Crow, K. D., Smith, C. D., Cheng, J. F., Wagner, G. P. & Amemiya, C. T. An independent genome duplication inferred from Hox paralogs in the American paddlefish-a representative basal ray-finned fish and important comparative reference. *Genome Biol. Evol.* **4**, 937–953 (2012).
34. Robertson, F. M. *et al.* Lineage-specific rediploidization is a mechanism to explain time-lags between genome duplication and evolutionary diversification. *Genome Biol.* **18**, 1–14 (2017).
35. Khaner, O. Evolutionary innovations of the vertebrates. *Integr. Zool.* **2**, 60–67 (2007).
36. Venkatesh, B. *et al.* Elephant shark genome provides unique insights into gnathostome evolution. *Nature* **505**, 174–179 (2014).
37. Huminiecki, L. & Conant, G. C. Polyploidy and the evolution of complex traits. *Int. J. Evol. Biol.* **2012**, 292068 (2012).
38. Huminiecki, L. *et al.* Emergence, development and diversification of the TGF-signalling pathway within the animal kingdom. *BMC Evol. Biol.* **9**, 1–17 (2009).
39. Holland, P. W. H. Evolution of homeobox genes. *Wiley Interdiscip. Rev. Dev. Biol.* **2**, 31–45 (2013).
40. Chen, Y. *et al.* Evolution of vertebrate central nervous system is accompanied by novel expression changes of duplicate genes. *J. Genet. Genomics* **38**, 577–584

(2011).
41. Roux, J., Liu, J. & Robinson-Rechavi, M. Selective constraints on coding sequences of nervous system genes are a major determinant of duplicate gene retention in vertebrates. *Mol. Biol. Evol.* **34**, 2773–2791 (2017).
42. Marlétaz, F. *et al.* Amphioxus functional genomics and the origins of vertebrate gene regulation. *Nature* **564**, 64–70 (2018).
43. Regier, J. C. *et al.* Arthropod relationships revealed by phylogenomic analysis of nuclear protein-coding sequences. *Nature* **463**, 1079–1083 (2010).
44. Schwager, E. E. *et al.* Chelicerata. in *Evolutionary Developmental Biology of Invertebrates* 99–139 (Springer, 2015). doi:10.1007/978-3-7091-1865-8_5
45. Sharma, P. P. *et al.* Phylogenomic interrogation of arachnida reveals systemic conflicts in phylogenetic signal. *Mol. Biol. Evol.* **31**, 2963–2984 (2014).
46. Ma, Y. *et al.* Extreme diversity of scorpion venom peptides and proteins revealed by transcriptomic analysis: Implication for proteome evolution of scorpion venom arsenal. *J. Proteomics* **75**, 1563–1576 (2012).
47. Clarke, T. H. *et al.* Multi-tissue transcriptomics of the black widow spider reveals expansions, co-options, and functional processes of the silk gland gene toolkit. *BMC Genomics* **15**, 1–17 (2014).
48. Haney, R. A. *et al.* Effects of gene duplication, positive selection, and shifts in gene expression on the evolution of the venom gland transcriptome in widow spiders. *Genome Biol. Evol.* **8**, 228–242 (2016).
49. Haney, R. A., Ayoub, N. A., Clarke, T. H., Hayashi, C. Y. & Garb, J. E. Dramatic expansion of the black widow toxin arsenal uncovered by multi-tissue transcriptomics and venom proteomics. *BMC Genomics* **15**, 1–18 (2014).
50. Sanggaard, K. W. *et al.* Spider genomes provide insight into composition and evolution of venom and silk. *Nat. Commun.* **5**, (2014).
51. Clarke, T. H., Garb, J. E., Hayashi, C. Y., Arensburger, P. & Ayoub, N. A. Spider Transcriptomes Identify Ancient Large-Scale Gene Duplication Event Potentially Important in Silk Gland Evolution. *Genome Biol. Evol.* **7**, 1856–1870 (2015).
52. Hilbrant, M., Damen, W. G. M. & McGregor, A. P. Evolutionary crossroads in developmental biology: the spider Parasteatoda tepidariorum. *Development* **139**, 2655–2662 (2012).
53. Mittmann, B. & Wolff, C. Embryonic development and staging of the cobweb spider Parasteatoda tepidariorumC. L. Koch, 1841 (syn.: Achaearanea tepidariorum; Araneomorphae; Theridiidae). *Dev. Genes Evol.* **222**, 189–216 (2012).
54. Akiyama-Oda, Y. Early patterning of the spider embryo: a cluster of mesenchymal cells at the cumulus produces Dpp signals received by germ disc epithelial cells. *Development* **130**, 1735–1747 (2003).
55. Akiyama-Oda, Y. & Oda, H. Axis specification in the spider embryo: dpp is required forradial-to-axial symmetry transformation and sog for ventralpatterning. *Development* **133**, 2347–2357 (2006).
56. Pechmann, M. *et al.* Novel function of Distal-less as a gap gene during spider segmentation. *PLoS Genet.* **7**, (2011).
57. McGregor, A. P. *et al.* Wnt8 Is Required for Growth-Zone Establishment and Development of Opisthosomal Segments in a Spider. *Curr. Biol.* **18**, 1619–1623 (2008).
58. Pechmann, M., McGregor, A. P., Schwager, E. E., Feitosa, N. M. & Damen, W. G. M.

Dynamic gene expression is required for anterior regionalization in a spider. *Proc. Natl. Acad. Sci.* **106**, 1468–1472 (2009).
59. Turetzek, N., Pechmann, M., Schomburg, C., Schneider, J. & Prpic, N. M. Neofunctionalization of a Duplicate dachshund Gene Underlies the Evolution of a Novel Leg Segment in Arachnids. *Mol. Biol. Evol.* **33**, 109–121 (2016).
60. Garwood, R. J., Sharma, P. P., Dunlop, J. A. & Giribet, G. A paleozoic stem group to mite harvestmen revealed through integration of phylogenetics and development. *Curr. Biol.* **24**, 1017–1023 (2014).
61. Sharma, P. P., Schwager, E. E., Extavour, C. G. & Giribet, G. Hox gene expression in the harvestman Phalangium opilio reveals divergent patterning of the chelicerate opisthosoma. *Evol. Dev.* **14**, 450–463 (2012).
62. Sharma, P. P., Schwager, E. E., Extavour, C. G. & Giribet, G. Evolution of the chelicera: A dachshund domain is retained in the deutocerebral appendage of Opiliones (Arthropoda, Chelicerata). *Evol. Dev.* **14**, 522–533 (2012).
63. Sharma, P. P., Schwager, E. E., Giribet, G., Jockusch, E. L. & Extavour, C. G. Distal-less and dachshund pattern both plesiomorphic and apomorphic structures in chelicerates: RNA interference in the harvestman Phalangium opilio (Opiliones). *Evol. Dev.* **15**, 228–242 (2013).
64. Allard, C. M. & Yeargan, K. V. Effect of Diet on Development and Reproduction of the Harvestman *Phalangium opilio* (Opiliones: Phalangiidae). *Environ. Entomol.* **34**, 6–13 (2005).
65. Bonatto Paese, C. L., Leite, D. J., Schönauer, A., McGregor, A. P. & Russell, S. Duplication and expression of Sox genes in spiders. *BMC Evol. Biol.* **18**, 205 (2018).
66. Paese, C. L. B., Schoenauer, A., Leite, D. J., Russell, S. & McGregor, A. P. A SoxB gene acts as an anterior gap gene and regulates posterior segment addition in a spider. *Elife* **7**, 1–43 (2018).
67. Schomburg, C. *et al.* Molecular characterization and embryonic origin of the eyes in the common house spider Parasteatoda tepidariorum. *Evodevo* **6**, (2015).
68. Samadi, L., Schmid, A. & Eriksson, B. J. Differential expression of retinal determination genes in the principal and secondary eyes of Cupiennius salei Keyserling (1877). *Evodevo* **6**, (2015).
69. Legent, K. & Treisman, J. E. Wingless Signaling in Drosophila Eye Development. in *Business* **469**, 141–161 (2008).
70. Fuhrmann, S. Wnt signaling in eye organogenesis. *Landes Biosci.* **4**, 60–67 (2008).
71. Janssen, R. *et al.* Conservation, loss, and redeployment of Wnt ligands in protostomes: implications for understanding the evolution of segment formation. *BMC Evol. Biol.* **10**, 374 (2010).
72. Leite, D. J. *et al.* Homeobox gene duplication and divergence in arachnids. *Mol. Biol. Evol.* **35**, (2018).
73. Morehouse, N. I., Buschbeck, E. K., Zurek, D. B., Steck, M. & Porter, M. L. Molecular evolution of spider vision: New opportunities, familiar players. *Biol. Bull.* **233**, 21–38 (2017).
74. Zhong, Y.-F., Butts, T. & Holland, P. W. H. HomeoDB: a database of homeobox gene diversity. *Evol. Dev.* **10**, 516–518 (2008).
75. Zhong, Y. & Holland, P. W. H. HomeoDB2: functional expansion of a comparative homeobox gene database for evolutionary developmental biology. *Evol. Dev.* **13**,

567–568 (2011).
76. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
77. Prpic, N. M., Schoppmeier, M. & Damen, W. G. M. Whole-mount in situ hybridization of spider embryos. *Cold Spring Harb. Protoc.* **3**, 2–6 (2008).
78. Clark, E. & Akam, M. Odd-paired controls frequency doubling in Drosophila segmentation by altering the pair-rule gene regulatory network. *Elife* **5**, 1–42 (2016).
79. Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).
80. Schönauer, A. *et al.* The Wnt and Delta-Notch signalling pathways interact to direct pair-rule gene expression via *caudal* during segment addition in the spider *Parasteatoda tepidariorum*. *Development* **143**, 2455–2463 (2016).
81. Holland, P. W. H., Booth, H. A. F. & Bruford, E. A. Classification and nomenclature of all human homeobox genes. *BMC Biol.* **5**, 1–28 (2007).
82. Quinonez, S. C. & Innis, J. W. Human HOX gene disorders. *Mol. Genet. Metab.* **111**, 4–15 (2014).
83. Zagozewski, J. L., Zhang, Q., Pinto, V. I., Wigle, J. T. & Eisenstat, D. D. The role of homeobox genes in retinal development and disease. *Dev. Biol.* **393**, 195–208 (2014).
84. Pick, L. Hox genes, evo-devo, and the case of the ftz gene. *Chromosoma* **125**, 535–551 (2016).
85. Davis, T. L. & Rebay, I. Master regulators in development: Views from the Drosophila retinal determination and mammalian pluripotency gene networks. *Dev. Biol.* **421**, 93–107 (2017).
86. Kozmik, Z. Pax genes in eye development and evolution. *Curr. Opin. Genet. Dev.* **15**, 430–438 (2005).
87. Kumar, J. P. The sine oculis homeobox (SIX) family of transcription factors as regulators of development and disease. *Cell. Mol. Life Sci.* **66**, 565–583 (2009).
88. Burke, A. C., Nelson, C. E., Morgan, B. A. & Tabin, C. Hox genes and the evolution of vertebrate axial morphology. *Development* **121**, 333–46 (1995).
89. Hughes, C. L. & Kaufman, T. C. Hox genes and the evolution of the arthropod body plan. *Evol. Dev.* **4**, 459–499 (2002).
90. Pascual-Anaya, J., D'Aniello, S., Kuratani, S. & Garcia-Fernàndez, J. Evolution of Hox gene clusters in deuterostomes. *BMC Dev. Biol.* **13**, (2013).
91. Lawrence, P. A. & Morata, G. Homeobox genes: Their function in Drosophila segmentation and pattern formation. *Cell* **78**, 181–189 (1994).
92. Durston, A. J. Vertebrate hox temporal collinearity: does it exist and what is it's function? *Cell Cycle* **18**, 523–530 (2019).
93. Zhu, K., Spaink, H. P. & Durston, A. J. Collinear Hox-Hox interactions are involved in patterning the vertebrate anteroposterior (A-P) axis. *PLoS One* **12**, 1–15 (2017).
94. Mallo, M. Reassessing the Role of Hox Genes during Vertebrate Development and Evolution. *Trends Genet.* **34**, 209–217 (2018).
95. Schneuwly, S., Klemenz, R. & Gehring, W. J. Redesigning the body plan of Drosophila by ectopic expression of the homoeotic gene Antennapedia. *Nature* **325**, 816–818 (1987).
96. Casaca, A., Santos, A. C. & Mallo, M. Controlling Hox gene expression and activity to build the vertebrate axial skeleton. *Dev. Dyn.* **243**, 24–36 (2014).

97. Vinagre, T. *et al.* Evidence for a Myotomal Hox/Myf Cascade Governing Nonautonomous Control of Rib Specification within Global Vertebral Domains. *Dev. Cell* **18**, 655–661 (2010).
98. Malicki, J., Schughart, K. & McGinnis, W. Mouse Hox-2.2 specifies thoracic segmental identity in Drosophila embryos and larvae. *Cell* **63**, 961–967 (1990).
99. McGinnis, N., Kuziora, M. A. & McGinnis, W. Human Hox-4.2 and Drosophila Deformed encode similar regulatory specificities in Drosophila embryos and larvae. *Cell* **63**, 969–976 (1990).
100. King, M. & Wilson, A. Evolution at two levels in humans and chimpanzees. *Science (80-.).* **188**, 107–116 (1975).
101. Averof, M. & Patel, N. H. Crustacean appendage evolution associated with changes in Hox gene expression. *Nature* **388**, 682–686 (1997).
102. Chan, Y. F. *et al.* Adaptive Evolution of Pelvic Reduction in Sticklebacks by Recurrent Deletion of a Pitx1 Enhancer. *Science (80-.).* **327**, 302–305 (2010).
103. Garcia-Fernàndez, J. The genesis and evolution of homeobox gene clusters. *Nat. Rev. Genet.* **6**, 881–892 (2005).
104. Ferrier, D. E. K. Evolution of Homeobox Gene Clusters in Animals: The Giga-Cluster and Primary vs. Secondary Clustering. *Front. Ecol. Evol.* **4**, 1–13 (2016).
105. Genikhovich, G. & Technau, U. On the evolution of bilaterality. *Development* **144**, 3392–3404 (2017).
106. Holland, P. W. H., Marlétaz, F., Maeso, I., Dunwell, T. L. & Paps, J. New genes from old: asymmetric divergence of gene duplicates and the evolution of development. *Philos. Trans. R. Soc. B Biol. Sci.* **372**, 20150480 (2017).
107. Ferguson, L. *et al.* Ancient Expansion of the Hox Cluster in Lepidoptera Generated Four Homeobox Genes Implicated in Extra-Embryonic Tissue Formation. *PLoS Genet.* **10**, (2014).
108. Soshnikova, N., Dewaele, R., Janvier, P., Krumlauf, R. & Duboule, D. Duplications of hox gene clusters and the emergence of vertebrates. *Dev. Biol.* **378**, 194–199 (2013).
109. Schoppmeier, M. & Damen, W. G. M. Expression of Pax group III genes suggests a single-segmental periodicity for opisthosomal segment patterning in the spider Cupiennius salei. *Evol. Dev.* **7**, 160–169 (2005).
110. Prpic, N. M., Janssen, R., Wigand, B., Klingler, M. & Damen, W. G. M. Gene expression in spider appendages reveals reversal of exd/hth spatial specificity, altered leg gap gene dynamics, and suggests divergent distal morphogen signaling. *Dev. Biol.* **264**, 119–140 (2003).
111. Schwager, E. E., Schoppmeier, M., Pechmann, M. & Damen, W. G. M. Duplicated Hox genes in the spider Cupiennius salei. *Front. Zool.* **4**, 1–11 (2007).
112. Sharma, P. P., Santiago, M. A., González-Santillán, E., Monod, L. & Wheeler, W. C. Evidence of duplicated Hox genes in the most recent common ancestor of extant scorpions. *Evol. Dev.* **17**, 347–355 (2015).
113. Cao, Z. *et al.* The genome of Mesobuthus martensii reveals a unique adaptation model of arthropods. *Nat. Commun.* **4**, 1–10 (2013).
114. Sharma, P. P., Schwager, E. E., Extavour, C. G. & Wheeler, W. C. Hox gene duplications correlate with posterior heteronomy in scorpions. *Proc. R. Soc. B Biol. Sci.* **281**, (2014).
115. Pechmann, M., Schwager, E. E., Turetzek, N. & Prpic, N. M. Regressive evolution of the arthropod tritocerebral segment linked to functional divergence of the Hox gene labial. *Proc. R. Soc. B Biol. Sci.* **282**, (2015).

116. Khadjeh, S. *et al.* Divergent role of the Hox gene Antennapedia in spiders is responsible for the convergent evolution of abdominal limb repression. *Proc. Natl. Acad. Sci.* **109**, 4921–4926 (2012).
117. Damen, W. G. M. Parasegmental organization of the spider embryo implies that the parasegment is an evolutionary conserved entity in arthropod embryogenesis. *Development* **129**, 1239–50 (2002).
118. Sharma, P. P. *et al.* A conserved genetic mechanism specifies deutocerebral appendage identity in insects and arachnids. *Proc. R. Soc. B Biol. Sci.* **282**, 20150698–20150698 (2015).
119. Castro, L. F. C., Rasmussen, S. L. K., Holland, P. W. H., Holland, N. D. & Holland, L. Z. A Gbx homeobox gene in amphioxus: Insights into ancestry of the ANTP class and evolution of the midbrain/hindbrain boundary. *Dev. Biol.* **295**, 40–51 (2006).
120. Reichert, H. A tripartite organization of the urbilaterian brain: Developmental genetic evidence from Drosophila. *Brain Res. Bull.* **66**, 491–494 (2005).
121. Su, C. Y., Kemp, H. A. & Moens, C. B. Cerebellar development in the absence of Gbx function in zebrafish. *Dev. Biol.* **386**, 181–190 (2014).
122. Chiang, C., Young, K. E. & Beachy, P. A. Control of Drosophila tracheal branching by the novel homeodomain gene unplugged, a regulatory target for genes of the bithorax complex. *Development* **121**, 3901–3912 (1995).
123. Kameda, Y. Hoxa3 and signaling molecules involved in aortic arch patterning and remodeling. *Cell Tissue Res.* **336**, 165–178 (2009).
124. Izpisúa-Belmonte, J. C. & Duboule, D. Homeobox genes and pattern formation in the vertebrate limb. *Dev. Biol.* **152**, 26–36 (1992).
125. Gordon, J. Hox genes in the pharyngeal region: How hoxa3 controls early embryonic development of the pharyngeal organs. *Int. J. Dev. Biol.* **62**, 775–783 (2018).
126. Shoji, H. *et al.* Regionalized expression of the Dbx family homeobox genes in the embryonic CNS of the mouse. *Mech. Dev.* **56**, 25–39 (1996).
127. Lacin, H., Zhu, Y., Wilson, B. A. & Skeath, J. B. Dbx mediates neuronal specification and differentiation through cross-repressive, lineage-specific interactions with eve and hb9. *Development* **136**, 3257–3266 (2009).
128. Cecchi, C., Mallamaci, A. & Boncinelli, E. Otx and Emx homeobox genes in brain development. *Int. J. Dev. Biol.* **44**, 663–668 (2000).
129. Hartmann, B., Hirth, F., Walldorf, U. & Reichert, H. Expression, regulation and function of the homeobox gene empty spiracles in brain and ventral nerve cord development of Drosophila. *Mech. Dev.* **90**, 143–153 (2000).
130. Sen, S., Reichert, H. & VijayRaghavan, K. Conserved roles of ems/Emx and otd/Otx genes in olfactory and visual system development in Drosophila and mouse. *Open Biol.* **3**, (2013).
131. Ebner, A. *et al.* Tracheal development in Drosophila melanogaster as a model system for studying the development of a branched organ. *Gene* **287**, 55–66 (2002).
132. Galle, S., Yanze, N. & Seipel, K. The homeobox gene Msx in development and transdifferentiation of jellyfish striated muscle. *Int. J. Dev. Biol.* **49**, 961–967 (2005).
133. Ramos, C. & Robert, B. msh/Msx gene family in neural development. *Trends Genet.* **21**, 624–632 (2005).
134. Bendall, A. J. & Abate-Shen, C. Roles for Msx and Dlx homeoproteins in

vertebrate development. *Gene* **247**, 17–31 (2000).
135. Friedmann, Y. & Daniel, C. W. Regulated expression of homeobox genes Msx-1 and Msx-2 in mouse mammary gland development suggests a role in hormone action and epithelial-stromal interactions. *Dev. Biol.* **177**, 347–355 (1996).
136. Alappat, S., Zhang, Z. Y. & Chen, Y. P. Msx homeobox gene family and craniofacial development. *Cell Res.* **13**, 429–442 (2003).
137. Knirr, S., Azpiazu, N. & Frasch, M. The role of the NK-homeobox gene slouch (S59) in somatic muscle patterning. *Development* **126**, 4525–4535 (1999).
138. Bae, Y. K. *et al.* Expression of sax1lnkx1.2 adn sax2lnk1.1 in zebrafish. *Gene Expr. Patterns* **4**, 481–486 (2004).
139. Holz, A. *et al.* The transcription factors Nkx2.2 and Nkx2.9 play a novel role in floor plate development and commissural axon guidance. *Development* **137**, 4249–4260 (2010).
140. Ranade, S. S. *et al.* Analysis of the Otd-dependent transcriptome supports the evolutionary conservation of CRX/OTX/OTD functions in flies and vertebrates. *Dev. Biol.* **315**, 521–534 (2008).
141. Steinmetz, P. R. H., Kostyuchenko, R. P., Fischer, A. & Arendt, D. The segmental pattern of otx, gbx, and Hox genes in the annelid Platynereis dumerilii. *Evol. Dev.* **13**, 72–79 (2011).
142. Knopp, P., Figeac, N., Fortier, M., Moyle, L. & Zammit, P. S. Pitx genes are redeployed in adult myogenesis where they can act to promote myogenic differentiation in muscle satellite cells. *Dev. Biol.* **377**, 293–304 (2013).
143. Angotzi, A. R., Ersland, K. M., Mungpakdee, S., Stefansson, S. & Chourrout, D. Independent and dynamic reallocation of pitx gene expression during vertebrate evolution, with emphasis on fish pituitary development. *Gene* **417**, 19–26 (2008).
144. Vorbrüggen, G. *et al.* Embryonic expression and characterization of a Ptx1 homolog in Drosophila. *Mech. Dev.* **68**, 139–147 (1997).
145. Winchell, C. J. & Jacobs, D. K. Expression of the Lhx genes apterous and lim1 in an errant polychaete: Implications for bilaterian appendage evolution, neural development, and muscle diversification. *Evodevo* **4**, 1–19 (2013).
146. Chou, S. J. & Tole, S. Lhx2, an evolutionarily conserved, multifunctional regulator of forebrain development. *Brain Res.* **1705**, 1–14 (2019).
147. Bieli, D. *et al.* The drosophila melanogaster mutants apblot and apXasta affect an essential apterous wing enhancer. *G3 Genes, Genomes, Genet.* **5**, 1129–1143 (2015).
148. Pueyo, J. I. & Couso, J. P. Chip-mediated partnerships of the homeodomain proteins Bar and Aristaless with the LIM-HOM proteins Apterous and Lim1 regulate distal leg development. *Development* **131**, 3107–3120 (2004).
149. Müller, D. *et al.* Regulation and functions of the lms homeobox gene during development of embryonic lateral transverse muscles and direct flight muscles in Drosophila. *PLoS One* **5**, (2010).
150. Cavodeassi, F., Modolell, J. & Gómez-Skarmeta, J. L. The Iroquois family of genes: From body building to neural patterning. *Development* **128**, 2847–2855 (2001).
151. Gómez-Skarmeta, J. L. & Modolell, J. Iroquois genes: Genomic organization and function in vertebrate neural development. *Curr. Opin. Genet. Dev.* **12**, 403–408 (2002).
152. Askary, A. *et al.* Iroquois Proteins Promote Skeletal Joint Formation by

Maintaining Chondrocytes in an Immature State. *Dev. Cell* **35**, 358–365 (2015).
153. Kim, K. H., Rosen, A., Bruneau, B. G., Hui, C. C. & Backx, P. H. Iroquois homeodomain transcription factors in heart development and function. *Circ. Res.* **110**, 1513–1524 (2012).
154. Atkey, M. R., Lachance, J. F. B., Walczak, M., Rebello, T. & Nilson, L. A. Erratum: Capicua regulates follicle cell fate in the Drosophila ovary through repression of mirror (Development vol. 133 (2115-2123)). *Development* **133**, 4794 (2006).
155. McNeill, H., Yang, C. H., Brodsky, M., Ungos, J. & Simon, M. A. Mirror encodes a novel PBX-class homeoprotein that functions in the definition of the dorsal-ventral border in the Drosophila eye. *Genes Dev.* **11**, 1073–1082 (1997).
156. Weiss, L. A. & Nieto, M. The crux of Cux genes in neuronal function and plasticity. *Brain Res.* **1705**, 32–42 (2019).
157. Nepveu, A. Role of the multifunctional CDP/Cut/Cux homeodomain transcription factor in regulating differentiation, cell growth and development. *Gene* **270**, 1–15 (2001).
158. Postigo, A. A., Ward, E., Skeath, J. B. & Dean, D. C. zfh-1, the Drosophila Homologue of ZEB, Is a Transcriptional Repressor That Regulates Somatic Myogenesis . *Mol. Cell. Biol.* **19**, 7255–7263 (1999).
159. Postigo, A. A. & Dean, D. C. ZEB, a vertebrate homolog of Drosophila Zfh-1, is a negative regulator of muscle differentiation. *EMBO J.* **16**, 3935–3943 (1997).
160. Komine, Y., Nakamura, K., Katsuki, M. & Yamamori, T. Novel transcription factor zfh-5 is negatively regulated by its own antisense RNA in mouse brain. *Mol. Cell. Neurosci.* **31**, 273–283 (2006).
161. Lai, Z., Fortini, M. E. & Rubin, G. M. The embryonic expression patterns of zfh-1 and zfh-2, two Drosophila genes encoding novel zinc-finger homeodomain proteins. *Mech. Dev.* **34**, 123–134 (1991).
162. Broihier, H. T., Moore, L. A., Van Doren, M., Newman, S. & Lehmann, R. zfh-1 is required for germ cell migration and gonadal mesoderm development in Drosophila. *Development* **125**, 655–666 (1998).
163. Holland, P. W. H., Garcia-Fernandez, J., Williams, N. A. & Sidow, A. Gene duplications and the origins of vertebrate development. *Development* **120**, 125–133 (1994).
164. Holland, L. Z. *et al.* The amphioxus genome illuminates vertebrate origins and cephalochordate biology (Genome Research (2008) 18, (1100-1111)). *Genome Res.* **18**, 1380 (2008).
165. Holland, L. Z. & Gibson-Brown, J. J. The Ciona intestinalis genome: When the constraints are off. *BioEssays* **25**, 529–532 (2003).
166. Takatori, N. *et al.* Comprehensive survey and classification of homeobox genes in the genome of amphioxus, Branchiostoma floridae. *Dev. Genes Evol.* **218**, 579–590 (2008).
167. Gulia-Nuss, M. *et al.* Genomic insights into the Ixodes scapularis tick vector of Lyme disease. *Nat. Commun.* **7**, (2016).
168. Grbić, M. *et al.* The genome of Tetranychus urticae reveals herbivorous pest adaptations. *Nature* **479**, 487–492 (2011).
169. Xu, P. X., Woo, I., Her, H., Beier, D. R. & Maas, R. L. Mouse Eya homologues of the Drosophila eyes absent gene require Pax6 for expression in lens and nasal placode. *Development* **124**, 219–231 (1997).
170. Noro, M., Sugahara, F. & Kuraku, S. Reevaluating Emx gene phylogeny: Homopolymeric amino acid tracts as a potential factor obscuring orthology

signals in cyclostome genes Genome evolution and evolutionary systems biology. *BMC Evol. Biol.* **15**, 1–13 (2015).
171. Kamachi, Y. & Kondoh, H. Sox proteins: regulators of cell fate specification and differentiation. *Development* **140**, 4129–4144 (2013).
172. Lefebvre, V., Dumitriu, B., Penzo-Méndez, A., Han, Y. & Pallavi, B. Control of cell fate and differentiation by Sry-related high-mobility-group box (Sox) transcription factors. *Int. J. Biochem. Cell Biol.* **39**, 2195–2214 (2007).
173. Bowles, J., Schepers, G. & Koopman, P. Phylogeny of the SOX family of developmental transcription factors based on sequence and structural indicators. *Dev. Biol.* **227**, 239–255 (2000).
174. Wilson, M. J. & Dearden, P. K. Evolution of the insect Sox genes. *BMC Evol. Biol.* **8**, 1–13 (2008).
175. Janssen, R. *et al.* Embryonic expression patterns and phylogenetic analysis of panarthropod sox genes: insight into nervous system development, segmentation and gonadogenesis. *BMC Evol. Biol.* **18**, 88 (2018).
176. Phochanukul, N. & Russell, S. No backbone but lots of Sox: Invertebrate Sox genes. *Int. J. Biochem. Cell Biol.* **42**, 453–464 (2010).
177. Reiprich, S. & Wegner, M. From CNS stem cells to neurons and glia: Sox for everyone. *Cell Tissue Res.* **359**, 111–124 (2015).
178. Sarkar, A. & Hochedlinger, K. The Sox family of transcription factors: Versatile regulators of stem and progenitor cell fate. *Cell Stem Cell* **12**, 15–30 (2013).
179. Clark, E. & Peel, A. D. Evidence for the temporal regulation of insect segmentation by a conserved sequence of transcription factors. *Dev.* **145**, 1–15 (2018).
180. Meng, F. W. & Biteau, B. A Sox Transcription Factor Is a Critical Regulator of Adult Stem Cell Proliferation in the Drosophila Intestine. *Cell Rep.* **13**, 906–914 (2015).
181. McKimmie, C., Woerfel, G. & Russell, S. Conserved genomic organisation of Group B Sox genes in insects. *BMC Genet.* **6**, 1–15 (2005).
182. Kirilly, D. *et al.* A genetic pathway composed of Sox14 and Mical governs severing of dendrites during pruning. *Nat. Neurosci.* **12**, 1497–1505 (2009).
183. Li, A. *et al.* Silencing of the Drosophila ortholog of SOX5 leads to abnormal neuronal development and behavioral impairment. *Hum. Mol. Genet.* **26**, 1472–1482 (2017).
184. Nanda, S. *et al.* Sox100B, a drosophila group e sox-domain gene, is required for somatic testis differentiation. *Sex. Dev.* **3**, 26–37 (2009).
185. Dichtel-Danjoy, M. L., Caldeira, J. & Casares, F. SoxF is part of a novel negative-feedback loop in the wingless pathway that controls proliferation in the Drosophila wing disc. *Development* **136**, 761–769 (2009).
186. Guth, S. I. E. & Wegner, M. Having it both ways: Sox protein function between conservation and innovation. *Cell. Mol. Life Sci.* **65**, 3000–3018 (2008).
187. Heenan, P., Zondag, L. & Wilson, M. J. Evolution of the Sox gene family within the chordate phylum. *Gene* **575**, 385–392 (2016).
188. Ritter, A. R. & Beckstead, R. B. Sox14 Is required for transcriptional and developmental responses to 20-hydroxyecdysone at the onset of drosophila metamorphosis. *Dev. Dyn.* **239**, 2685–2694 (2010).
189. Voldoire, E., Brunet, F., Naville, M., Volff, J. N. & Galiana, D. Expansion by whole genome duplication and evolution of the sox gene family in teleost fish. *PLoS One* **12**, 1–20 (2017).

190. King, N. *et al.* The genome of the choanoflagellate Monosiga brevicollis and the origin of metazoans. *Nature* **451**, 783–788 (2008).
191. Katoh, K. & Miyata, T. A heuristic approach of maximum likelihood method for inferring phylogenetic tree and an application to the mammalian SOX-3 origin of the testis-determining gene SRY. *FEBS Lett.* **463**, 129–132 (1999).
192. Kamachi, Y. *et al.* Evolution of non-coding regulatory sequences involved in the developmental process: Reflection of differential employment of paralogous genes as highlighted by Sox2 and group B1 Sox genes. *Proc. Japan Acad. Ser. B Phys. Biol. Sci.* **85**, 55–68 (2009).
193. Lee, H.-J. *et al.* Sox15 Is Required for Skeletal Muscle Regeneration. *Mol. Cell. Biol.* **24**, 8428–8436 (2004).
194. Schwager, E. E., Meng, Y. & Extavour, C. G. Vasa and piwi are required for mitotic integrity in early embryogenesis in the spider Parasteatoda tepidariorum. *Dev. Biol.* **402**, 276–290 (2015).
195. Zhong, L., Wang, D., Gan, X., Yang, T. & He, S. Parallel expansions of sox transcription factor group b predating the diversifications of the arthropods and jawed vertebrates. *PLoS One* **6**, (2011).
196. Tautz, D. Segmentation. *Dev. Cell* **7**, 301–312 (2004).
197. Chipman, A. D. Parallel evolution of segmentation by co-option of ancestral gene regulatory networks. *BioEssays* **32**, 60–70 (2010).
198. Damen, W. G. M. Evolutionary conservation and divergence of the segmentation process in arthropods. *Dev. Dyn.* **236**, 1379–1391 (2007).
199. Fritzenwanker, J. H., Uhlinger, K. R., Gerhart, J., Silva, E. & Lowe, C. J. Untangling posterior growth and segmentation by analyzing mechanisms of axis elongation in hemichordates. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 8403–8408 (2019).
200. Graham, A., Butts, T., Lumsden, A. & Kiecker, C. What can vertebrates tell us about segmentation? *Evodevo* **5**, 1–8 (2014).
201. Kaufholz, F. & Turetzek, N. Sox enters the picture. *Elife* **7**, 1–4 (2018).
202. Seaver, E. C., Yamaguchi, E., Richards, G. S. & Meyer, N. P. Expression of the pair-rule gene homologs runt, Pax3/7, even-skipped-1 and even-skipped-2 during larval and juvenile development of the polychaete annelid Capitella teleta does not support a role in segmentation. *Evodevo* **3**, 1–18 (2012).
203. Liu, P. Z. & Kaufman, T. C. Short and long germ segmentation: Unanswered questions in the evolution of a developmental mode. *Evol. Dev.* **7**, 629–646 (2005).
204. Palmeirim, I., Pais-de-Azevedo, T., Magno, R. & Duarte, I. Recent advances in understanding vertebrate segmentation. *F1000Research* **7**, 1–9 (2018).
205. Schroeder, M. D., Greer, C. & Gaul, U. How to make stripes: Deciphering the transition from nonperiodic to periodic patterns in Drosophila segmentation. *Development* **138**, 3067–3078 (2011).
206. Damen, W. G. M., Janssen, R. & Prpic, N. M. Pair rule gene orthologs in spider segmentation. *Evol. Dev.* **7**, 618–628 (2005).
207. Kanayama, M. *et al.* Travelling and splitting of a wave of hedgehog expression involved in spider-head segmentation. *Nat. Commun.* **2**, 1–7 (2011).
208. Schwager, E. E., Pechmann, M., Feitosa, N. M., McGregor, A. P. & Damen, W. G. M. hunchback Functions as a Segmentation Gene in the Spider Achaearanea tepidariorum. *Curr. Biol.* **19**, 1333–1340 (2009).
209. Oda, H. *et al.* Progressive activation of Delta-Notch signaling from around the

blastopore is required to set up a functional caudal lobe in the spider Achaearanea tepidariorum. *Development* **134**, 2195–2205 (2007).
210. Russell, S. R. H., Sanchez-Soriano, N., Wright, C. R. & Ashburner, M. The Dichaete gene of Drosophila melanogaster encodes a SOX-domain protein required for embryonic segmentation. *Development* **122**, 3669–3676 (1996).
211. Feitosa, N. M. *et al.* Molecular control of gut formation in the spider Parasteatoda tepidariorum. *Genesis* **55**, 1–17 (2017).
212. Darwin, C. *On the Origin of Species by Means of Natural Selection, or Preservation of Favoured Races in the Struggle for Life*. (John Murray, 1859).
213. Gehring, W. J. The evolution of vision. *Wiley Interdiscip. Rev. Dev. Biol.* **3**, 1–40 (2014).
214. Nilsson, D. E. & Bok, M. J. Low-resolution vision'at the hub of eye evolution. *Integr. Comp. Biol.* **57**, 1066–1070 (2017).
215. Oakley, T. H. & Speiser, D. I. How Complexity Originates: The Evolution of Animal Eyes. *Annu. Rev. Ecol. Evol. Syst.* **46**, 237–260 (2015).
216. Gavelis, G. S., Keeling, P. J. & Leander, B. S. How exaptations facilitated photosensory evolution: Seeing the light by accident. *BioEssays* **39**, 1–8 (2017).
217. de Jong, W. W., Hendriks, W., Mulders, J. W. M. & Bloemendal, H. Evolution of eye lens crystallins: the stress connection. *Trends Biochem. Sci.* **14**, 365–368 (1989).
218. Gehring, W. J. The master control gene for morphogenesis and evolution of the eye. *Genes to Cells* **1**, 11–15 (1996).
219. Tomarev, S. I. *et al.* Squid Pax-6 and eye development. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 2421–2426 (1997).
220. Su, H. *et al. A C-terminally truncated form of β-catenin acts as a novel regulator of Wnt/β-catenin signaling in planarians*. *PLoS Genetics* **13**, (2017).
221. Suga, H. *et al.* Flexibly deployed Pax genes in eye development at the early evolution of animals demonstrated by studies on a hydrozoan jellyfish. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 14263–14268 (2010).
222. Kozmik, Z. *et al.* Role of pax genes in eye evolution: A Cnidarian PaxB gene uniting Pax2 and Pax6 functions. *Dev. Cell* **5**, 773–785 (2003).
223. Silver, S. J. Signaling circuitries in development: insights from the retinal determination gene network. *Development* **132**, 3–13 (2004).
224. Treisman, J. E. & Heberlein, U. 4 Eye Development in Drosophila: Formation of the Eye Field and Control of Differentiation. in 119–158 (1998). doi:10.1016/S0070-2153(08)60454-8
225. Czerny, T. *et al.* Twin of eyeless, a second Pax-6 gene of Drosophila, acts upstream of eyeless in the control of eye development. *Mol. Cell* **3**, 297–307 (1999).
226. Wang, C. W. & Henry Sun, Y. Segregation of eye and antenna fates maintained by mutual antagonism in Drosophila. *Dev.* **139**, 3413–3421 (2012).
227. Kumar, J. P. & Moses, K. EGF Receptor and Notch signaling act upstream of Eyeless/Pax6 to control eye specification. *Cell* **104**, 687–697 (2001).
228. Halder, G. *et al.* Eyeless initiates the expression of both sine oculis and eyes absent during Drosophila compound eye development. *Development* **125**, 2181–2191 (1998).
229. Li, Y. *et al.* Optix functions as a link between the retinal determination network and the dpp pathway to control morphogenetic furrow progression in Drosophila. *Dev. Biol.* **381**, 50–61 (2013).

230. Yao, J. G. *et al.* Differential requirements for the Pax6(5a) genes eyegone and twin of eyegone during eye development in Drosophila. *Dev. Biol.* **315**, 535–551 (2008).
231. Blanco, J., Seimiya, M., Pauli, T., Reichert, H. & Gehring, W. J. Wingless and Hedgehog signaling pathways regulate orthodenticle and eyes absent during ocelli development in Drosophila. *Dev. Biol.* **329**, 104–115 (2009).
232. Zhou, Q., DeSantis, D. F., Friedrich, M. & Pignoni, F. Shared and distinct mechanisms of atonal regulation in Drosophila ocelli and compound eyes. *Dev. Biol.* **418**, 10–16 (2016).
233. Melicharek, D. *et al.* Identification of novel regulators of atonal expression in the developing Drosophila retina. *Genetics* **180**, 2095–2110 (2008).
234. Zhou, Q., DeSantis, D. F., Friedrich, M. & Pignoni, F. Shared and distinct mechanisms of atonal regulation in Drosophila ocelli and compound eyes. *Dev. Biol.* **418**, 10–16 (2016).
235. Zuber, M. E., Gestri, G., Viczian, A. S., Barsacchi, G. & Harris, W. A. Specification of the vertebrate eye by a network of eye field transcription factors. *Development* **130**, 5155–5167 (2003).
236. Acampora, D. *et al.* Forebrain and midbrain regions are deleted in Otx2(-/-) mutants due to a defective anterior neuroectoderm specification during gastrulation. *Development* **121**, 3279–3290 (1995).
237. Carl, M., Loosli, F. & Wittbrodt, J. Six3 inactivation reveals its essential role for the formation and patterning of the vertebrate eye. *Development* **129**, 4057–4063 (2002).
238. Grindley, J. C., Davidson, D. R. & Hill, R. E. The role of Pax-6 in eye and nasal development. *Development* **121**, 1433–1442 (1995).
239. Bernier, G. *et al.* Expanded retina territory by midbrain transformation upon overexpression of Six6 (Optx2) in Xenopus embryos. *Mech. Dev.* **93**, 59–69 (2000).
240. Heanue, T. A. *et al.* Dach1, a vertebrate homologue of Drosophila dachshund, is expressed in the developing eye and ear of both chick and mouse and is regulated independently of Pax and Eya genes. *Mech. Dev.* **111**, 75–87 (2002).
241. Davis, R. J. *et al.* Dach1 Mutant Mice Bear No Gross Abnormalities in Eye, Limb, and Brain Development and Exhibit Postnatal Lethality. *Mol. Cell. Biol.* **21**, 1484–1490 (2001).
242. Logan, C. Y. & Nusse, R. THE WNT SIGNALING PATHWAY IN DEVELOPMENT AND DISEASE. *Annu. Rev. Cell Dev. Biol.* **20**, 781–810 (2004).
243. Taciak, B., Pruszynska, I., Kiraga, L., Bialasek, M. & Krol, M. Wnt signaling pathway in development and cancer. *J. Physiol. Pharmacol.* **69**, 185–196 (2018).
244. Baonza, A. & Freeman, M. Control of Drosophila eye specification by Wingless signalling. *Development* **129**, 5313–5322 (2002).
245. Royet, J. & Finkelstein, R. Establishing primordia in the Drosophila eye-antennal imaginal disc: The roles of decapentaplegic, wingless and hedgehog. *Development* **124**, 4793–4800 (1997).
246. Magri, M. S., Dominguez-Cejudo, M. A. & Casares, F. Wnt controls the medial - lateral subdivision of the Drosophila head. *Biol. Lett.* **14**, (2018).
247. Cvekl, A. & Zhang, X. Signaling and Gene Regulatory Networks in Mammalian Lens Development. *Trends Genet.* **33**, 677–702 (2017).
248. Gunhaga, L. The lens: A classical model of embryonic induction providing new insights into cell determination in early development. *Philos. Trans. R. Soc. B*

Biol. Sci. **366**, 1193–1203 (2011).
249. Smith, A. N., Miller, L. A. D., Song, N., Taketo, M. M. & Lang, R. A. The duality of β-catenin function: A requirement in lens morphogenesis and signaling suppression of lens fate in periocular ectoderm. *Dev. Biol.* **285**, 477–489 (2005).
250. Kreslova, J. *et al.* Abnormal lens morphogenesis and ectopic lens formation in the absence of β-catenin function. *genesis* **45**, 157–168 (2007).
251. Lagutin, O. V. Six3 repression of Wnt signaling in the anterior neuroectoderm is essential for vertebrate forebrain development. *Genes Dev.* **17**, 368–379 (2003).
252. Sugiyama, Y., Lovicu, F. J. & McAvoy, J. W. Planar cell polarity in the mammalian eye lens. *Organogenesis* **7**, 191–201 (2011).
253. Maurus, D., Héligon, C., Bürger-Schwärzler, A., Brändli, A. W. & Kühl, M. Noncanonical Wnt-4 signaling and EAF2 are required for eye development in Xenopus laevis. *EMBO J.* **24**, 1181–1191 (2005).
254. Blackburn, D. C. *et al.* Isolation and expression of Pax6 and atonal homologues in the American horseshoe crab, Limulus polyphemus. *Dev. Dyn.* **237**, 2209–2219 (2008).
255. Vetter, M. L. & Brown, N. L. The role of basic helix-loop-helix genes in vertebrate retinogenesis. *Semin. Cell Dev. Biol.* **12**, 491–498 (2001).
256. Leite, D. J. & McGregor, A. P. Arthropod evolution and development: recent insights from chelicerates and myriapods. *Curr. Opin. Genet. Dev.* **39**, 93–100 (2016).
257. Fernández, R. & Gabaldón, T. Gene gain and loss across the metazoan tree of life. *Nat. Ecol. Evol.* (2020). doi:10.1038/s41559-019-1069-x
258. Dennis, D. J., Han, S. & Schuurmans, C. bHLH transcription factors in neural development, disease, and reprogramming. *Brain Res.* **1705**, 48–65 (2019).
259. Golson, M. L. & Kaestner, K. H. Fox transcription factors: From development to disease. *Dev.* **143**, 4558–4570 (2016).
260. Heslip, T. R., Theisen, H., Walker, H. & Marsh, J. L. SHAGGY and DISHEVELLED exert opposite effects on wingless and decapentaplegic expression and on positional identity in imaginal discs. *Development* **124**, 1069–1078 (1997).

Supplementary Material

Table S1. List of primers used in Chapter I.

Gene Name	Forward primer	Reverse primer	Fragment size
Po-Gbx	GTTCTTCGCCGGCTTCTTTC	TTTTCGGAGCGTTAGGGTTC	393 bp
Pt-Gbx1	ACGAACAACCTTTACTAGCG	CCAAGCTGATTATGAACTGC	240 bp
Pt-Gbx2	GGGACTTTATTCAACCGGAGC	CAGAGCATCGGAAAGGGAGA	651 bp
Pt-Hox3a	TGGCACATTCCATCATCACA	TTGTTTGCTGTTGGGTGACG	837 bp
Pt-Hox3b	TCCAAAACCGTCGCATGAAG	CACCGCTGTACAAAGTAGCC	770 bp
Po-Dbx	TTGCAGGTCAAAATATGGTTCC	GGAGGGTTGCTATTGACTCC	158 bp
Pt-Dbx1	ACAGTTACGCCTTTTCTGGT	ACAGCCCTCCTCATCATTCC	322 bp
Pt-Dbx2	GCAAGTCCACATGTTACCCA	ACGATGTCACTCAAATCAGGC	422 bp
Po-Emx	GTTAGCCGTACCCGTTTCAC	TGGAGTATGAAACCGTGGCT	648 bp
Pt-Emx1	TGCTTCTCCGGTTTCTGTCA	CGCACTTTTCATCACTGGTG	737 bp
Pt-Emx2	TTCCAAAACCAACACGACCT	GGTGACTGTGAACGGATGTA	832 bp
Pt-Emx3	CAGCCGAAGACCACATGAAC	GGTAGTGCCGTCATTTCTGG	808 bp
Pt-Emx4	CACCAACAAGAGCACCAGAA	GACGCACGACACAAATGAAT	868 bp
Po-Msx	CCAAATTCACCCTCACCGAC	GTTTCCGGTTGCTCTTGTGT	829 bp
Pt-Msx1	GCAAGCCTCGAACTCCTTTC	CCGTTTGGCACCATCTATGA	993 bp
Pt-Msx2	GCGTGGATATGGAAGAGTCG	GGTGCCATCAGAGGTCCTAA	556 bp
Pt-Msx3	GAAGTCCAAGTCCTGCGTTA	TCGGTAAAGACACAATGGGG	756 bp
Pt-Slou1	CGGCTCTCCTCCTGAAAGAT	ACTGTTTGCATCCATTCCCG	682 bp
Pt-Slou2	ACAGCCTGTGGAGATGACAA	ACGCAGCAGGTAAATAAGGAC	406 bp
Po-Vnd	CCTACATCCCTCCAACAGCA	CCGAACCACCACCGTCTATA	783 bp
Pt-Vnd1	TCCAACAGACGCAACAACAA	TTGTTCTACCACCAACGCTG	657 bp
Pt-Vnd2	ACCTTTTGCACACATGGGTC	CTCTCTTTCTGGGGCTGACA	505 bp
Pt-Otx1	GTCCCAATCACAAGCCAGTC	AGATGGGGCTTGGAATCGTC	390 bp
Pt-Otx2	GGCCACCAGTTACATCCAAC	CTGATGAGTTCTTCGACGGC	438 bp
Po-Pitx	GAGATTCAACGTCGTCAGGC	AGGTGGTGTGGGATTGCTTA	572 bp
Pt-Pitx1	GACCCATCACAGACTGGACA	CCAGCATATGTCCCAAAGCC	804 bp
Pt-Pitx2	ACTTGCCTGACTCAGCATCT	TAAGGACAAGGAGGTGCAGG	629 bp
Po-Ap	TCAACAGCAAGGCATGACAC	CTCGTTTGTTTCTGGGTGCA	777 bp
Pt-Ap1	GCTGCCACCTATCCATCTCT	TGAGACGCCAGTTCAGAGAG	820 bp
Pt-Ap2	CTACGAGATGTTTTCCCGCG	CCAGTGGGGTCATAGGATCA	812 bp
Pt-Ap3	AACCCATCCAGTGCTTCTCA	CCTCCAAGTTACGGGTAAAG	540 bp
Po-Irx1	AAATTACCAACGCCTAGCCG	CATCGTCGCACCTGTTTCTG	512 bp
Pt-Irx1	CACCACAAAGTAGCCAGCAG	TCCCAAGTGAAAACAGCGTG	920 bp
Pt-Irx2	CGTAGAACAGCATCGTGTGA	TTTATTGCAGGGATGTCGCC	862 bp
Pt-Irx3	GCTCGCCATCATCACCAAAA	GAGAACCTCTTGACAGCACT	826 bp
Pt-Irx4	AATGCCAGGAAAGTGACTCC	TGTTAGCTCTCACACCGACT	835 bp
Po-ct	CGCCTTCCAGTTCTTCATCG	GGGGTAGGGGTCCATTGAAA	686 bp
Pt-ct1	GTTGGGCCTATCTATCCGCACT	CCACTGCTTGATGGAGGTGGTA	756 bp
Pt-ct2	GGAGCAGAAGGAAGCCTTGAGA	TCCTGTTCGTTACCGCTGTCAT	774 bp
Po-Zfh	GTCTGAACGCCTCGATGAAC	GAATTTCGGGTCTTGCGGTT	983 bp
Pt-Zfh1	CCAGAGACTACCCCAACACCTG	TAGGCAATGGCGTAACACCAGA	568 bp
Pt-Zfh2	AAGTTCCTTTGAGCATCGCAGC	GTCTCGGTCGAGGAAATGGACA	501 bp

Table S2. List of primers used in Chapter II and III.

Gene Name	Forward primer	Reverse primer	Fragment size
Pt-Dichaete	GCCCAAAGAAGGAAGATCGC	GGAGTAGAGTCCGAGCAACT	858 bp
Po-Sox21b	CCTCATCATCTTAGCCCCGT	GGCATGGACATTGGGTTCTG	806 bp
Pt-Sox21b-1 (ISH)	ACGCTTTCATGGTTTGGTCA	ATGCTGGGACCTGGAGAAAT	790 bp
Pt-Sox21b-1 (dsRNA)	TGCAAGCTCCGCAAATCGTAC	AGAAGAGGCAGGATAGCCGC	548 bp
Pt-Sox21b-2	CGCGAGTTCAGAGAAGACGA	AGAGCTTGAAATGGCAGACA	458 bp
Po-SoxN	TGAACGGTTACATGCCCAAC	ACGCCGAGATAACAACATCC	670 bp
Pt-SoxN	GACCGCAAGCCATCATTCTC	GGTCGTGAGTTGGAAATCGG	819 bp
Pt-Sox21a-1	GCGGTGAAGGAAGAAGAACG	AGTGATGACAGGTGGTGAGG	822 bp
Pt-Sox21a-2	GAAGCTGAGAAACGCCCTTT	AGGACTTGGTGGCAATGTTG	603 bp
Po-SoxC	AAAAGGCCCATGAACGCTTT	GTGGTGGTGGTGGATTCAAC	941 bp
Pt-SoxC1	TGTGAAAAGGCCCATGAACG	CGTTAAGTCGCTGATGTCCG	760 bp
Pt-SoxC2	CCAAAATCCCGGCATTCCAT	AGCAGCCCAATCATCTCTGT	667 bp
Po-SoxD	CGACGCAAAATTCTCAAGGC	AACCTCCGTCCGCCATTAAT	515 bp
Pt-SoxD1	TGACCGAAGCTCAGTTTCCA	TATTGTTTGGTGGCGGTTGG	1047 bp
Pt-SoxD2	CGTCTCCTATAGCTCCACCC	CACGATACCATAGAGTGCGC	927 bp
Po-SoxE1	GGCGAGTTTTAAGCGACGAA	GAAAGGCGGTTCAGGGAATC	848 bp
Po-SoxE2	AACGCTTTCATGGTATGGGC	GAGGAGGAGGAGTGTTGCTT	250 bp
Pt-SoxE1	CGGGGTTATGACTGGAGTGT	TGTCCACTGTTCGTCCATGA	753 bp
Pt-SoxE2	GACGATGACCTGGACAATGC	TGGATTGATGATGCTGGGGT	1186 bp
Pt-SoxF1	GTCACCCAACTGCGACTTTT	ACGTAATGAAATCGGAGCGC	1015 bp
Pt-SoxF2	ATGGAGAGGGTTGAGTCACG	TGCTGGTAGTACACGACGTT	960 bp
Pt-Dll	TCCGTACCTGGGTTCGTATC	TGTCCCATGAGGAGATAGGC	653 bp
Pt-hb	GCCGAGACCTTTCAAGTGTC	TCTGTTCCTCGATTGGCTGT	823 bp
Pt-Msx1	GCAAGCCTCGAACTCCTTTC	CCGTTTGGCACCATCTATGA	993 bp
Pt-Dl	ACAAACCACACGGCTTTTTC	GCTTGGTCAAGCAGTCATCA	967 bp

Table S3. List of primers used in Chapter IV.

Gene Name	Forward primer	Reverse primer	Fragment size
Pt-ato1	ACGTGCTTTTGATGCGGAAA	TCATAACAGCAGGCGAAGGA	862 bp
Pt-ato2	TGCATTCAAGTATCATTGCTCAG	GCTGTGATATAAGTTTGGGCCA	629 bp
Pt-Toy	GGCGAACCGTAGATCAAGTG	CTGTGGTCTAAGGCAGGTGT	589 bp
Pt-Eyg	TACATCACCAGTCACCCCAC	TCTACGAAACTTGGGCCGAT	868 bp
Pt-so1	TCACGCCAAACTCTCATCCT	ATGGTTCTGCATCGAGTCCA	575 bp
Pt-Wnt2	TTGATGGCGTTCTGTTGTCC	CCCAGAGTACCCAACGACTT	836 bp
Pt-Wnt5	TGACACCTGTTGAGCGAAAC	ATCCTGTCTCTTCGTTGGGG	1028 bp
Pt-Wnt7.2	CCCTGAAGGTGGATGGAAGT	TGGATGAATGGTGCGAGTCA	663 bp
Pt-Wnt8	TGTCCGGTTCAAAAGCTGTG	AACTTGCTCTCTCTGCGTCT	795 bp
Pt-Wnt16	AAAGGGATGTGAGGGCTCTT	TTGTCGCCAGGTTTGTTTGA	654 bp

Table S4. List of universal primers used in this thesis.

Primer Name	Primer sequence
M13 Forward	GTAAAACGACGGCCAG
M13 Reverse	CAGGAAACAGCTATGAC
3' T7 Reverse	AGGGATCCTAATACGACTCACTATAGGGCCCGGGGC
T7 Forward	TAATACGACTCACTATAGGG
T7/T3 Reverse	TAATACGACTCACTATAGGGAATTAACCCTCACTAAAGGGA

Table S5. List of PCR results and summary of key expression domains from *in situ* hybridization staining for each gene analysed in Chapter I

Phalangium opilio			*Parasteatoda tepidariorum*		
Gene	**PCR**	**Expression domains**	**Gene**	**PCR**	**Expression domains**
Gbx	Successful	Prosomal appendages, Neuroectoderm	***Gbx1***	Successful	No expression
			Gbx2	Successful	Prosomal appendages, Opisthosomal appendages
Hox3	NA	Prosomal appendages, Neuroectoderm	***Hox3a***	Successful	Early segmental
			Hox3b	Successful	Segmental expression, Prosomal appendages, Neuroectoderm
Dbx	Successful	Prosomal appendages, Neuroectoderm	***Dbx1***	Successful	Neuroectoderm
			Dbx2	Successful	No expression
Emx	Successful	Segmental expression, Prosomal appendages, Neuroectoderm	***Emx1***	Successful	Segmental expression
			Emx2	Successful	Segmental expression, Prosomal appendages
			Emx3	Successful	Segmental expression, Prosomal appendages, Neuroectoderm
			Emx4	Successful	Segmental expression, Prosomal appendages, Neuroectoderm
Msx	Successful	Segmental expression, Prosomal appendages, Neuroectoderm	***Msx1***	Successful	Segmental expression, Neuroectoderm
			Msx2	Successful	Chelicerae
			Msx3	Successful	Prosomal appendages, Opisthosomal appendages
Slou	Successful	No expression	***Slou1***	Successful	Prosomal appendages, Opisthosomal appendages
			Slou2	Successful	No expression
Vnd	Successful	Neuroectoderm	***Vnd1***	Successful	Neuroectoderm
			Vnd2	Successful	No expression
Otx	NA	NA	***Otx1***	Successful	Early segmental, Neuroectoderm, Opisthosomal appendages
			Otx2	Successful	Neuroectoderm, Median eye primordia
Pitx	Successful	Neuroectoderm	***Pitx1***	Successful	Neuroectoderm
			Pitx2	Successful	Neuroectoderm
Ap	Successful	Segmental expression, Neuroectoderm	***Ap1***	Successful	Prosomal appendages, Opisthosomal appendages
			Ap2	Successful	Segmental expression, Opisthosomal appendages
			Ap3	Successful	No expression
Irx1	Successful	Segmental expression, Prosomal appendages, Neuroectoderm, Germ band edge	***Irx1***	Successful	Prosomal appendages, Neuroectoderm
			Irx2	Successful	Germ band edge
			Irx3	Successful	Prosomal appendages
Irx2	Unsuccessful	NA	***Irx4***	Successful	Segmental expression, Prosomal appendages, Neuroectoderm
Irx3	Unsuccessful	NA			
Ct	Successful	Segmental expression, Prosomal appendages, Neuroectoderm	***Ct1***	Successful	Prosomal appendages
			Ct2	Successful	Segmental expression, Prosomal appendages, Opisthosomal appendages
Zfh	Successful	Prosomal appendages, Neuroectoderm	***Zfh1***	Successful	Prosomal appendages, Neuroectoderm
			Zfh2	Successful	Prosomal appendages

Table S6. List of PCR results and summary of key expression domains from *in situ* hybridization staining for each gene analysed in Chapter II

Phalangium opilio			*Parasteatoda tepidariorum*		
Gene	**PCR**	**Expression domains**	**Gene**	**PCR**	**Expression domains**
Dichaete	Unsuccessful	NA	***Dichaete***	Successful	Ubiquitous, Segmental expression
Sox21a	NA	NA	***Sox21a1***	Successful	Segmental expression, Neuroectoderm, Chelicerae
			Sox21a2	Successful	Neuroectoderm, Prosomal appendages
Sox21b	Successful	Segmental expression, Neuroectoderm	***Sox21b1***	Successful	Segmental expression, Neuroectoderm
			Sox21b2	Successful	Neuroectoderm
SoxN	Successful	Prosomal appendages, Neuroectoderm	***SoxN***	Successful	Prosomal appendages, Neuroectoderm
SoxC	Successful	Prosomal appendages, Neuroectoderm	***SoxC1***	Successful	Prosomal appendages, Neuroectoderm
			SoxC2	Successful	No expression
SoxD	Successful	Segmental expression, Prosomal appendages, Pre-cheliceral region edge	***SoxD1***	Unsuccessful	NA
			SoxD2	Successful	Segmental expression, Prosomal appendages, Opisthosomal appendages, Pre-cheliceral region edge
SoxE1	Unsuccessful	NA	***SoxE1***	Successful	Prosomal appendages, Opisthosomal appendages, Pre-cheliceral region
SoxE2	Unsuccessful	NA	***SoxE2***	Successful	Prosomal appendages, Opisthosomal appendages, Pre-cheliceral region, Germline progenitors
SoxF	NA	NA	***SoxF1***	Successful	Prosomal appendages, Opisthosomal appendages, Lateral eye primordia
			SoxF2	Successful	Prosomal appendages, Opisthosomal appendages, Non-neurogenic ectoderm edge

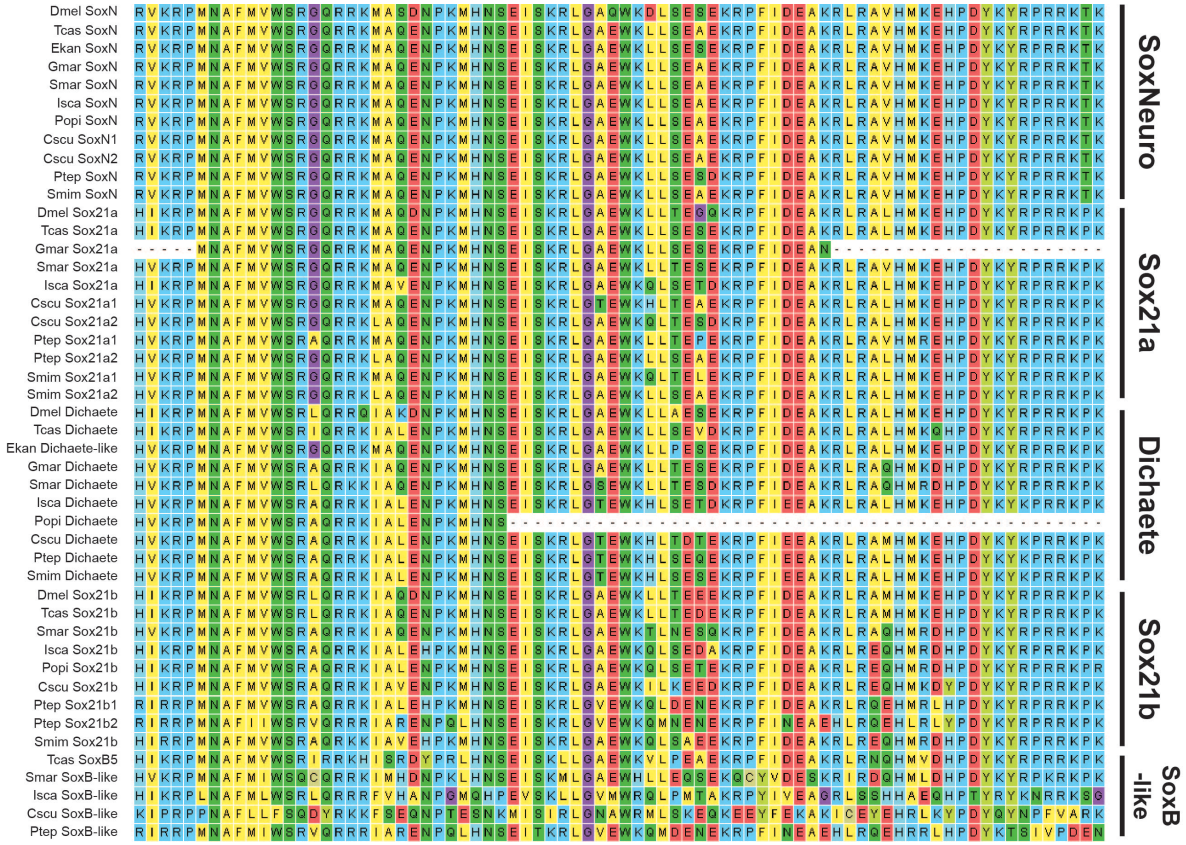

Figure S1: Multiple sequence alignment of Panarthropoda Sox Group B HMG domain sequences. Alignment was performed in the MEGA software using the MUSCLE algorithm.

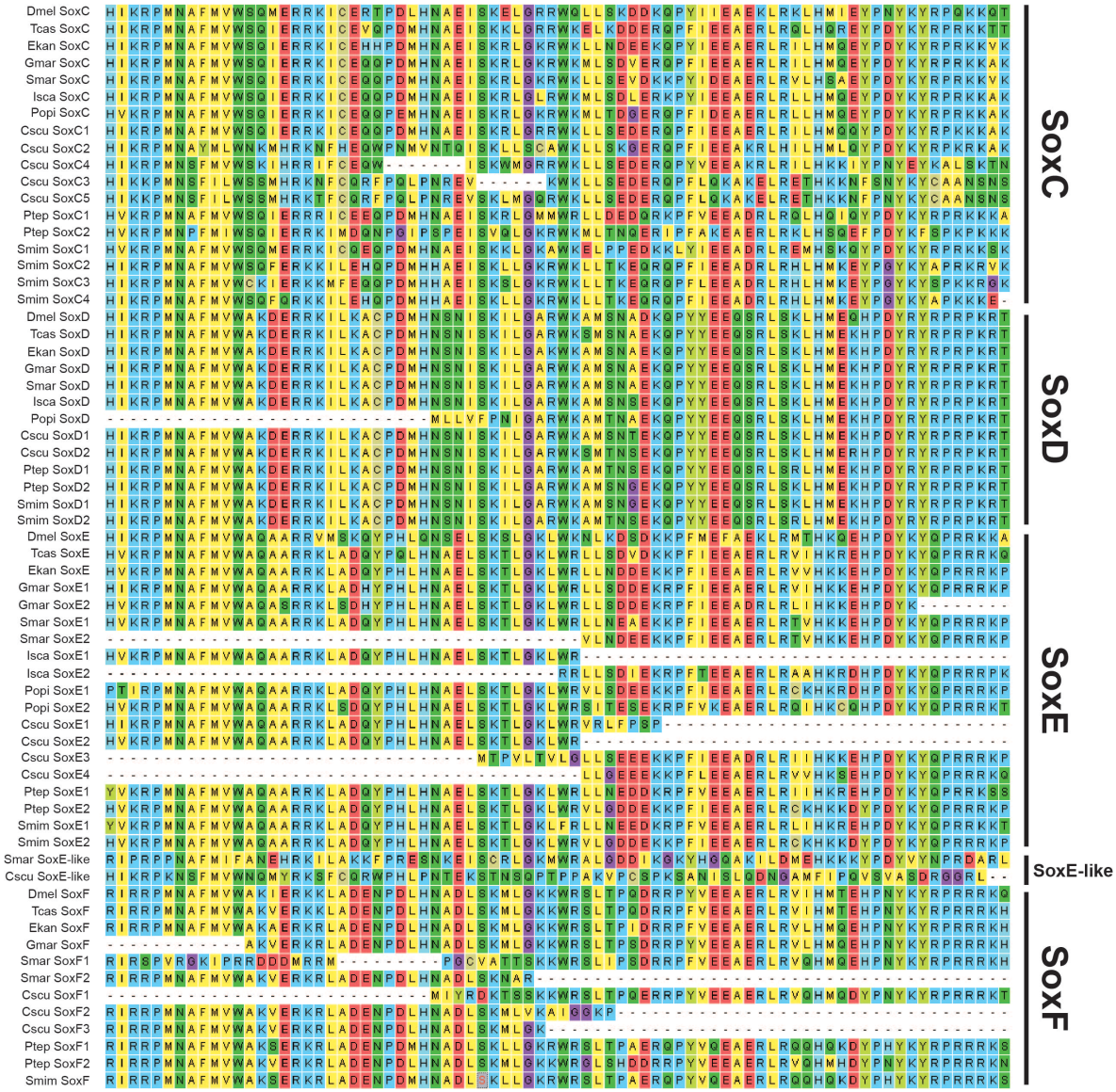

Figure S2: Multiple sequence alignment of Panarthropoda Sox Group C-F HMG domain sequences. Alignment was performed in the MEGA software using the MUSCLE algorithm.

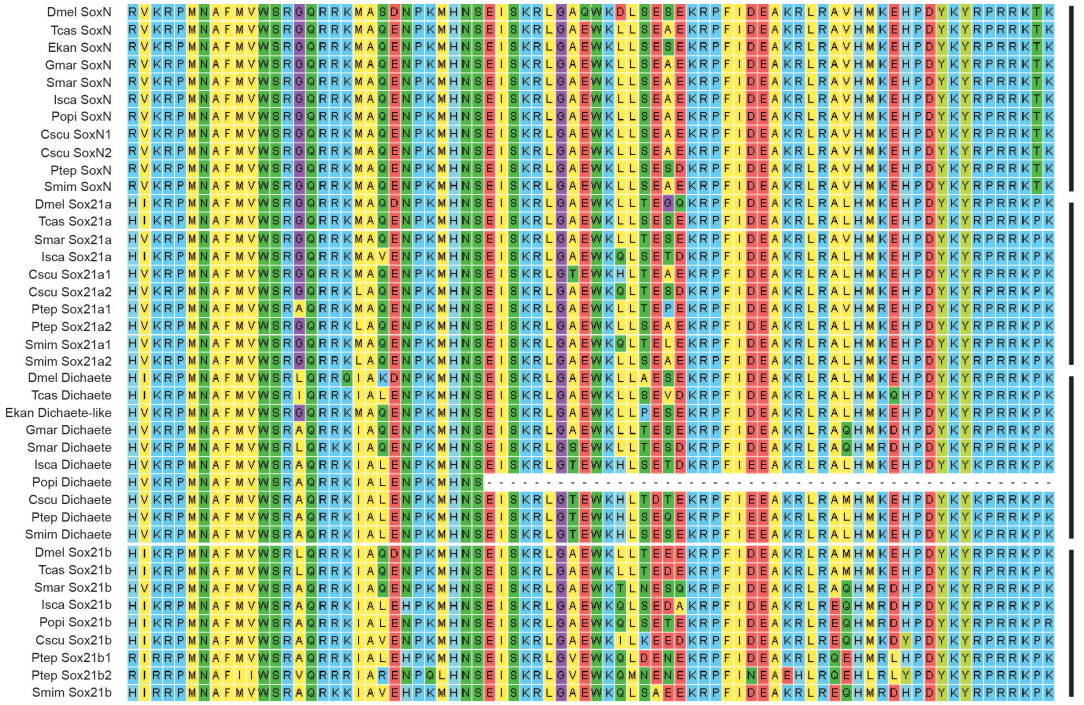

Figure S3: Multiple sequence alignment of Panarthropoda Sox Group B HMG domain sequences after removing highly divergent sequences. Alignment was performed in the MEGA software using the MUSCLE algorithm.

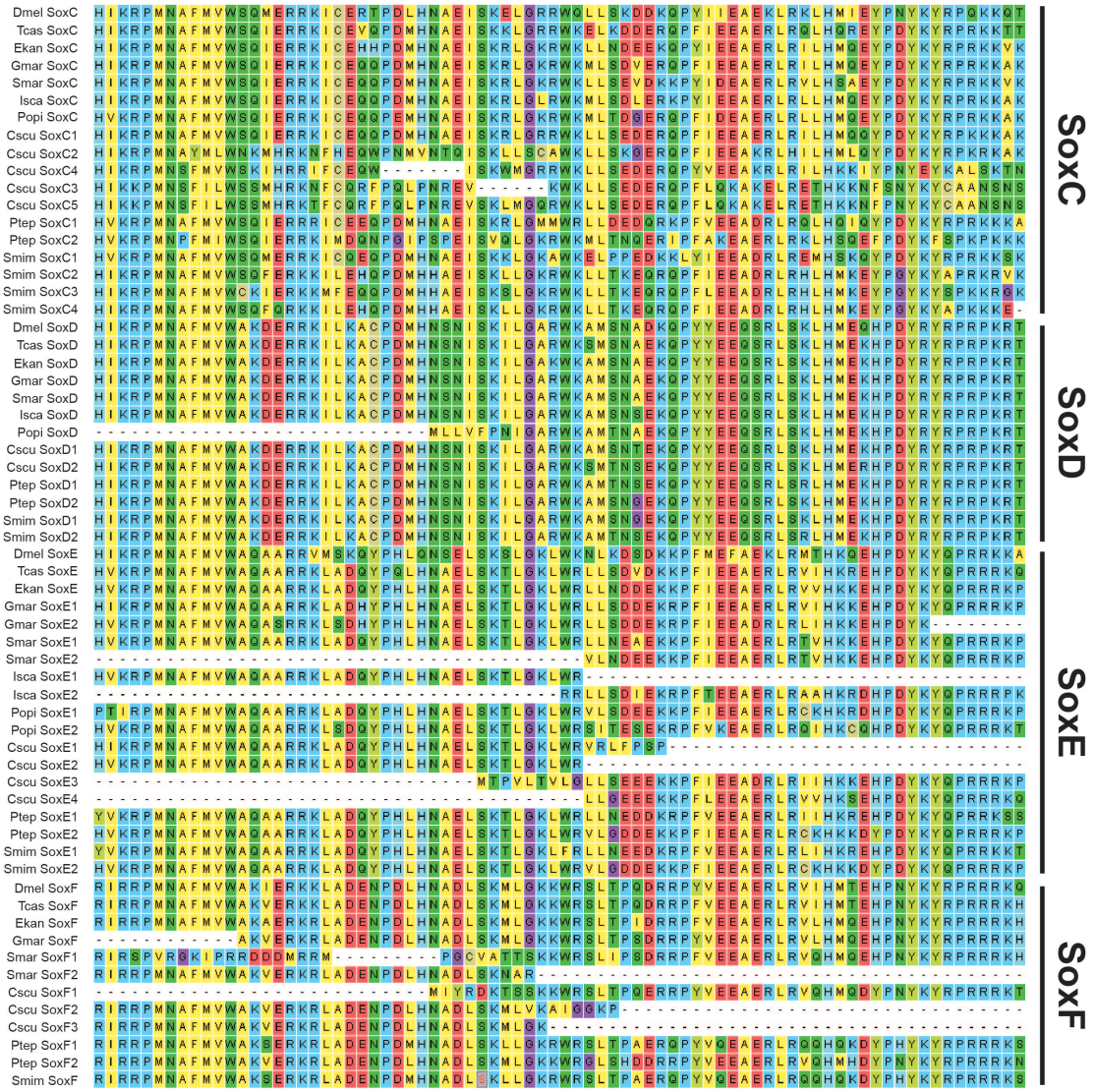

Figure S4: Multiple sequence alignment of Panarthropoda Sox Group C-F HMG domain sequences after removing highly divergent sequences. Alignment was performed in the MEGA software using the MUSCLE algorithm.

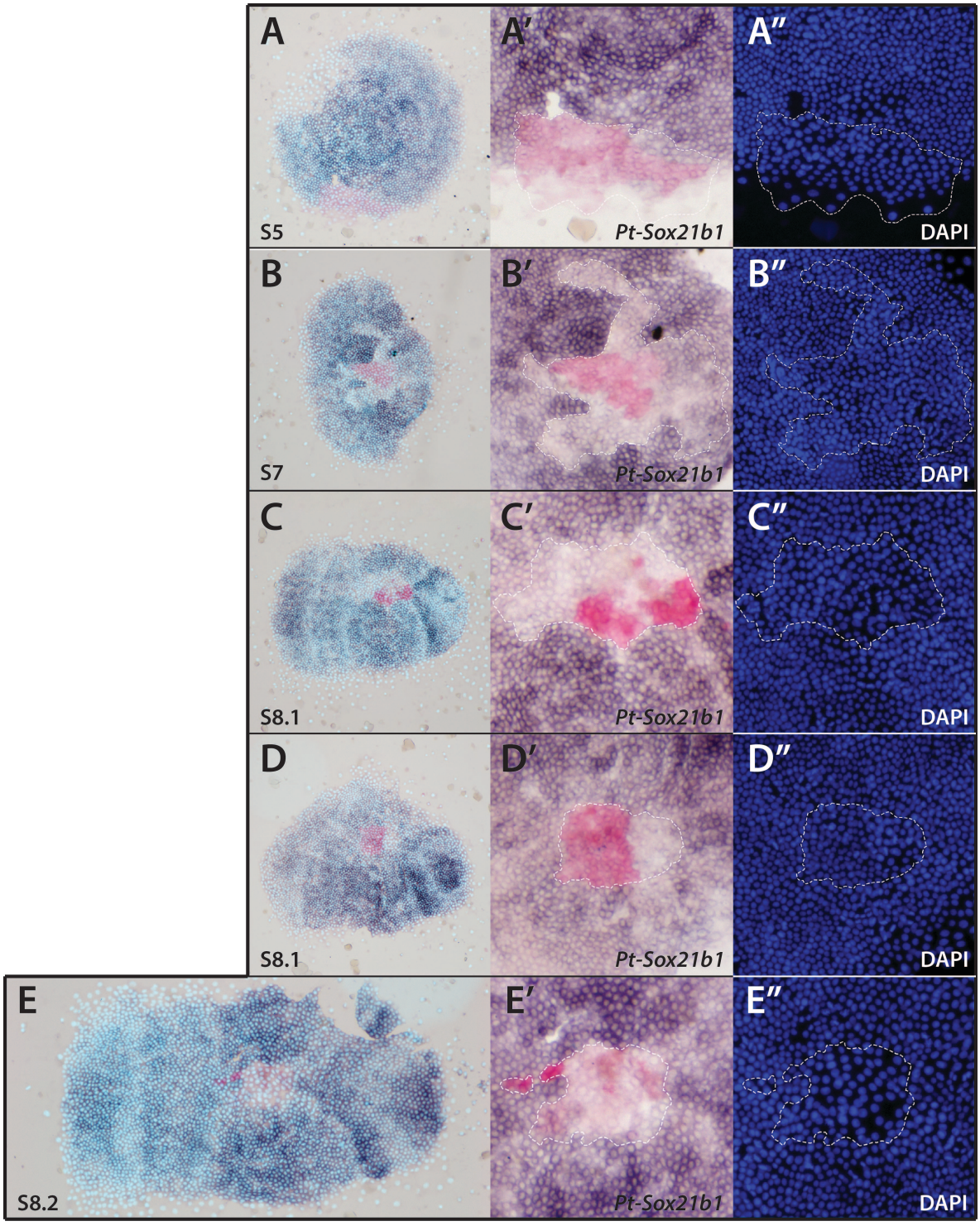

Figure S5: Knockdown clones of *Pt-Sox21b-1* exhibit a decreased number of cells in the affected area. (A-E) Injected embryos with *Pt-Sox21b-1* eRNAi knockdown clones stained for *Pt-Sox21b-1* and counterstained with DAPI. (A'-E") Zoom in of the area affected by *Pt-Sox21b-1* knockdown (white dashed line) showing *Pt-Sox21b-1* staining (A'-E') and DAPI counterstain (A"-E"). Anterior is to the left in all images. Red staining indicates clone location.

Homeobox Gene Duplication and Divergence in Arachnids

Daniel J. Leite,[1] Luís Baudouin-Gonzalez,[1] Sawa Iwasaki-Yokozawa,[2] Jesus Lozano-Fernandez,[3,4] Natascha Turetzek,[5] Yasuko Akiyama-Oda,[2,6] Nikola-Michael Prpic,[†,5] Davide Pisani,[3,4] Hiroki Oda,[2,7] Prashant P. Sharma,[8] and Alistair P. McGregor*[,1]

[1]Department of Biological and Medical Sciences, Oxford Brookes University, Oxford, United Kingdom

[2]JT Biohistory Research Hall, Takatsuki, Osaka, Japan

[3]School of Earth Sciences, University of Bristol, Life Sciences Building, Bristol, United Kingdom

[4]School of Biological Sciences, University of Bristol, Life Sciences Building, Bristol, United Kingdom

[5]Department of Cellular Neurobiology, Johann-Friedrich-Blumenbach Institute for Zoology and Anthropology, Georg-August-University, Göttingen, Germany

[6]Microbiology and Infection Control, Osaka Medical College, Takatsuki, Osaka, Japan

[7]Department of Biological Sciences, Graduate School of Science, Osaka University, Osaka, Japan

[8]Department of Integrative Biology, University of Wisconsin-Madison, Madison, WI, USA

[†]Present address: Department of Biology, Zoology and Developmental Biology, Justus-Liebig-Universität Gieeßn, Gießen, Germany

***Corresponding author:** E-mail: amcgregor@brookes.ac.uk.

Associate editor: Mary J. O'Connell

Abstract

Homeobox genes are key toolkit genes that regulate the development of metazoans and changes in their regulation and copy number have contributed to the evolution of phenotypic diversity. We recently identified a whole genome duplication (WGD) event that occurred in an ancestor of spiders and scorpions (Arachnopulmonata), and that many homeobox genes, including two Hox clusters, appear to have been retained in arachnopulmonates. To better understand the consequences of this ancient WGD and the evolution of arachnid homeobox genes, we have characterized and compared the homeobox repertoires in a range of arachnids. We found that many families and clusters of these genes are duplicated in all studied arachnopulmonates (*Parasteatoda tepidariorum*, *Pholcus phalangioides*, *Centruroides sculpturatus*, and *Mesobuthus martensii*) compared with nonarachnopulmonate arachnids (*Phalangium opilio*, *Neobisium carcinoides*, *Hesperochernes* sp., and *Ixodes scapularis*). To assess divergence in the roles of homeobox ohnologs, we analyzed the expression of *P. tepidariorum* homeobox genes during embryogenesis and found pervasive changes in the level and timing of their expression. Furthermore, we compared the spatial expression of a subset of *P. tepidariorum* ohnologs with their single copy orthologs in *P. opilio* embryos. We found evidence for likely subfunctionlization and neofunctionalization of these genes in the spider. Overall our results show a high level of retention of homeobox genes in spiders and scorpions post-WGD, which is likely to have made a major contribution to their developmental evolution and diversification through pervasive subfunctionlization and neofunctionalization, and paralleling the outcomes of WGD in vertebrates.

Key words: **homeobox genes, development, gene duplication.**

Article

Introduction

Developmental programs precisely orchestrate proliferation and differentiation to build multicellular organisms. Many of the key regulatory factors and pathways utilized in development are conserved between species like the Wnt and Delta/Notch signaling pathways and transcription factors (TF) such as those encoded by the homeobox genes (Reviewed by Carroll *et al.* 2005; Rokas 2008). Many studies in recent decades have shown that changes in the expression and copy number of these tool kit genes can lead to the evolution of phenotypic differences among species (Averof and Patel 1997; Stern 1998; Ronshaugen *et al.* 2002; Carroll *et al.* 2005; Liubicich *et al.* 2009; Werner *et al.* 2010; Guerreiro *et al.* 2013; Koshikawa *et al.* 2015; Kvon *et al.* 2016). Therefore, understanding the evolution of these genes can provide important insights into the development and evolution of metazoans.

The homeobox genes encode a large superclass of TFs (Garcia-Fernandez 2005; Hoegg and Meyer 2005; Pascual-Anaya *et al.* 2012; Holland 2015; Ferrier 2016). They are characterized by encoding a homeodomain, which is usually 60 amino acids in length and folds to form a structure with three α-helices and an N-terminal domain (Ortiz-Lombardia *et al.* 2017). The third α-helix and N-terminal domain confer the specificity to the binding of the homeodomain to the major and minor groove of the DNA double helix,

respectively (Hanes and Brent 1991; Chu *et al.* 2012; Ortiz-Lombardia *et al.* 2017). This conservation of sequence facilitates the characterization of many homeobox genes based solely on their homeodomain sequence (Holland *et al.* 2007), although there are also a variety of other DNA binding domains found in metazoan homeobox genes, which provide additional identification characteristics and biological functions (Burglin and Affolter 2016).

During the evolution of metazoans the expansion of homeobox gene number via duplication has been associated with multicellularity and the increase in morphological complexity (Garcia-Fernandez 2005; Hoegg and Meyer 2005; Pascual-Anaya *et al.* 2012; Holland 2015). The initial multiplication and divergence of proto-homeobox genes started early in evolution and created several classes of homeobox genes (Pascual-Anaya *et al.* 2012; Ferrier 2016). In the urbilaterian, the homeobox genes are hypothesized to have formed a large "Giga-homeobox" cluster, containing several homeobox families (Ferrier 2016). In metazoans, this Giga-cluster also included the metazoan specific ANTP class of homeobox genes (Ferrier 2016). Subsequent tandem duplications of each of the different classes generated clusters of similar homeobox class genes such as the ParaHox, SuperHox, SINE/Six, TALE/Irx, PRD/HRO clusters (Ferrier 2016). These clusters were then fragmented in the genome of the bilaterian ancestor, and have been subject to lineage specific retention, loss, and further duplication during bilaterian evolution (Ferrier 2016).

We recently found that in arachnids there had been a whole genome duplication (WGD) in a common ancestor of arachnopulmonates (spiders, scorpions, and Pedipalpi [Uropygi and Amblypygi]; Sharma, Kaluziak, *et al.* 2014; Schwager *et al.* 2017). Like the independent WGDs in vertebrates, after this event many duplicated homeobox genes have been retained in spiders and scorpions, including two clusters of Hox genes (Lynch *et al.* 2006; Putnam *et al.* 2008; Cao *et al.* 2013; Sharma, Schwager, *et al.* 2014; Di *et al.* 2015; Qu *et al.* 2015; Sharma et al. 2015; Schwager *et al.* 2017). Furthermore, divergence in the expression of ohnologs in spiders, including the Hox genes, suggests there has been neofunctionalization and subfunctionalization of many of these genes since the WGD (Pechmann *et al.* 2015; Turetzek *et al.* 2016, 2017; Schwager *et al.* 2017).

Here, we systematically compare the repertoires of homeobox genes between the arachnopulmonates with an ancestral WGD, the spiders *Parasteatoda tepidariorum* and *Pholcus phalangioides*, and the scorpions *Centruroides sculpturatus* and *Mesobuthus martensii* (Di *et al.* 2015), with arachnids that have no evidence for an ancestral WGD, the harvestman *Phalangium opilio*, the pseudoscorpions *Neobisium carcinoides* and *Hesperochernes* sp., and the tick *Ixodes scapularis*, as well as several mandibulate arthropods. We find pervasive duplication and retention of homeobox genes in arachnopulmonates, and synteny analysis of homeobox genes in *P. tepidariorum* also revealed several more duplicated ancient homeobox clusters (Ferrier 2016), in addition to the Hox clusters. To explore the fate and role of these duplicated genes further we compared the expression profiles of ohnologs during spider embryogenesis and found striking differences in their levels and temporal expression. Furthermore, comparison of the spatial expression of duplicated homeobox genes between *P. tepidariorum* and their single copy homologues in *P. opilio* suggests that there has been extensive neofunctionalization and subfunctionalization in embryogenesis during arachnopulmonate evolution. Taken together, our work shows that WGD greatly expanded the repertoire of homeobox genes in arachnopulmonates and that this contributed to diversification in their developmental gene regulatory networks and may have contributed to evolutionary innovations in these animals as has been postulated in other animal lineages (Van de Peer *et al.* 2009; Huminiecki and Conant 2012).

Results

Comparison of Homeobox Gene Families in Arachnids and Other Arthropods

To systematically identify homeobox repertoires we searched for the characteristic homeodomain sequence in a range of available and new arachnid transcriptomes. In a transcriptome of the spider *P. phalangioides* (Turetzek *et al.*, in preparation), we identified 78 homeobox families (fig. 1 and supplementary file 1, Supplementary Material online), which is similar to the 80 families identified previously in the spider *P. tepidariorum* (Schwager *et al.* 2017) and to the 82 families found in the scorpions *C. sculpturatus* and *M. martensii* (Di *et al.* 2015; Schwager *et al.* 2017).

For lineages that were thought not to have an ancestral WGD, we surveyed existing transcriptomes from the tick *I. scapularis*, the harvestman *P. opilio*, and the pseudoscorpion *Hesperochernes* sp., as well as sequencing a transcriptome for another pseudoscorpion *N. carcinoides*. The number of homeobox families found in *I. scapularis* (70) (fig. 1 and supplementary file 1, Supplementary Material online) was comparable to arachnopulmonates and mandibulates (*S. maritima*—83; *A. mellifera*—77; *T. castaneum*—80; *D. melanogaster*—80) (Zhong *et al.* 2008; Zhong and Holland 2011; Chipman *et al.* 2014). However, we only managed to recover genes from 65 families in *P. opilio* and just 27 and 16 families in *N. carcinoides* and *Hesperochernes* sp., respectively, which likely represent only a subset of families present in these arachnids (fig. 1 and supplementary file 1, Supplementary Material online).

The assignment of homeobox genes into families was verified using a maximum likelihood tree constructed using the homeodomain sequences (supplementary fig. 1, Supplementary Material online). This analysis provided good support for the annotation of each homeodomain to a homeobox gene family, as families were generally monophyletic and had >70% bootstrap support. The general topology of the tree also grouped the homeobox classes together consistent with Holland *et al.* (2007).

Comparisons of the repertoires of homeobox families between these species suggest particular patterns of retention and loss of homeobox families in arthropod lineages (fig. 1). Overall, excluding the harvestman and pseudoscorpion data

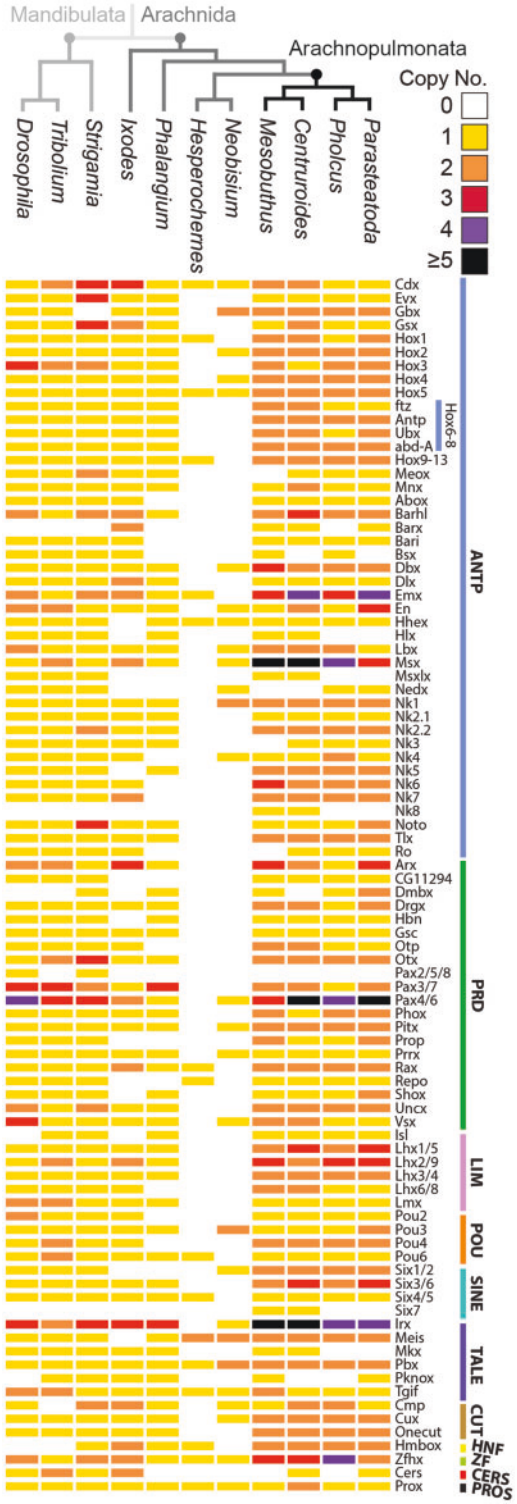

FIG. 1. Comparison of homeobox repertoires in arthropods reveals pervasive duplication in arachnopulmonates. The copy number of homeobox families is generally greater in arachnopulmonates compared with other arthropods across all classes, except Cers and Pros. Homeobox genes are classified based on Holland *et al.* (2007) and the number of paralogs in each family is color coded. The Hox6-8 family has been broken down further to show the specific copy numbers of *ftz*, *Antp*, *Ubx*, and *abdA*.

due to incompleteness, 60 of the known 87 homeobox families were present in all species surveyed, indicating a reasonable retention of most families.

Families that were present in vertebrates, arachnids and the myriapod, but absent in insects were the HNF and Dmbx families (Zhong *et al.* 2008; Zhong and Holland 2011; Chipman *et al.* 2014). Another family that was present in vertebrates and arachnids but missing from the mandibulates surveyed was the Barx family (Zhong *et al.* 2008; Zhong and Holland 2011; Chipman *et al.* 2014). The only family not present in arachnids but present in mandibulates and vertebrates was the Pax2/5/8 family.

There were also some retention/loss differences among arachnid species. While Nedx is present in spiders, it appears to have been lost in the scorpions and *I. scapularis*, although there is a single copy in the pseudoscorpion *N. carcinoides* (fig. 1 and supplementary file 1, Supplementary Material online). The Hlx, Msxlx, and Mkx families also appear to be missing from spiders but present in the scorpions and the mandibulates surveyed (fig. 1).

Pervasive Duplication of Homeobox Genes in Arachnopulmonates

Although the number of homeobox families is fairly similar between arthropod species surveyed, except the harvestman and pseudoscorpions, the actual number of genes varied considerably between arachnopulmonates and non-arachnopulmonate arthropods. The spider *P. phalangioides* had a total of 132 homeobox genes (supplementary file 1, Supplementary Material online), which is comparable to the 145 in *P. tepidariorum* and the 156 found in the scorpions *C. sculpturatus* and *M. martensii* (Di *et al.* 2015; Schwager *et al.* 2017). In contrast, the nonarachnopulmonate species *I. scapularis*, *P. opilio*, *N. carcinoides*, and *Hesperochernes* sp. had 96, 69, 32, and 17 homeobox genes, respectively (supplementary file 1, Supplementary Material online). The most complete nonarachnopulmonate data set represented by *I. scapularis* compared well to the number of homeobox genes previously identified in *S. maritima* (113), *T. castaneum* (105), and *D. melanogaster* (104) (Zhong *et al.* 2008; Zhong and Holland 2011; Chipman *et al.* 2014).

We found that 58%, 50%, 59%, 57% of homeobox families in *P. tepidariorum*, *P. phalangioides*, *C. sculpturatus*, and *M. martensii* are duplicated, compared with 24% in the tick, 3% in the harvestman, 19% in the centipede, beetle, and fly. This shows that many more of the arachnopulmonate homeobox families are comprised of multiple genes copies compared with other arthropods. In total, 34 families are duplicated in all four arachnopulmonate species (fig. 1), which may indicate that these were duplicated in a single event and subsequently retained in the ancestor of the Araneae and Scorpiones lineages. 18 of these 34 families are not duplicated in any of the nonarachnopulmonate species surveyed. Furthermore, 38 families are duplicated in both spiders, whereas 46 families are duplicated in both scorpions (fig. 1).

The families in arachnopulmonates that contain more than two copies, such as Pax4/6 and Irx, are also duplicated in the mandibulate species surveyed. This perhaps suggests

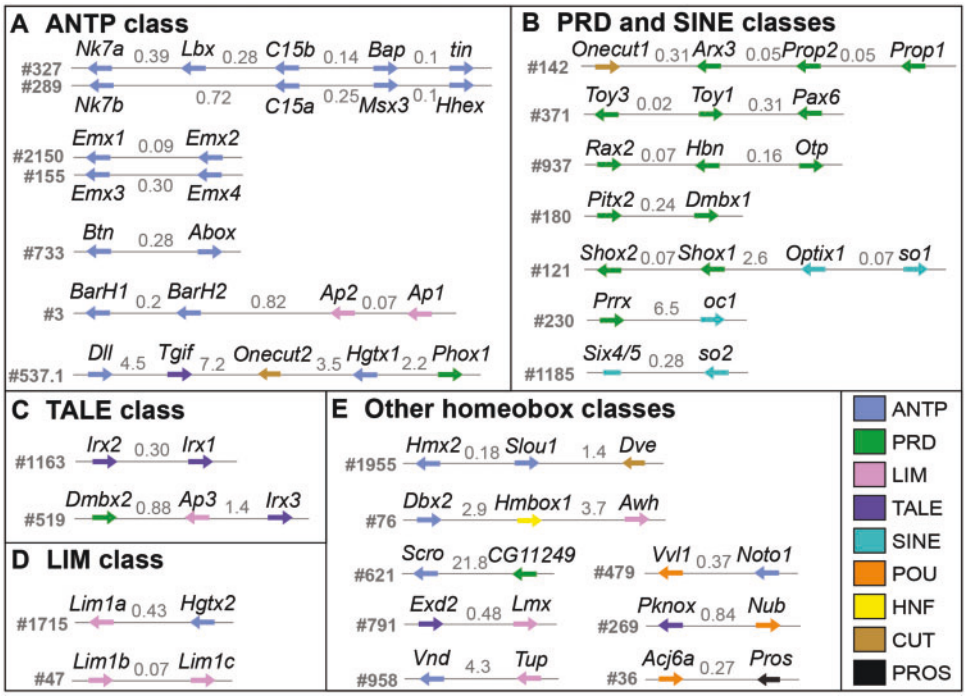

Fig. 2. Homeobox gene clustering in the *Parasteatoda tepidariorum* genome. (*A*) Scaffolds containing at least two ANTP class genes. (*B*) Scaffolds containing PRD and SINE class gene clusters. (*C*) Scaffolds containing the Irx family of the TALE class. (*D*) Scaffolds with Lhx1/5 family of the LIM class. (*E*) Other scaffolds with at least two homeobox genes. All other homeobox genes were localized to individual scaffolds. The intergenic distances are indicated in Mb. *Parasteatoda tepidariorum* DoveTail assembly scaffold numbers are to the left of each cluster. Arrows depict the direction of transcription. Nonhomeobox genes are not shown.

that these were duplicated in the arthropod ancestor and that further paralogs were generated in arachnopulmonates due to the WGD (fig. 1).

Homeobox Gene Ohnologs and Tandem Duplicates in *P. tepidariorum*

It has already been shown that duplicated Hox clusters were retained after the ancestral WGD in arachnopulmonates (Schwager *et al.* 2017). Therefore, we next investigated if other homeobox gene clusters have also been retained. Of the 45 homeobox gene families that are duplicated in *P. tepidariorum*, 40 families are represented by paralogs that are located on different scaffolds, hereafter called dispersed paralogs. Some of these dispersed paralogs are present as duplicated clusters in the genome.

One homeobox cluster that is present across protostomes and deuterostomes is the NK cluster (Garcia-Fernandez 2005; Ferrier 2016). In *P. tepidariorum*, we identified scaffolds that contained duplicated remnants of this cluster. There were two clusters that contained *Nk7* and *Tlx*/*C15* paralogs, which have the same transcriptional orientation on each scaffold (fig. 2*A*). Each of these clusters also contained other ANTP class genes that are usually found in the NK cluster (*Lbx*, *Bap*, *tin*, *Hhex*, and *Msx*). However, of these five genes only *Msx* is duplicated, though the other two *Msx* paralogs are not located in the NK clusters. This indicates differential retention/loss between these duplicate NK clusters in *P. tepidariorum*.

We also identified other clusters of homeobox genes that are duplicated and retained to various extent in *P. tepidariorum*. There is evidence for a duplication of the SINE/Six cluster on scaffolds #121 and #1185 (fig. 2*B*). This cluster, found in both protostomes and deuterostomes, is usually composed of three genes commonly arranged in the order *Optix*, *sine oculis* (*so*), and *Six4*/*5* (Ferrier 2016). On both scaffolds there are *so* genes followed by one paralog of *Optix* on scaffold #121 and the single *Six4*/*5* gene on scaffold #1185. There are also other paralogs of *Optix* in *P. tepidariorum* but they are dispersed in the genome. We also identified clusters of ANTP, TALE, and LIM class genes. There are two scaffolds that each contained two tandem paralogs of *Emx* genes, and these clusters have maintained the same transcriptional orientation (fig. 2*A*). For the TALE class, two *Irx*/*mirr* paralogs were identified on one scaffold and a single copy of *Irx* was present on another scaffold along with *Dmbx2* and *Ap3* (fig. 2*C*). We also identified a scaffold containing two *Lhx1*/*5* paralogs and another with a single copy of *Lhx1*/*5* and one of the *Hgtx* paralogs (fig. 2*D*).

We also found eight homeobox families with tandemly duplicated paralogs: the BarH, Lhx5/9, Pax4/6, Prop, and Shox families as well as the aforementioned mentioned Emx, Irx, and Lhx1/5 families (fig. 2). These tandem duplicates were all found in the same transcriptional orientation apart from the Pax4/6 cluster. This means that of the retained duplicate homeobox families, 50% were found as dispersed paralogs, whereas only 6% have conclusively resulted from tandem duplications. Collectively this implies that there has been a

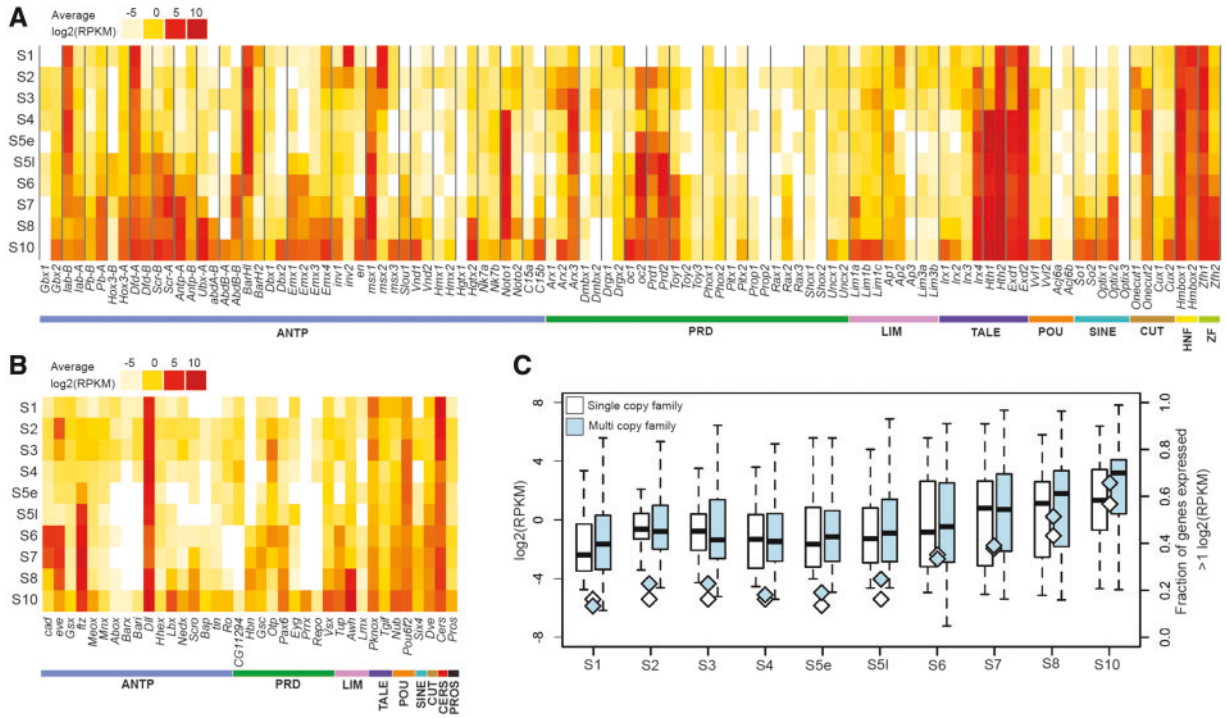

Fig. 3. Expression of homeobox genes in *Parasteatoda tepidariorum* from S1 to S10. The transcriptome profile of *P. tepidariorum* AUGUSTUS gene models for (*A*) duplicated and (*B*) single copy Homeobox genes. (*C*) The average expression of all homeobox genes increases from S1 to S2, likely corresponding to onset of zygotic transcription (Pechmann *et al.* 2017). The numbers of families expressed >1 log2(RPKM) also increase from S1 to S2. The mean expression level is lower and fewer families are expressed around S4/S5e. After which the mean expression level and number of families continues to increase.

greater contribution of WGD than tandem duplication to the expansion of arachnopulmonate homeobox repertoires.

Expression of Homeobox Genes in a *P. tepidariorum* Embryogenesis

We next investigated the expression of homeobox genes in *P. tepidariorum* by quantifying their levels in RNA-Seq data covering the first ten stages of embryogenesis of this spider. All 145 annotated homeobox genes were found to be expressed in at least one of the ten embryonic stages assayed, with the exception of *Slou2* (fig. 3*A* and *B*).

There is an increase in the average expression of single copy and duplicated homeobox genes from S1 to S2 (fig. 3*C*). The number of homeobox genes expressed >1 log2(RPKM) also increases between these first two stages, especially in the case of the multicopy genes. This observation is likely to be explained by the onset of zygotic transcription at S2 (Pechmann *et al.* 2017). After S2 both the average expression level and number of genes expressed decreases to the lowest levels around early S5 after which the number of genes and the average expression also increases (fig. 3*C*).

Interestingly, one homeobox gene that is highly expressed at S1 was *Distal-less* (*Dll*) (fig. 3*B*). This is much earlier than previously reported at S5 (detected by ISH) and its roles in segment specification and limb development (Pechmann *et al.* 2011). Furthermore, expression of *Pt-cad* and *Pt-eve* was also earlier detected at S1 and then increased at S2, again earlier than previously detected using ISH (fig. 3*B*) (Schönauer *et al.* 2016). Therefore, it is possible that *Pt-Dll*, *Pt-cad*, *and Pt-eve* are maternally deposited in this spider and are involved in presently unknown functions during early embryogenesis.

Expression Divergence of Duplicated *P. tepidariorum* Homeobox Genes in the Embryonic Transcriptome

To assess the divergence in the expression of duplicated *P. tepidariorum* homeobox genes, the RNA-Seq profiling was then analyzed to compare the expression levels of dispersed and tandem paralogs during embryogenesis in this spider (fig. 3*A*).

The spatial and temporal expression of Hox paralogs in *P. tepidariorum* was previously analyzed using ISH and showed that Hox genes from both clusters are expressed in the classical collinear fashion across the AP axis (Schwager *et al.* 2017). Interestingly, both the previous ISHs and our RNA-Seq profiling reveal that one paralog of each Hox gene is always expressed earlier than the other, except for the *Pt-abdA* paralogs (Schwager *et al.* 2017). Overall, the timing of Hox expression in the RNA-Seq data matches well with onset of expression detected by ISH (fig. 3*A*). However, both *Pt-lab-A* and *Pt-Dfd-A* were highly expressed from S1 onward, indicating earlier expression than detected by ISH (Pechmann *et al.* 2015; Schwager *et al.* 2017). These results are consistent with previous findings that *P. tepidariorum* Hox paralogs have probably been subject to subfunctionalization and/or neofunctionalization (Pechmann *et al.* 2015; Schwager *et al.* 2017).

Other dispersed paralogs that were present in clusters were the NK class families Nk7 and Tlx/C15 (fig. 2*A*). The

Pt-Nk7 paralogs are both expressed at very low levels throughout most of embryogenesis apart from S10 when they both increase in expression (fig. 3A). The *Pt-C15* paralogs, however, exhibit divergence in their timing and level of expression, with *Pt-C15b* showing increased expression around S7 to S10, compared with *Pt-C15a*, which is barely expressed at any of the ten stages (fig. 3A).

There were also several cases of dispersed (nonclustered) paralogs, which have diverged in the level and timing of their expression (fig. 3A). For example, *Pt-Hth2* is expressed throughout all ten stages, whereas *Pt-Hth1* is only expressed from S4 to S10 and these genes have demonstrably different expression patterning during limb development in this spider (Turetzek *et al.* 2017). Other dispersed paralogs that show aspects of divergence including *Pt-Gbx*, *Pt-Msx*, *Pt-Noto*, *Pt-Arx*, *Pt-Onecut*, *Pt-Hmbox*, and *Pt-Zfh* (fig. 3A), as well as the en/Inv family. *Pt-en* is expressed at S7 in the RNA-Seq data (fig. 3A), which is consistent with ISHs that show expression of *en* starts at early S8 in forming segments in *P. tepidariorum* (Schwager 2008). The *Pt-Inv1* paralog shows similar expression, however *Pt-Inv2* appears to be maternally loaded and down regulated at S2 when zygotic transcription starts (fig. 3A). Therefore, the timing of expression between *Pt-en*/*Pt-Inv* paralogs suggests that they have diverged in function.

A few dispersed paralogs exhibited very similar expression profiles such as *Pt-Pitx*, *Pt-Phox*, and *Pt-Vvl* (fig. 3A). However, it is possible that expression difference may occur later in development or during adult stages and this analysis does not account for any differences in the spatial expression pattern of these genes that may have occurred. This suggests that overall there has been evolutionary changes in the *cis*-regulation of most dispersed paralogs resulting in divergence in expression levels and transcriptional timing between paralogs.

Divergence of Tandem Paralog Expression during *P. tepidariorum* Embryogenesis

Tandem duplicates, like dispersed duplicates, also exhibit both conserved and divergent expression profiles. The Emx family contains four paralogs, of which pairs of paralogs are found on two different scaffolds (fig. 2A). Paralogs *Pt-Emx1* and *Pt-Emx2* have similar expression, which increases from S6 to S10 (fig. 3A). In contrast the other two paralogs, *Pt-Emx3* and *Pt-Emx4*, are both expressed later from S7/S8 to S10 (fig. 3A). There is some early expression of *Pt-Emx4*, however, overall it appears that *Pt-Emx* paralogs that are on the same scaffold have more similar expression profiles.

The Irx family is also represented by four paralogs, two found in tandem (*Pt-Irx1* and *Pt-Irx2*) and two dispersed (*Pt-Irx3* and *Pt-Irx4*) (fig. 2C). The tandem duplicates are both expressed only at S10 (fig. 3A), while *Pt-Irx3* is expressed only at S3 and the *Pt-Irx4* paralog is expressed from S2 to S10 at fairly consistent levels (fig. 3A).

The Lim1/5 family is represented by two paralogs on one scaffold and a third paralog on a separate scaffold (fig. 2D). The two *Pt-Lim1/5* paralogs on the same scaffold had very similar expression, with low levels at S3 but stronger expression at S10 (fig. 3A). In contrast the single *Pt-Lim1/5* paralog on the other scaffold was expressed from S7 to S10 (fig. 3A).

The remaining tandem duplicates, *Pt-BarH*, *Pt-Prop*, and *Pt-Shox*, all showed divergent expression between paralogs (figs. 2A and B and 3A). For example, the *Pt-BarH1* paralog is strongly expressed from S1 to S6, whereas the other paralog appears to be expressed only in S1 and then again at S10 (fig. 3A).

Comparison of Duplicated *P. tepidariorum* Homeobox Gene Expression with Single Copy Orthologs in *P. opilio*

To polarize the expression patterns of duplicated homeobox genes in a phylogenetic context, we analyzed the embryonic expression patterns of a subset of duplicated homeobox gene families in *P. tepidariorum* and compared the expression of selected spider genes to their single copy orthologs in *P. opilio*.

The Msx family provides a likely example of neofunctionalization in the spider (fig. 4A–F). The likely ancestral expression pattern of this gene, possibly represented by *Po-Msx*, is mostly maintained in *Pt-Msx1* (fig. 4A–D). *Pt-Msx2* has probably gained a new expression domain in the chelicerae (fig. 4E). *Pt-Msx3* is also expressed in a conserved pattern at the base of the prosomal appendages (fig. 4F).

While we observed an apparent case of neofunctionalization in the Msx family there were several families (*Emx*, *Irx*, *Pitx*, *Zfh*, and *Cux*) that appear to have undergone subfunctionalization. In the Emx family, the expression pattern of the single copy of *Po-Emx* is subdivided between the four paralogs found in *P. tepidariorum* (fig. 4G–M′). Expression of the tandem paralogs *Pt-Emx1* and *Pt-Emx2* was observed in stripes in the anterior of each opisthosomal segment and *Pt-Emx2* also has expression at the base of prosomal appendages. In contrast, both *Pt-Emx3* and *Pt-Emx4* are expressed in the precheliceral segment and in patches in each segment along the ventral midline, which collectively form a similar expression seen for *Po-Emx* (fig. 4G′ and J–M′). Therefore, expression of *Pt-Emx* paralogs is most similar between the tandem paralogs (fig. 3A) consistent with the RNA-Seq profiles of these genes in *P. tepidariorum*. Nevertheless, some differences are still present between tandem duplicates, mostly in their prosomal appendage domain.

Another likely case of subfunctionalization occurs in the Irx family (fig. 4N–T′). In this family, *Pt-Irx1*, *Pt-Irx2*, and *Pt-Irx4* appear to have subdivided the expression pattern between them compared with *Po-Irx* (fig. 4N–P′). *Pt-Irx1* and *Pt-Irx4* have very similar expression domains, with expression in patches in the precheliceral segment and along the anterior boarder of prosomal and opisthosomal segments (fig. 4Q, Q′, T, and T′). However, *Pt-Irx4* expression extends more laterally in the opisthosomal segments, compared with *Pt-Irx1*. Furthermore, the onset of *Pt-Irx4* expression is earlier and continues until later in embryogenesis compared with *Pt-Irx1*. The other expression domain of *Po-Irx* around the dorsal boundary edge of the germ band is shared with the *Pt-Irx2* paralog. Finally, *Pt-Irx3* has possibly gained a completely new domain in the prosomal appendages of later stages and therefore possibly represents another case of neofunctionalization in *P. tepidariorum* (fig. 4S).

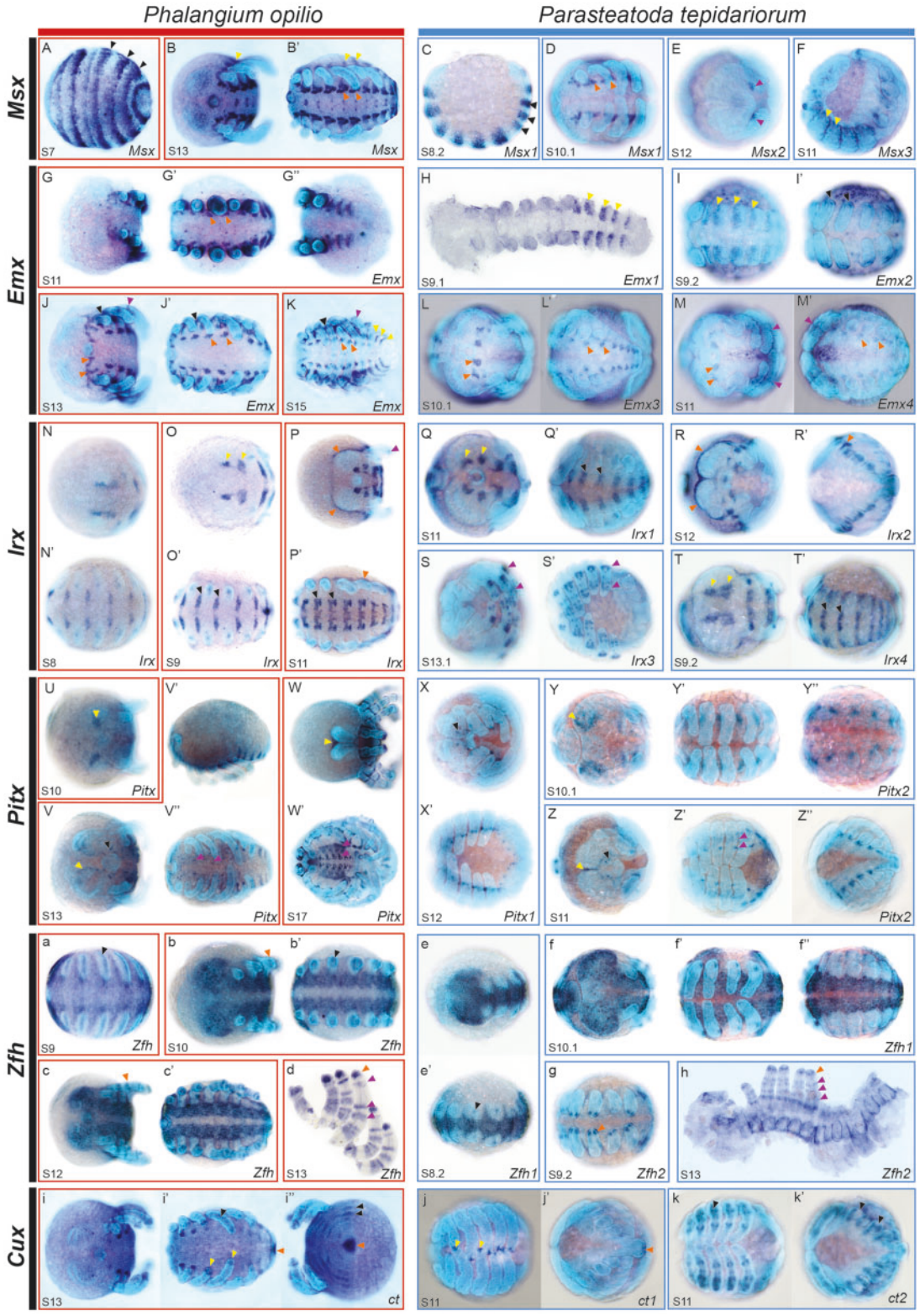

Fig. 4. Expression of *Parasteatoda tepidariorum* paralogs compared with single copy orthologs in *Phalangium opilio*. Expression patterns of *Msx* (*A*–*F*), *Emx* (*G*–*M*′), *Irx* (*N*–*T*′), *Pitx* (*U*–*Z*″), *Zfh* (*a*–*h*), and *Cux* (*i*–*k*′) genes in *P. tepidariorum* (blue boxes) and *P. opilio* (red boxes). The early striped expression of *Po-Msx* (*A*) matches that of *Pt-Msx1* (*C*), indicated by black arrows. The patches of *Po-Msx* expression (*B*′) in each segment along the ventral midline are similar to *Pt-Msx1* (*D*), shown with orange arrows. Expression of *Po-Msx* and *Pt-Msx3* (*B*, *B*′ and *F*) are similar in the region around the base of the appendages (yellow arrows). *Pt-Msx2* has undergone possible neofunctionalization (*E*, purple arrows), with expression in the chelicerae that is not seen for *Po-Msx*. There is similar expression of *Po-Emx* (*K*) in the lateral parts of the opisthosoma compared with *Pt-Emx1* (*H*) and *Pt-Emx2* (*I*), shown with yellow arrows. The expression of *Po-Emx* around the base of the appendages is only seen for *Pt-Emx2* (*I*′), black arrows. The other two *P. tepidariorum* paralogs, *Pt-Emx3* and *Pt-Emx4*, both have expression in the precheliceral region and in patches along the ventral midline, which is also present in *P. opilio* (*G*′ and *J*–*M*′), indicated by orange arrows. The *Po-Emx* expression in the limbs (*J* and *K*) is

For the single copy of *Po-Pitx* (fig. 4*U*–*W*′), expression during embryogenesis was observed in the precheliceral region, which closely resembled the expression of *Pt-Pitx2* in precheliceral region of this spider (fig. 4*Y* and *Z*). Other expression domains of *Po-Pitx* around the stomodeum and the patches along the ventral midline were shared between the *Pt-Pitx* paralogs (fig. 4*X*–*Z*″), indicating possible subfunctionalization of *Pt-Pitx* paralogs.

The *Pt-Zfh* paralogs have also undergone expression divergence (fig. 4*e*–*h*) such that they represent subfunctionalization compared with *Po-Zfh* expression (fig. 4*a*–*d*). *Po-Zfh* expression initially starts along the ventral midline of the germ band with emerging expression within the opisthosomal appendages (fig. 4*a*–*b*′). This expression is mirrored by the *Pt-Zfh1* paralog (fig. 4*e*′), whereas the expression of *Pt-Zfh2* matches that of later *Po-Zfh* expression, with bands of expression in the limbs and faint expression surrounding the coxa and opisthosomal organs (fig. 4*d* and *h*).

Po-ct expression has also probably been subfunctionalized between *Pt-ct* paralogs (fig. 4*i*–*k*′). *Po-ct* is expressed in the tips of the prosomal appendages and at the very posterior of the germ band matching the expression of *Pt-ct1*, while the expression of *Po-ct* in the mesoderm of prosomal appendages, and opisthosoma matches *Pt-ct2* expression (fig. 4*i*–*k*′).

We also found that expression of *Hmx* paralogs in *P. tepidariorum* were highly divergent and again these paralogs perhaps represents an additional example of subfunctionalization (supplementary fig. 2*I*–*K*, Supplementary Material online). *Pt-Hmx1* is mainly expressed in the prosomal appendages while *Pt-Hmx2* is expressed in a pair of cell clusters in the precheliceral region (supplementary fig. 2*I*–*K*, Supplementary Material online).

Loss of embryonic expression was found in three of the ten families analyzed (Gbx, Dbx, and Vnd), where one paralog has retained the likely ancestral pattern as compared with *P. opilio*, while the expression of the other could not be detected during *P. tepidariorum* embryogenesis by ISH (supplementary fig. 2*A*–*H*′, Supplementary Material online). Additionally, in the case of *Pt-Gbx2*, only the prosomal appendage expression observed in *Po-Gbx* is conserved (supplementary fig. 2*C*–*E*, Supplementary Material online), while this gene has also possibly gained a novel expression domain in the opisthosomal limb buds (supplementary fig. 2*F*, Supplementary Material online). It remains possible that the paralogs for which we did not detect expression during embryogenesis are expressed later during juvenile or adult stages.

Discussion

Homeobox Gene Repertoires in Chelicerates

Homeobox genes encode an important group of transcription factors that regulate a wide range of developmental processes (Zagozewski *et al.* 2014; Bataille *et al.* 2015; Du and Taylor 2015; Zuniga 2015; Krumlauf 2016). Consequently they have received substantial attention and are often characterized and compared within and between animal genomes to better understand their evolution and development. Among arthropods, the insects have been sampled the most extensively and robustly, but there has been limited characterization of these genes in other arthropod groups. For example, among the chelicerates, systematic analysis of the homeobox gene repertoires has only been carried out previously for horseshoe crabs and the scorpion *M. martensii* (Di *et al.* 2015; Kenny *et al.* 2015). Therefore, in order to better understand the homeobox repertoires in chelicerates, we surveyed the two spiders *P. tepidariorum* and *P. phalangioides*; another scorpion, *C. sculpturatus*; the pseudoscorpions *N. carcinoides*, *Hesperochernes* sp.; the harvestman *P. opilio*, and the tick *I. scapularis*.

Overall we found a similar complement of homeobox classes and families verifying that chelicerates share and have retained similar homeobox repertoires to other arthropods (fig. 1). However several families were observed that are possibly specific to scorpions (Nk8 and Six7), and the Nedx family in spiders was not found in other arachnids except one of the pseudoscorpions. These particular families may therefore regulate lineage specific features during scorpion and spider development. Furthermore, the Barx family, which is found in chelicerates but not in other arthropods, may coordinate specific aspects of chelicerate development.

Aside from the incomplete data set from the pseudoscorpions and the harvestman, we found the fewest homeobox

Fig. 4. Continued

similar to *Pt-Emx4* (*M* and *M*′), purple arrows. The expression of *Po-Irx* in the precheliceral region (*N*, *O*, and *P*) is seen for *Pt-Irx1* (*Q*) and *Pt-Irx4* (*T*), shown by yellow arrows. These two paralogs also have expression in the opisthosoma (*Q*′ and *T*′), which matches *Po-Irx* (*N*′, *O*′ and *P*′), black arrows. The expression of *Po-Irx* around the germ band (*P* and *P*′) can be seen for *Pt-Irx2* (*R* and *R*′), indicated by orange arrows. There is possibly more elaborate expression of *Pt-Irx3* (*S* and *S*′) in the limbs compared with *Po-Irx* (*P*), shown by purple arrows. *Pt-Pitx1* and *Pt-Pitx2* expression along the ventral midline (*X*′, *Y*′ and *Z*′), which in combination are similar to that seen for *Po-Pitx* (*V*″ and *W*′) (purple arrows). Expression of *Po-Pitx* in the precheliceral anterior furrows (*U*, *V*, and *W*, yellow arrows) is seen for *Pt-Pitx2* (*Y* and *Z*, yellow arrows). However the small dots of *Po-Pitx* expression around the stomodeum (*V*, black arrows) is shared between the two *Pt-Pitx* paralogs (*X* and *Z*, black arrows). Expression of *Po-Zfh* along the ventral midline (*a*–*c*′) is seen for the *Pt-Zfh1* paralog (*e*–*f*″). The later expression of *Po-Zfh* in the distal tips of limbs (*b* and *d*, orange arrows), in bands along the limbs (*d*, purple arrows) and faint expression throughout the embryo is mirrored by *Pt-Zfh2* (*g* and *h*). The expression of *Po-ct* (*i*–*i*″) has clearly subfunctionalized in *P. tepidariorum* with *Pt-ct1* having expression in distal tips of limbs (*j*, yellow arrows) and in the posterior of the germ band (*j*′, orange arrows). The expression of *Pt-ct2* (*k* and *k*′) resembles the striped expression of *Po-ct* in the opisthosoma (*i*″) and in the mesoderm of the appendages (*i*′), indicated by black arrows. All embryos are orientated with the anterior to the left. Images within a box are different views of the same embryo. Images *H* and *h* are flat mounted embryos. One side of the prosoma has been removed in (*h*) to aid flat mounting. Opisthosomal limbs have been dissected from *P. opilio* in (*d*).

families in the tick *I. scapularis* indicating that they have either been lost in this arachnid or there is incomplete sequence information for all families. However, the lineage of parasitiforms, and their putative sister group, the acariforms, also exhibit a greater loss of conserved miRNA families compared with other arachnid lineages (Leite *et al.* 2016). Mite genomes in particular can exhibit marked genome compaction, dynamic rearrangements of homeobox clusters, and associated loss of many transcription factors (Hoy 2009; Grbic *et al.* 2011). Therefore it is likely that there is actual loss of homeobox genes in *I. scapularis*. Interestingly, we also observed long-branch lengths for several tick homeodomains, but it is not known if these functional changes are related to the loss of genes, to rapid evolution of gene function, or to the underlying accelerated rate of evolution inherent to this order (Sharma, Schwager, et al. 2014). Note that while we found only a few families in the two pseudoscorpion species, this likely reflects their representation in the transcriptomes analyzed rather than true losses in this lineage.

Expansion of Homeobox Genes after WGD in the Ancestor of Arachnopulmonates

Previous work identified duplicated homeobox genes in chelicerates (Nossa *et al.* 2014; Di *et al.* 2015; Kenny *et al.* 2015), such as Hox genes in spiders and scorpions (Schwager *et al.* 2007; Sharma, Schwager, *et al.* 2014; Sharma et al. 2015), as well as other homeobox genes involved in spider eye development (Samadi *et al.* 2015; Schomburg *et al.* 2015). However, apart from a scorpion and horseshoe crabs there was no previous systematic analysis of homeobox duplication in chelicerates and in particular how these repertoires have been shaped by WGD in the ancestor of arachnopulmonates.

We found many more duplicated homeobox families in arachnopulmonate species (51–59%) compared with other arthropods surveyed, including *I. scapularis* (24%), *P. opilio* (3%), pseudoscorpions (19% and 6%), and several mandibulates (19%) (fig. 1). Indeed, the proportion of duplicated homeobox families found in *P. tepidariorum* or *C. sculpturatus* is greater than found in either the BUSCO (41%) or OMA (20.5%) data sets (Schwager *et al.* 2017). In fact 18 homeobox families were represented by two paralogs in all four arachnopulmonates but were only single copy in all other arthropods surveyed. This makes up a considerable proportion of the 63–78 duplicates identified in *P. tepidariorum* and *C. sculpturatus* compared with mandibulates and ticks with respect to the BUSCO-Ar database.

It was previously shown that two clusters of Hox genes have been retained in arachnopulmonates following WGD, whereas only one Hox cluster with single copies of most Hox genes is found in *P. opilio*, *I. scapularis*, and *T. urticae* (Sharma *et al.* 2012; Pace *et al.* 2016). Indeed this appears to be a general consequence of WGD: there are two complete and two partial clusters of Hox genes in horseshoe crabs (Nossa *et al.* 2014). In addition, in vertebrate lineages multiple clusters of Hox genes have been produced by several WGD events (Hoegg and Meyer 2005; Mungpakdee *et al.* 2008; Pascual-Anaya *et al.* 2012).

We also found evidence for the duplication of clusters of other homeobox genes in arachnopulmonates in the form of duplicated ANTP (NK cluster), SINE, TALE, and LIM class genes (fig. 2A, C, and D). The inferred ancestral order of arachnopulmonate NK cluster genes (*Nk7*, *Lbx*, *Tlx*, *bap*, *tin*, *Msx*) is consistent with their predicted order in the protostome–deuterostome ancestor (Garcia-Fernandez 2005; Ferrier 2016), requiring just an inversion containing *Lbx* and *Tlx* (fig. 2A). Other ANTP class genes in *P. tepidariorum* are also clustered, which is suggestive of remnants of the mega cluster, however these were not retained as duplicates (fig. 2). A HRO cluster containing *Hbn*, *Rax2*, and *Otp* was also present, and provides further evidence, along with data from *S. maritima*, that this cluster is a feature of arthropods and other protostomes (fig. 2B) (Mazza *et al.* 2010; Chipman *et al.* 2014; Ferrier 2016). However, the order of the three genes in *P. tepidariorum* is different to other arthropods, suggesting that there has been an inversion in the lineage leading to this spider (Mazza *et al.* 2010).

In insects and myriapods, the SINE/Six cluster has degraded and all three genes are dispersed in the genome (Chipman *et al.* 2014; Ferrier 2016). This suggests that the SINE/Six cluster was present in the arthropod ancestor and then has subsequently been degraded in mandibulates but retained in chelicerates. The clusters of ANTP, PRD, SINE, TALE, and LIM class genes in *P. tepidariorum* suggests that spiders have retained many features of the hypothetical clustering of homeobox genes in the bilaterian ancestor (Ferrier 2016). Furthermore, several of these clusters are duplicated and there are different patterns of gene loss/retention and rearrangements, for example, fewer genes have been lost in the Hox cluster compared with the NK cluster.

Retention of gene duplicates in arachnopulmonates has also been observed for other important developmental genes including Wnts and *frizzled4*, and *dachshund* (*dac*), as well as venom and silk genes (Schwager *et al.* 2007, 2017; Janssen *et al.* 2010, 2015; Haney *et al.* 2014, 2016; Clarke *et al.* 2015; Pechmann *et al.* 2015; Samadi *et al.* 2015; Schomburg *et al.* 2015; Turetzek *et al.* 2016, 2017). Furthermore, miRNAs are also pervasively duplicated in arachnopulmonate genomes (Leite *et al.* 2016). This suggests that the retention of duplicated homeobox genes and other developmental toolbox genes after WGD in arachnopulmonates has played an important role in the evolution of development of these animals. The high rate of retention of duplicated homeobox genes after WGD in arachnopulmonates is similar to that observed after the two rounds of WGD in vertebrates (Dehal and Boore 2005; Maere *et al.* 2005; Holland *et al.* 2008; McGrath *et al.* 2014; Schwager *et al.* 2017). Indeed most of the homeobox gene families duplicated in arachnopulmonates are also duplicated in vertebrates, but interestingly the Noto, Drgx, Hmbox families are only duplicated in the former (supplementary fig. 3, Supplementary Material online) (Zhong *et al.* 2008; Zhong and Holland 2011). This indicates that arachnopulmonates and vertebrates have independently retained and utilized duplicated copies of these important transcription factors and this likely contributed to the developmental evolution, novel phenotypes, and

adaptation of these two phyla. Furthermore, families that were only present as single copies in vertebrates and arachnopulmonates were Bsx, Hlx, and Mkx, which indicates that these families fail to retain duplicates in both lineages after WGDs. An intriguing counterpoint for future investigation is therefore horseshoe crabs, which have been shown to have likely undergone two rounds of WGD, but exemplify morphological external stasis and evolutionary relictualism (Sharma, Schwager, *et al.* 2014; Kenny *et al.* 2015; Schwager *et al.* 2017).

Divergence in the Expression of Homeobox Paralogs

How has the ancestral WGD in arachnopulmonates contributed to their evolution and the development of lineage specific features? It has already been shown that one paralog of *dac* in the spider has a distinct and novel role (by comparison to the ancestral function of this gene within Arthropoda), being responsible for patterning the distal boundary of the arachnid-specific podomere, the patella (Turetzek *et al.* 2016). Furthermore, the arrangement of structures in the opisthosoma of scorpions coincides with the staggered expression of paralogous Hox gene expression (Sharma, Schwager, *et al.* 2014), suggesting that divergences in Hox paralogs may in part be responsible for innovations of the scorpion body. Moreover, the Hox paralogs of spiders have also divergences in their temporal and spatial expression (Schwager *et al.* 2007, 2017), while other homeobox paralogs also show differential expression among the developing eyes (Samadi *et al.* 2015; Schomburg *et al.* 2015).

In our study we did not identify any homeobox gene paralogs in *P. tepidariorum* with the same temporal expression profile (fig. 3A), and ISHs on a subset of paralogs also showed divergence in the spatial expression between *P. tepidariorum* paralogs. Comparisons of gene expression of *P. tepidariorum* paralogs with their single copy ortholog in *P. opilio* suggest that most of the surveyed paralogs have likely undergone subfunctionalization, usually in the developing appendages and nervous system. The *P. tepidariorum Msx* genes have also apparently been subject to possible neofunctionalization in the case of *Msx2* in developing chelicerae.

The expression of *Pitx* and *Zfh* genes between *P. opilio* and *P. tepidariorum*, in conjunction with their known expression in *Drosophila*, provide particularly strong evidence for subfunctionalization. In *Drosophila*, *Pitx* is expressed in several tissues including a subset of ventral somatic muscles and in neural cells (Vorbrüggen *et al.* 1997). These expression patterns in *Drosophila* are consistent with *Po-Pitx* expression (fig. 4), suggesting that this expression pattern is ancestral. In *P. tepidariorum*, both *Pitx* paralogs also show metameric patterning along the ventral neuroectoderm, with *Pt-Pitx1* most similar to the *Drosophila* and *P. opilio* CNS expression and *Pt-Pitx2* showing both CNS and mesodermal expression (fig. 4*U*–*Z″*).

The expression patterns of *Zfh* in *Drosophila* and *P. opilio* are also very similar, again implying the ancestral expression of this gene (fig. 4*a*–*d*). Early expression of the *Drosophila Zfh2* ortholog is seen in the brain and ventral CNS (Lai *et al.* 1991), with later expression seen in leg imaginal discs as an initially broad domain at the centre of the disc that develops into rings of expression in each leg segment and in a domain throughout the tarsus (Guarner *et al.* 2014). These patterns are similar to that seen for *P. opilio* (fig. 4*a*–*d*). However in *P. tepidariorum*, early CNS expression and later limb expression has been divided between the *Zfh* paralogs. *Pt-Zfh1* is strongly expressed in the CNS and initial limb buds, while *Pt-Zfh2* is observed in the later pattern expression in rings and at the distal tips of the limbs (fig. 4*e*–*h*). This expression divergence observed for the *Pitx* and *Zfh* paralogs exemplify the pervasive temporal and spatial subfunctionization of genes that has likely occurred in the spider and probably other arachnopulmonates.

Conclusion

Our study has revealed the first comparative genomic picture of the repertoires of homeobox genes in arachnids. This shows that there has been a high level of gene retention of these developmental genes since the WGD in the common ancestor of arachnopulmonates. Furthermore, most of the *P. tepidariorum* homeobox gene paralogs exhibit differences in their timing and spatial expression, and when compared with their single copy homologues in *P. opilio*. This suggests there has been pervasive subfunctionalization and/or neofunctionalization of these genes since WGD. It will be interesting to further investigate the roles of these genes in spider development to ascertain their contribution to the evolution of development and diversification of these arachnids, especially with respect to emergence of novel traits including silk glands and book lungs. Furthermore, future comparisons of ohnologs between arachnopulmonates and vertebrates should provide exciting new insights into the general consequences of WGD in animals.

Materials and Methods

Identification of Homeobox Genes in Arachnids

To identify homeobox genes in arachnid species, we analyzed both existing resources and also new transcriptomic data generated in this study. Existing protein predictions were collected for the tick *Ixodes scapularis* (PRJNA16232), the harvestman *Phalangium opilio* (PRJNA236471), and the pseudoscorpion *Hesperochernes* sp. (PRJNA254752).

For further characterization of homeobox genes in arachnids we also generated de novo transcriptomes for the spider *Pholcus phalangioides* and the pseudoscorpion *Neobisium carcinoides*. For *P. phalangioides* RNA isolation, library preparation and sequencing with Illumina HiSeq2000 was previously described (Janssen *et al.* 2015). A de novo transcriptome assembly (Turetzek N, Torres-Oliva M, Kaufholz F, Prpic NM, Posnien N, in preparation) was performed with Trinity version r20140717 (Haas *et al.* 2013) with the following settings: –seqType fq –JM 240 G – run_as_paired –CPU 6 and using Trimmomatic for quality trimming and filtering (Bolger *et al.* 2014). For the pseudoscorpion *N. carcinoides*, RNA was extracted from the whole body, sequenced with Illumina HiSeqll and de novo assembly of the transcriptome was carried out using Trinity v 2.0.3 (Grabherr *et al.* 2011) under

default parameters and using Trimmomatic for quality control. The raw sequence reads for *P. phalangioides* and the pseudoscorpion *N. carcinoides* have been deposited in the SRA with accession numbers PRJNA448805 and PRJNA438779, respectively.

Longest open reading frames (ORFs) were predicted from the transcriptomes of *P. phalangioides* and the pseudoscorpion *N. carcinoides* as well as from the existing nucleotide transcriptome of the harvestman *P. opilio* (PRJNA236471) and the pseudoscorpion *Hesperochernes* sp. (PRJNA254752) using TransDecoder v3.0.0 (Haas *et al.* 2013). To retain putative proteins the sequence homology and protein domains of predicted ORFs were then analyzed, respectively, with BLASTP v2.2.28+ (e-value $1e^{-6}$) (Altschul *et al.* 1990) using the UniProt Swiss-Prot database (UniProt 2015), and HMMER v3.1 (Wheeler and Eddy 2013) using the Pfam v30.0 database (Finn *et al.* 2016).

The protein sequences from *P. phalangioides*, *I. scapularis*, *P. opilio*, and the two pseudoscorpions were then searched for the presence of homeodomain sequences using BLASTP v2.2.28+ (Altschul *et al.* 1990) with query amino acid homeodomain sequences from all ten species in HomeoDB (Zhong *et al.* 2008; Zhong and Holland 2011) combined with homeodomain sequences from *P. tepidariorum* (Schwager *et al.* 2017), *C. sculpturatus* (Schwager *et al.* 2017), *M. martensii* (Di *et al.* 2015), *Strigamia maritima* (Chipman *et al.* 2014). All the initial BLASTP hits with >30% percentage identity were retained. Next, the full protein sequences of the BLASTP hits were then analyzed using the Conserved Domain Database (CDD) search tool (Marchler-Bauer *et al.* 2015) to confirm the presence of homeodomains as well as annotate other functional domains. BLASTP hits that did not have homeodomains identified by CDD were removed. Transcripts within a species that had identical protein sequences predicted to encode homeodomains were manually checked and identical nucleotide transcripts or isoforms were removed. Specific BLAST searches for PROS class genes also identified a *Pros* gene (MMa30254) in *M. martensii* not reported previously by Di *et al.* (2015). All identified homeobox genes and their sequences are given in supplementary file 1, Supplementary Material online. By concentrating on the detection of homeobox genes based on the presence of homeodomains some partial transcripts of homeobox genes that lack this domain, or may have highly divergent homeodomains, may be missing in our data set.

Phylogenetic Analysis of Arachnid Homeodomains

The predicted homeobox genes were then classified based on phylogenetic analysis of the homeodomain sequences they encode. Amino acid sequences of homeodomains from two spiders (*P. tepidariorum* and *P. phalangioides*), two scorpions (*C. sculpturatus* and *M. martensii*) two pseudoscorpions (*Hesperochernes* sp. and *N. carcinoides*), the harvestman *P. opilio*, the tick *I. scapularis*, the myriapod (centipede) *S. maritima* and three insects *Apis mellifera*, *Tribolium castaneum*, and *Drosophila melanogaster* were aligned with ClustalW (Larkin *et al.* 2007), excluding unusual PROS HPD sequences and the *Cs-Emx1* homeodomain because it has a large insertion.

Phylogenetic analyses, using only unique homeodomain sequence alignments, were performed in RAxML, with support levels estimated using the rapid bootstrap algorithm (1000 replicates) (Stamatakis *et al.* 2008), under the PROTGAMMALG model of amino acid substitution—that was identified as best fitting using a custom Perl script from the Exelixis Lab website (https://sco.h-its.org/exelixis/web/software/raxml/hands_on.html). Homeodomain proteins were classified based on the homology of their homeodomains to known homeodomain containing proteins and annotated with nomenclature following that of Holland *et al.* (2007).

Synteny Analysis of Homeobox Genes in *P. tepidariorum*

To investigate the arrangement of homeobox genes in *P. tepidariorum* we used the high quality HiRise/DoveTail genome assembly (Schwager *et al.* 2017). The scaffold location and coordinates of the previously identified homeobox genes (Schwager *et al.* 2017) were extracted from the GFF file, which contains coordinates of AUGUSTUS gene models relative to the HiRise/DoveTail genome, and were used to calculate the gaps between genes.

Analysis of Homeobox Gene Expression in *P. tepidariorum* Embryogenesis

Homeobox gene expression levels were analyzed during *P. tepidariorum* embryogenesis using RNA sequencing. RNA was extracted using the Dynabeads mRNA DIRECT Kit (Ambion) from 10–100 embryos of each successive developmental stage (stage [S]1–S4, S5 early and S5 late, S6–S8 and S10; Akiyama-Oda and Oda 2003; Mittmann and Wolff 2012). Two replicate sets of mRNAs were independently obtained from two pairs of parents. The mRNAs were fragmented using the NEBNext RNase III RNA Fragmentation Module (New England BioLabs) and then used to construct DNA libraries with the NEBNext Ultra Directional RNA Library Prep Kit for Illumina (New England BioLabs) and NEBNext Multiplex Oligos for Illumina (Index Primers Set 1, New England Biolabs). The libraries were sequenced using the 150-cycle format of the Illumina MiSeq Reagent Kit v3. The resulting sequence reads were subjected to adaptor trimming using the CLC Genomics Workbench 7.0.3 (Qiagen), and quality of the sequences was confirmed with FastQC v0.11.2 (Andrews 2011). The trimmed raw reads have been deposited in the SRA with accession number PRJNA448775 (Iwasaki-Yokozawa *et al.* 2018). Replicates for each stage were aligned to the *P. tepidariorum* reference transcriptome (Schwager *et al.* 2017) using TopHat v2 (Kim *et al.* 2013). Outputs files were sorted and indexed with Samtools v1.2 (Li *et al.* 2009) and RPKM expression levels were quantified using HTSeq-count (Anders *et al.* 2015) and custom Perl scripts. Heatmaps were generated in R v3.2.3 (R Core Team 2015) using the ComplexHeatmap package (Gu *et al.* 2016).

Parasteatoda tepidariorum and *P. opilio* Cultures

An inbred culture of *P. tepidariorum* (from a strain collected in Göttingen, Germany) was maintained at Oxford Brookes University and fed on a diet of *Drosophila vestigial* mutants and *Gryllodes sigillatus*, with a 12:12 h light:dark cycle at 25°C. The culture of *P. opilio* was maintained at the University of Wisconsin, Madison, WI and fed on a diet of fish flakes supplemented with *Acheta domesticus* nymphs, with a 14: 10 light: dark cycle at 20°C.

Cloning of Gene Fragments and Probe Synthesis

cDNA was generated using QuantiTech (Qiagen) with RNA extracted (Qiazol) from S1 to S14 *P. tepidariorum* embryos and from S7 to S17 for *P. opilio*. Gene fragments were amplified by PCR and cloned into the TOPO-TA vector (ThermoFisher Scientific). Primer sequences are provided in supplementary table 1, Supplementary Material online. RNA probes were transcribed with T3 (11031163001—Roche) or T7 polymerase (10881775001—Roche), with DIG RNA labeling mix (11277073910—Roche), from PCR fragments generated from TOPO-TA clones following standard protocols.

In Situ Hybridization in *P. tepidariorum* and *P. opilio*

Colourmetric in situ hybridization (ISH) for *P. tepidariorum* and *P. opilio* was performed as previously described (Akiyama-Oda and Oda 2003; Sharma *et al.* 2012). Embryos were counterstained with DAPI (Roche—10236276001) for ~20 mins to visualize nuclei. Embryos were imaged using a Zeiss Axio Zoom V.16 and a Nikon SMZ25, and overlays were generated in Photoshop CS6.

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

This work was supported by a Leverhulme Trust grant (RPG-2016-234) to A.P.M., Nigel Groome studentships to D.J.L. and L.B.G., a Marie Skłodowska-Curie Fellowship for J.L.-F. (655814), and the Japan Society for the Promotion of Science (JSPS) Grants-in-Aid for Scientific Research (KAKENHI) awards to H.O. (15K07139) and Y.A.-O. (26440130). P.P.S. was supported by National Science Foundation grant IOS-1552610. N.T. and N.M.P. were supported by grants from the Deutsche Forschungsgemeinschaft grant numbers PR 1109/4-1 and PR 1109/6-1.

References

Akiyama-Oda Y, Oda H. 2003. Early patterning of the spider embryo: a cluster of mesenchymal cells at the cumulus produces *Dpp* signals received by germ disc epithelial cells. *Development* 130(9):1735–1747.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol*. 215(3):403–410.

Anders S, Pyl PT, Huber W. 2015. HTSeq–a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31(2):166–169.

Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. Available from http://www.bioinformatics.babraham.ac.uk/projects/fastqc. Last accessed January 25, 2015.

Averof M, Patel NH. 1997. Crustecean appendage evolution associated with changes in Hox gene expression. *Nature* 388(6643):682–686.

Bataille L, Frendo JL, Vincent A. 2015. Hox control of *Drosophila* larval anatomy; the alary and thoracic alary-related muscles. *Mech Dev*. *138*(*Pt* 2):170–176.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30(15):2114–2120.

Burglin TR, Affolter M. 2016. Homeodomain proteins: an update. *Chromosoma* 125(3):497–521.

Cao Z, Yu Y, Wu Y, Hao P, Di Z, He Y, Chen Z, Yang W, Shen Z, He X. 2013. The genome of *Mesobuthus martensii* reveals a unique adaptation model of arthropods. *Nat Commun*. 4:2602.

Carroll SB, Grenier JK, Weatherbee SD. 2005. From DNA to diversity. Oxford: Blackwell Publishing.

Chipman AD, Ferrier DE, Brena C, Qu J, Hughes DS, Schroder R, Torres-Oliva M, Znassi N, Jiang H, Almeida FC, et al. 2014. The first myriapod genome sequence reveals conservative arthropod gene content and genome organisation in the centipede *Strigamia maritima*. *PLoS Biol*. 12(11):e1002005.

Chu SW, Noyes MB, Christensen RG, Pierce BG, Zhu LJ, Weng Z, Stormo GD, Wolfe SA. 2012. Exploring the DNA-recognition potential of homeodomains. *Genome Res*. 22(10):1889–1898.

Clarke TH, Garb JE, Hayashi CY, Arensburger P, Ayoub NA. 2015. Spider transcriptomes identify ancient large-scale gene duplication event potentially important in silk gland evolution. *Genome Biol Evol*. 7(7):1856–1870.

Dehal P, Boore JL. 2005. Two rounds of whole genome duplication in the ancestral vertebrate. *PLoS Biol*. 3(10):e314.

Di Z, Yu Y, Wu Y, Hao P, He Y, Zhao H, Li Y, Zhao G, Li X, Li W, et al. 2015. Genome-wide analysis of homeobox genes from *Mesobuthus martensii* reveals Hox gene duplication in scorpions. *Insect Biochem Mol Biol*. 61:25–33.

Du H, Taylor HS. 2015. The role of Hox genes in female reproductive tract development, adult function, and fertility. *Cold Spring Harb Perspect Med*. 6(1):a023002.

Ferrier DEK. 2016. Evolution of homeobox gene clusters in animals: the Giga-cluster and primary vs. secondary clustering. *Front Ecol Evol*. 4:1–13.

Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, et al. 2016. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res*. 44(D1):D279–D285.

Garcia-Fernandez J. 2005. The genesis and evolution of homeobox gene clusters. *Nat Rev Genet*. 6(12):881–892.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol*. 29(7):644–652.

Grbic M, Van Leeuwen T, Clark RM, Rombauts S, Rouze P, Grbic V, Osborne EJ, Dermauw W, Ngoc PC, Ortego F, et al. 2011. The genome of *Tetranychus urticae* reveals herbivorous pest adaptations. *Nature* 479(7374):487–492.

Gu Z, Eils R, Schlesner M. 2016. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* 32(18):2847–2849.

Guarner A, Manjon C, Edwards K, Steller H, Suzanne M, Sanchez-Herrero E. 2014. The *zinc finger homeodomain-2* gene of *Drosophila* controls *Notch* targets and regulates apoptosis in the tarsal segments. *Dev Biol*. 385(2):350–365.

Guerreiro I, Nunes A, Woltering JM, Casaca A, Nóvoa A, Vinagre T, Hunter ME, Duboule D, Mallo M. 2013. Role of a polymorphism in a Hox/Pax-responsive enhancer in the evolution of the vertebrate spine. *Proc Natl Acad Sci U S A*. 110(26):10682–10686.

Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, et al. 2013. *De novo* transcript

sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc*. 8(8):1494–1512.
Hanes SD, Brent R. 1991. A genetic model for interaction of the homeodomain recognition helix with DNA. *Science* 251(4992):426–430.
Haney RA, Ayoub NA, Clarke TH, Hayashi CY, Garb JE. 2014. Dramatic expansion of the black widow toxin arsenal uncovered by multitissue transcriptomics and venom proteomics. *BMC Genomics* 366(15):1–18.
Haney RA, Clarke TH, Gadgil R, Fitzpatrick R, Hayashi CY, Ayoub NA, Garb JE. 2016. Effects of gene duplication, positive selection and shifts in gene expression on the evolution of the venom gland transcriptome in widow spiders. *Genome Biol Evol*. 8(1):228–242.
Hoegg S, Meyer A. 2005. Hox clusters as models for vertebrate genome evolution. *Trends Genet*. 21(8):421–424.
Holland PW. 2015. Did homeobox gene duplications contribute to the Cambrian explosion? *Zool Lett*. 1:1.
Holland LZ, Albalat R, Azumi K, Benito-Gutierrez E, Blow MJ, Bronner-Fraser M, Brunet F, Butts T, Candiani S, Dishaw LJ, et al. 2008. The amphioxus genome illuminates vertebrate origins and cephalochordate biology. *Genome Res*. 18(7):1100–1111.
Holland PW, Booth HA, Bruford EA. 2007. Classification and nomenclature of all human homeobox genes. *BMC Biol*. 5:47.
Hoy MA. 2009. The predatory mite *Metaseiulus occidentalis*: mitey small and mitey large genomes. *BioEssays* 31(5):581–590.
Huminiecki L, Conant GC. 2012. Polyploidy and the evolution of complex traits. *Int J Evol Biol*. 2012:292068.
Iwasaki-Yokozawa S, Akiyama-Oda Y, Oda H. 2018. Genome-scale embryonic developmental profile of gene expression in the common house spider *Parasteatoda tepidariorum*. *Data Brief*. 19:865–867.
Janssen R, Le Gouar M, Pechmann M, Poulin F, Bolognesi R, Schwager EE, Hopfen C, Colbourne JK, Budd GE, Brown SJ, et al. 2010. Conservation, loss, and redeployment of Wnt ligands in protostomes: implications for understanding the evolution of segment formation. *BMC Evol Biol*. 10(372):1–21.
Janssen R, Schönauer A, Weber M, Turetzek N, Hogvall M, Goss GE, Patel NH, McGregor AP, Hilbrant M. 2015. The evolution and expression of panarthropod *frizzled* genes. *Front Ecol Evol*. 3:1–14.
Kenny NJ, Chan KW, Nong W, Qu Z, Maeso I, Yip HY, Chan TF, Kwan HS, Holland PW, Chu KH, et al. 2015. Ancestral whole-genome duplication in the marine chelicerate horseshoe crabs. *Heredity*. 116(2):190–199.
Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol*. 14(4):R36.
Koshikawa S, Giorgianni MW, Vaccaro K, Kassner VA, Yoder JH, Werner T, Carroll SB. 2015. Gain of *cis*-regulatory activities underlies novel domains of *wingless* gene expression in *Drosophila*. *Proc Natl Acad Sci U S A*. 112(24):7524–7529.
Krumlauf R. 2016. Hox genes and the hindbrain: a study in segments. *Curr Top Dev Biol*. 116:581–596.
Kvon EZ, Kamneva OK, Melo US, Barozzi I, Osterwalder M, Mannion BJ, Tissieres V, Pickle CS, Plajzer-Frick I, Lee EA, et al. 2016. Progressive loss of function in a limb enhancer during snake evolution. *Cell* 167(3):633–642 e11.
Lai Z, Fortini ME, Rubin GM. 1991. The embryonic expression patterns of *zfh-1* and *zfh-2*, two *Drosophila* genes encoding novel zinc-finger homeodomain proteins. *Mech Dev*. 34(2–3):123–134.
Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, et al. 2007. ClustalW and ClustalX version 2. *Bioinformatics* 23(21):2947–2948.
Leite DJ, Ninova M, Hilbrant M, Arif S, Griffiths-Jones S, Ronshaugen M, McGregor AP. 2016. Pervasive microRNA duplication in chelicerates: insights from the embryonic microRNA repertoire of the spider *Parasteatoda tepidariorum*. *Genome Biol Evol*. 8(7):2133–2144.
Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25(16):2078–2079.
Liubicich DM, Serano JM, Pavlopoulos A, Kontarakis Z, Protas ME, Kwan E, Chatterjee S, Tran KD, Averof M, Patel NH. 2009. Knockdown of *Parhyale Ultrabithorax* recapitulates evolutionary changes in crustacean appendage morphology. *Proc Natl Acad Sci U S A*. 106(33):13892–13896.
Lynch VJ, Roth JJ, Wagner GP. 2006. Adaptive evolution of Hox-gene homeodomains after cluster duplications. *BMC Evol Biol*. 6:86.
Maere S, De Bodt S, Raes J, Casneuf T, Van Montagu M, Kuiper M, Van de Peer Y. 2005. Modeling gene and genome duplications in eukaryotes. *Proc Natl Acad Sci U S A*. 102(15):5454–5459.
Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY, Geer RC, He J, Gwadz M, Hurwitz DI. 2015. CDD: nCBI's conserved domain database. *Nucleic Acids Res*. 43(Database issue):D222–D226.
Mazza ME, Pang K, Reitzel AM, Martindale MQ, Finnerty JR. 2010. A conserved cluster of three PRD-class homeobox genes (*homeobrain*, *rx* and *orthopedia*) in the Cnidaria and Protostomia. *Evodevo* 1(1):3.
McGrath CL, Gout JF, Johri P, Doak TG, Lynch M. 2014. Differential retention and divergent resolution of duplicate genes following whole-genome duplication. *Genome Res*. 24(10):1665–1675.
Mittmann B, Wolff C. 2012. Embryonic development and staging of the cobweb spider *Parasteatoda tepidariorum* C. L. Koch, 1841 (syn.: *achaearanea tepidariorum*; Araneomorphae; Theridiidae). *Dev Genes Evol*. 222(4):189–216.
Mungpakdee S, Seo HC, Angotzi AR, Dong X, Akalin A, Chourrout D. 2008. Differential evolution of the 13 Atlantic salmon Hox clusters. *Mol Biol Evol*. 25(7):1333–1343.
Nossa CW, Havlak P, Yue JX, Vincent KY, Brockmann J, Putman NH. 2014. Joint assembly and genetic mapping of the Atlantic horshoe crab genome reveals ancient whole genome duplication. *Giga Sci*. 9(3):1–21.
Ortiz-Lombardia M, Foos N, Maurel-Zaffran C, Saurin AJ, Graba Y. 2017. Hox functional diversity: novel insights from flexible motif folding and plastic protein interaction. *Bioessays* 39(4):1600246.
Pace RM, Grbic M, Nagy LM. 2016. Composition and genomic organization of arthropod Hox clusters. *Evodevo* 7:11.
Pascual-Anaya J, D'Aniello S, Kuratani S, Garcia-Fernàndez J. 2012. Evolution of Hox gene clusters in deuterostomes. *BMC Dev Biol*. 13(26):1–14.
Pechmann M, Benton MA, Kenny NJ, Posnien N, Roth S. 2017. A novel role for *Ets4* in axis specification and cell migration in the spider *Parasteatoda tepidariorum*. *Elife* 6(e27590):1–20.
Pechmann M, Khadjeh S, Turetzek N, McGregor AP, Damen WG, Prpic NM. 2011. Novel function of *Distal-less* as a gap gene during spider segmentation. *PLoS Genet*. 7(10):e1002342.
Pechmann M, Schwager EE, Turetzek N, Prpic NM. 2015. Regressive evolution of the arthropod tritocerebral segment linked to functional divergence of the Hox gene *labial*. *Proc Biol Sci*. 282(1814):1–6.
Putnam NH, Butts T, Ferrier DE, Furlong RF, Hellsten U, Kawashima T, Robinson-Rechavi M, Shoguchi E, Terry A, Yu JK, et al. 2008. The amphioxus genome and the evolution of the chordate karyotype. *Nature* 453(7198):1064–1071.
Qu Z, Kenny NJ, Lam HM, Chan TF, Chu KH, Bendena WG, Tobe SS, Hui JH. 2015. How did arthropod sesquiterpenoids and ecdysteroids arise? Comparison of hormonal pathway genes in noninsect arthropod genomes. *Genome Biol Evol*. 7(7):1951–1959.
R Core Team 2015. R: a language and environment for statistical computing. Vienna (Austria): R Foundation for Statistical Computing.
Rokas A. 2008. The origins of multicellularity and the early history of the genetic toolkit for animal development. *Annu Rev Genet*. 42:235–251.
Ronshaugen M, McGinnis N, McGinnis W. 2002. Hox protein mutation and macroevolution of the insect body plan. *Nature* 415(6874):914–917.
Samadi L, Schmid A, Eriksson BJ. 2015. Differential expression of retinal determination genes in the principal and secondary eyes of *Cupiennius salei* Keyserling (1877). *Evodevo* 6:16.

Schomburg C, Turetzek N, Schacht MI, Schneider J, Kirfel P, Prpic NM, Posnien N. 2015. Molecular characterization and embryonic origin of the eyes in the common house spider *Parasteatoda tepidariorum*. *Evodevo* 6:15.

Schönauer A, Paese CL, Hilbrant M, Leite DJ, Schwager EE, Feitosa NM, Eibner C, Damen WG, McGregor AP. 2016. The Wnt and Delta-Notch signalling pathways interact to direct pair-rule gene expression via *caudal* during segment addition in the spider *Parasteatoda tepidariorum*. *Development* 143(13):2455–2463.

Schwager EE. 2008. Segmentation of the spider Achaearanea tepidariorum investigated by gene expression and functional analysis of the gap gene hunchback. PhD thesis, University of Cologne, Germany.

Schwager EE, Schoppmeier M, Pechmann M, Damen WG. 2007. Duplicated Hox genes in the spider *Cupiennius salei*. *Front Zool*. 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(1):1–27.

Sharma PP, Kaluziak ST, Perez-Porro AR, Gonzalez VL, Hormiga G, Wheeler WC, Giribet G. 2014. Phylogenomic interrogation of arachnida reveals systemic conflicts in phylogenetic signal. *Mol Biol Evol*. 31(11):2963–2984.

Sharma PP, Santiago MA, Gonzalez-Santillan E, Monod L, Wheeler WC. 2015. Evidence of duplicated Hox genes in the most recent common ancestor of extant scorpions. *Evol Dev*. 17(6):347–355.

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(5):450–463.

Sharma PP, Schwager EE, Extavour CG, Wheeler WC. 2014. Hox gene duplications correlate with posterior heteronomy in scorpions. *Proc Biol Sci*. 281(1792):20140661.

Stamatakis A, Hoover P, Rougemont J. 2008. A rapid bootstrap algorithm for the RAxML Web servers. *Syst Biol*. 57(5):758–771.

Stern DL. 1998. A role of *Ultrabithorax* in morphological differences between *Drosophila* species. *Nature* 396(6710):463–466.

Turetzek N, Khadjeh S, Schomburg C, Prpic NM. 2017. Rapid diversification of *homothorax* expression patterns after gene duplication in spiders. *BMC Evol Biol*. 17(1):168.

Turetzek N, Pechmann M, Schomburg C, Schneider J, Prpic NM. 2016. Neofunctionalisation of a duplicate *dachshund* gene underlies the evolution of a novel leg segment in arachnids. *Mol Biol Evol*. 33(1):109–121.

UniProt C. 2015. UniProt: a hub for protein information. *Nucleic Acids Res*. 43(Database issue):D204–D212.

Van de Peer Y, Maere S, Meyer A. 2009. The evolutionary significance of ancient genome duplications. *Nat Rev Genet*. 10(10):725–732.

Vorbrüggen G, Constien R, Zilian O, Wimmer EA, Dowe G, Taubert H, Noll M, Jäckle H. 1997. Embryonic expression and characterization of a *Ptx1* homolog in *Drosophila*. *Mech Dev*. 68(1–2):139–147.

Werner T, Koshikawa S, Williams TM, Carroll SB. 2010. Generation of a novel wing colour pattern by the *Wingless* morphogen. *Nature* 464(7292):1143–1148.

Wheeler TJ, Eddy SR. 2013. nhmmer: dNA homology search with profile HMMs. *Bioinformatics* 29(19):2487–2489.

Zagozewski JL, Zhang Q, Pinto VI, Wigle JT, Eisenstat DD. 2014. The role of homeobox genes in retinal development and disease. *Dev Biol*. 393(2):195–208.

Zhong Y-F, Butts T, Holland PWH. 2008. HomeoDB: a database of homeobox gene diversity. *Evol Dev*. 10(5):516–518.

Zhong Y-F, Holland PWH. 2011. HomeoDB2: functional expansion of a comparative homeobox gene database for evolutionary developmental biology. *Evol Dev*. 13(6):567–568.

Zuniga A. 2015. Next generation limb development and evolution: old questions, new perspectives. *Development* 142(22):3810–3820.