

Head and protocerebral patterning in the spider *Parasteatoda tepidariorum*

Amber Harper

Thesis submitted in partial fulfilment of the requirements of Oxford Brookes University for the
degree of Doctor of Philosophy.

June 2024

Abstract

Chelicerates (including spiders, scorpions, horseshoe crabs, mites and ticks) diverged from other mandibulate arthropods (insects, crustaceans, centipedes and millipedes) around 500 million years ago, making them good models to study whether traits and mechanisms of development are ancestral or derived with respect to the arthropod common ancestor. An arachnoplumonate-arachnid, the spider *Parasteatoda tepidarium* has emerged as one of the main chelicerate model species for evolutionary and developmental biology. The arthropod protocerebrum is derived from the neurogenetic ectoderm of the pre-antennal/chelicerar region. It gives rise to centres or neuropils, which in insects are essential for spatial orientation and object recognition (central complex) and olfactory learning and memory (mushroom bodies/MB). However, it is not clear whether the chelicerate arcuate body and spider MB are homologous to insect central complex and MB, respectively. The overall aim of this thesis is the homology of arthropod brain regions through a comparison of gene expression patterns during spider embryogenesis. To understand the regulation of the development of the spider pre-chelicerar region, chelicerar and pedipalp segments of the spider embryo, I first analysed marker genes in five putative anterior cell clusters from scRNA-seq analysis at three embryonic stages of *P. tepidariorum* (Leite *et al.*, 2024). I found two clusters corresponded to the pedipalpal segment, two clusters corresponded to the pre-chelicerar region: a *six3*-expressing anterior and *otd2*-expressing posterior cluster and one to the *hh1*-expressing pre-chelicerar/chelicerar boundary. This work provided new insights into pre-chelicerar patterning and supports conclusions made by Posnien *et al.* (2023), such as the location of the vertebrate mid- and hindbrain boundary homolog. I then characterised the development of neuroendocrine cell populations in the *six3*-expression region by analysing the expression of neuroendocrine marker genes from insects and centipedes during spider embryogenesis. I found that spiders have cell populations corresponding to the insect neuroendocrine centres the *pars intercerebralis* and *pars lateralis*, the first to be identified in a chelicerate and placing their origin to the arthropod common ancestor. This work further supports the homology of spiders and insects at the level of anterior neuroectoderm patterning and homology of their neuropils. Finally, I analysed the retention of ohnologs in key developmental genes (Hox, Wnt, Frizzled, and head patterning genes) following whole-genome duplication in arachnoplumonates. The improved taxonomic sampling provided further evidence for the ancestral arachnoplumonate whole genome duplication and allowed me to identify likely patterns of ohnolog retention and loss. Furthermore, I identified the first reported duplicates of *Wnt1/wg* in any animal and the first *Wnt10* in any arachnid, and the first *vax* in a spider.

Publications relevant to thesis

Publications

Leite, D. J., Schönauer, A., Blakeley, G., Harper, A., *et al.* (2024). An atlas of spider development at single-cell resolution provides new insights into arthropod embryogenesis. *Evodevo*, 15(1):5.

Baudouin Gonzalez, L., Schoenauer, A., Harper, A., *et al.* (2024). Development and patterning of a highly versatile visual system in spiders. *bioRxiv*:2023.12.22.572789.

Harper, A., Baudouin Gonzalez, L., Schonauer, A., *et al.* (2021). Widespread retention of ohnologs in key developmental gene families following whole-genome duplication in arachnopulmonates. *G3 (Bethesda)*, 11(12).

Publication associated with this thesis

Baudouin-Gonzalez, L., Schoenauer, A., Harper, A., *et al.* (2021). The Evolution of Sox Gene Repertoires and Regulation of Segmentation in Arachnids. *Molecular Biology and Evolution*, 38(8):3153-3169.

Baudouin-Gonzalez, L., Harper, A., McGregor, A. P., *et al.* (2022). Regulation of Eye Determination and Regionalization in the Spider *Parasteatoda tepidariorum*. *Cells*, 11(4).

Mulhair, P. O., Crowley, L., Boyes, D. H., Harper, A., *et al.* (2023). Diversity, duplication, and genomic organization of homeobox genes in Lepidoptera. *Genome Res*, 33(1):32-44.

Acknowledgements

To Alistair McGregor, thank you for encouraging me to pursue a PhD and for giving me this incredible opportunity. Your continued reassurance and support have meant so much. To Amber and Grace, the incredible women with whom I started my PhD, I feel so lucky to call you lifelong friends. Nothing bonds you like doing a PhD during a global pandemic, sharing a lab bench for three years, and spending hours each week feeding spiders. To Emily and Becky, the best housemates I could have asked for, thank you for some of my favourite evenings spent in Oxford whether it was deep or silly chats on the sofa, sunsets at the lake, or Tennis Tuesdays. To my partner Sam for his continued encouragement and support. I could not have completed this journey without you, and I look forward to the next chapter of our lives together. Thank you to my friends and family for always cheering me on, but most importantly, my parents, for teaching me that Harper's never give up.

Table of Contents

1. Chapter 1: General Introduction	13
1.1 The arthropod head.....	13
1.1.1 Development of the pre-antennal, antennal, and intercalary segments in <i>D. melanogaster</i>	14
1.2 Genes involved in arthropod head patterning.....	15
1.2.1 <i>otx/otd</i>	16
1.2.2 <i>emx/ems</i>	16
1.2.3 <i>Sp5/btd</i>	16
1.2.4 <i>col</i> and <i>cnc</i>	17
1.2.5 <i>hh</i> stripe-splitting.....	17
1.3 Arthropod visual system.....	18
1.3.1 Spider vision.....	19
1.3.2 Eye development in <i>P. tepidariorum</i>	19
1.4 Arthropod midline neuropils, the insect central complex and chelicerate arcuate body.....	21
1.5 Mushroom bodies (MB).....	22
1.6 Embryonic development of the insect central complex and MB.....	24
1.7 The spider model <i>Parasteatoda tepidariorum</i>	25
1.8 Aims	26
2. Chapter Two: Methods and Materials.....	28
2.1 <i>P. tepidariorum</i> culture, embryo staging and fixation	28
2.2 General molecular biology.....	28
2.2.1 RNA extraction and cDNA synthesis	28
2.2.2 Primer design	28
2.2.3 General PCR	28
2.2.4 <i>in situ</i> probe synthesis	29
2.3 <i>in situ</i> hybridisation protocol.....	29
2.4 Imaging and image processing.....	29
2.4 Bioinformatics	30
2.4.1 Identification of candidate genes	30
2.5 Phylogenetic analysis	31

3. Chapter Three Results: Investigating cell types and patterning of the spider pre-chelicer region	32
3.1 Background	32
3.1.1 Forming the spider head: <i>hh</i> stripe-splitting segmentation	32
3.1.2 Single-cell sequencing of three stages of spider embryogenesis	32
3.1.3 Chapter three aims	35
3.2 Results: Clusters 15 and 18: Pedipalpal segment.....	35
3.3 Cluster 13: Pre-chelicer posterior	38
3.3.1 <i>Pax6</i> paralogs mark the spider ocular segment.....	41
3.3.2 <i>Pax6.1</i> is not expressed in the PEP	42
3.3.3 <i>Pax6</i> paralogs are not involved in the formation of the chelicer and pedipalpal segments.....	42
3.4 Cluster 14: Pre-chelicer anterior.....	44
3.4.1 The <i>Pax6</i> -expressing ocular segment and the <i>six3</i> -expressing anterior median region represent early independent units in the pre-chelicer region	46
3.5 Cluster 16: Pre-chelicer/chelicer boundary	47
3.5.1 <i>lim1a</i> is likely involved in the formation of the chelicer segment	49
3.6 Differentially expressed genes between <i>so1</i>+ cells VS <i>so1</i>- cells.....	50
3.7 Chapter Three Discussion	51
3.7.1 Cluster 15 and 18: Pedipalpal segment	52
3.7.2 Cluster 14: Pre-chelicer anterior, insights into arcuate body development, <i>pars intercerebralis</i> , and labrum development	52
3.7.3 Cluster 13: Pre-chelicer posterior and insights into the initiation of eye development and MB development.....	54
3.7.4 Cluster 16: Pre-chelicer/chelicer boundary and insights into stripe-splitting segmentation.....	56
4. Chapter Four Results: Studying the development of the <i>pars intercerebralis</i> and <i>pars lateralis</i> in the spider brain.....	58
4.1 Background: Anterior neuroendocrine centres in insects and centipedes are established in the <i>six3</i> expressing neuroectoderm	58
4.1.1 Embryonic development of the <i>D. melanogaster</i> PI and PL	58
4.1.2 Embryonic development of the centipede <i>Strigamia maritima</i> PI and a central NSC population that pioneers the CNS	60
4.1.3 Chapter four aims	61

4.2 Results: Development of the spider <i>pars intercerebralis</i> from bilateral invaginating head placodes.....	62
4.3 <i>foxQ2</i> and <i>vsx/chx</i> are expressed non-overlapping domains in the developing brain and labrum	63
4.4 <i>fas2A</i> expression in an INP	64
4.5 <i>tll</i> marks the developing protocerebrum and encompasses the spider PI and PL.....	67
4.6 Spiders have a <i>col</i>-expressing anterior neuronal cell population	68
4.6.1 <i>P. tepidariorum</i> PC2 has no detectable embryonic expression	70
4.7 <i>nkx2.1/scro</i> is expressed in some cells of the spider PI placode.....	71
4.8 <i>hbn</i> and <i>raxA</i> are absent from the spider PI/PL.....	73
4.8.1 <i>P. tepidariorum</i> <i>raxB</i> and <i>otp</i> have no detectable embryonic expression.....	75
4.10 Discussion	75
4.10.1 <i>foxQ2</i> in the development of the arcuate body and PI.....	76
4.10.2 <i>fas2A</i> in the development of the principal eyes and pioneer neurones.....	77
4.10.3 <i>col</i> -expressing anterior neuronal cell population	77
4.10.4 The spider anterior furrow and its molecular similarities to the vertebrate cranial placodes and insect head placode	78
4.10.5 <i>nkx2.1/scro</i> is expressed in the primordia of arthropod, polychaete, and vertebrate neurosecretory centres	79
4.10.6 <i>raxA</i> and <i>hbn</i> candidates for spider arcuate body development.....	80
4.10.7 The role <i>otp</i> in spiders is unclear	81
5. Chapter Five Results: Analysis of the retention of ohnologs in key developmental genes following whole-genome duplication in arachnopulmonates	82
5.1 Background: Evidence of WGD in arachnopulmonates.....	82
5.1.1 Chapter five aims	83
5.2 Results: <i>Hox</i> repertoires and their origins	84
5.3 <i>Wnt</i> repertoires and their origins	88
5.3.1 Arachnopulmonates retained <i>Wnt1/wg</i> ohnologs	92
5.3.2 Arachnopulmonates retained <i>Wnt4</i> ohnologs.....	92
5.3.3 Scorpions and amblypygids retained <i>Wnt6</i> ohnologs.....	92
5.3.4 Spiders and amblypygids retained <i>Wnt7</i> ohnologs	93
5.3.5 Arachnopulmonates retained <i>Wnt11</i> ohnologs.....	93
5.3.6 Unclear if <i>P. pseudoannulata</i> <i>WntA</i> duplication is true	93

5.4 <i>frizzled</i> repertoires and their origins	94
5.4.1 Amblypygids retained <i>fz1</i> ohnologs.....	96
5.4.2 Origin of spider and amblypygid <i>fz2</i> duplicates is unclear.....	97
5.4.3 Amblypygids retained <i>fz3</i> ohnologs.....	97
5.4.4 Arachnopulmonates retained <i>fz4</i> ohnologs.....	97
5.5 Sensory/neurosecretory TF repertoires and their origins	97
5.4.4 Spiders and scorpions retained <i>rax</i> ohnologs and scorpions retained <i>vsx/chx</i> and <i>otp</i> ohnologs	99
5.6 Discussion	101
5.6.1 Hox duplicates are broadly retained.....	101
5.6.2 Limited retention of Wnt duplicates.....	102
5.6.3 <i>frizzled</i> repertoires vary substantially	103
5.6.4 Sensory/neurosecretory TFs expanded in scorpions	104
6 General Discussion.....	106
6.1 Anterior neuroectoderm patterning in <i>P. tepidariorum</i> and mapping of the vertebrate mid-hindbrain boundary homolog.....	106
6.2 Further analysis of SCS cell types in spiders	108
6.3 Further analysis of the PI and PL in spiders and other arachnids	109
7 Bibliography	111
8 Supplementary Materials	119
8.1 Chapter Two Supplementary Materials.....	119
8.2 Chapter Three Supplementary Materials	119
8.3 Chapter Five Supplementary Materials	121

List of Figures

Figure 1. Head structure in Arthropoda.	13
Figure 2. Regulatory cascade controlling formation of the intercalary segment in <i>D. melanogaster</i>	15
Figure 3. Examples of anterior segmentation through <i>hh</i> stripe-splitting in different arthropod taxa.	18
Figure 4. Dynamics of eye development in <i>Parasteatoda tepidariorum</i> and expression of retinal determination genes in the eye primordia.....	20
Figure 5. Generalised spider visual pathways.....	22
Figure 6. <i>Parasteatoda tepidariorum</i> embryogenesis, life cycle and morphology.	26
Figure 7. Single-cell RNA sequencing of three embryonic stages of <i>P. tepidariorum</i>	33
Figure 8. Marker genes for clusters 15 and 18.....	36
Figure 9. Expression of cluster 15 and 18 marker genes in <i>P. tepidariorum</i> embryos.....	37
Figure 10. Cluster 13 markers.	39
Figure 11. Expression of cluster 13 marker genes in <i>P. tepidariorum</i> embryos.	40
Figure 12. Expression of <i>Pax6.1</i> and <i>Pax6.2</i> in the developing head of <i>P. tepidariorum</i> embryos.	41
Figure 13. Expression of <i>Pax6.2</i> and <i>otd2</i> in the developing head of <i>P. tepidariorum</i> embryos.	42
Figure 14. Expression of <i>Pax6.2</i> and <i>hh1</i> in the developing head of <i>P. tepidariorum</i> embryos.....	43
Figure 15. Cluster 14 markers.	44
Figure 16. Expression of Cluster 14 marker genes in <i>P. tepidariorum</i> embryos.....	46
Figure 17. Expression of <i>six3.1</i> and <i>Pax6.1</i> in the developing head of <i>P. tepidariorum</i> embryos.....	47
Figure 18. Cluster 16 markers.	48
Figure 19. Expression of Cluster 16 marker genes in <i>P. tepidariorum</i> embryos.....	49
Figure 20. Expression of <i>lim1a</i> and <i>hh1</i> in the developing head of <i>P. tepidariorum</i> embryos.	50
Figure 21. Expression of <i>IrreC-rst</i> in <i>P. tepidariorum</i> embryos.	51
Figure 22: Anterior neurosecretory centres in insects and centipedes are established in the <i>six3</i> expressing neuroectoderm.....	59
Figure 23: Expression of <i>vsx/chx</i> and <i>six3.2</i> in the developing head of <i>P. tepidariorum</i> embryos.	62
Figure 24: Expression of <i>vsx/chx</i> and <i>foxQ2</i> in the developing head of <i>P. tepidariorum</i> embryos.....	64
Figure 25: Expression of <i>fas2A</i> , <i>fas2A</i> and <i>six3.2</i> , and <i>fas2A</i> and <i>vsx/chx</i> in <i>P. tepidariorum</i> embryos.....	66
Figure 26: Expression of <i>tll</i> and <i>hh1</i> in <i>P. tepidariorum</i> embryos.	68
Figure 27: Expression of <i>six3.2</i> and <i>foxQ2</i> in the developing head of <i>P. tepidariorum</i> embryos.....	69
Figure 28: Expression of <i>col1</i> in the developing head of <i>P. tepidariorum</i> embryos.	70
Figure 29: Expression of <i>nkx2.1/scro</i> in the developing head of <i>P. tepidariorum</i> embryos.	71
Figure 30: Expression of <i>nkx2.1/scro</i> and spider stomodeal markers in the developing head of <i>P.</i> <i>tepidariorum</i> embryos.	72
Figure 31: Expression of <i>hbn</i> and <i>raxA</i> in the developing head of <i>P. tepidariorum</i> embryos.....	74
Figure 32: Anterior neurosecretory centres in spiders are established in the <i>six3</i> expressing neuroectoderm.	75
Figure 33. Repertoires of Hox genes in arachnids and other selected arthropods.	85

Figure 34. Maximum-likelihood phylogeny of Hox amino acid sequences.....	88
Figure 35. Repertoires of <i>Wnt</i> subfamilies in arachnids and other selected arthropods.	89
Figure 36. Maximum-likelihood phylogeny of Wnt amino acid sequences.	91
Figure 37. Repertoires of <i>frizzled</i> subfamilies in arachnids and other selected arthropods.....	94
Figure 38. Maximum-likelihood phylogeny of Frizzled amino acid sequences.	96
Figure 39. Repertoires of sensory/neurosecretory genes in arachnids and other selected arthropods.	98
Figure 40. Maximum-likelihood phylogeny of <i>rax</i> , <i>otp</i> , and <i>vsx/chx</i> amino acid sequences.....	100
Figure 41. Mapping the spider head boundary (SHB) based on expression patterns of mid-hindbrain boundary (MHB) marker gene orthologs.	106
Supplementary Figure 42. Expression of <i>islet/tup</i> and <i>vsx/chx</i> in <i>Parasteatoda tepidariorum</i> embryos.	120
Supplementary Figure 43. Expression of <i>Pax2.1</i> and <i>Pax2.2</i> in <i>Parasteatoda tepidariorum</i> embryos.	121
Supplementary Figure 44. <i>otp</i> , <i>vsx/chx</i> , <i>rax</i> alignment.....	126

List of Tables

Table 1. Summary of chelicerates resources surveyed.	30
Supplementary Table 2. List of primer sequences.	119
Supplementary Table 3. Differentially expressed genes between <i>so1</i> -expressing and <i>so1</i> -negative cells. ...	119
Supplementary Table 4. <i>otp</i> , <i>nkx2.1/scro</i> , <i>rax</i> , <i>vsx/chx</i> , <i>vax</i> gene IDs and homeodomain sequence.....	121

Abbreviations

AMR: anterior median region

CNS: Central nervous system

INPs: Invaginating neural precursor groups

MB: Mushroom body

NSCs: neurosecretory cells

PNS: peripheral nervous system

PI: *pars intercerebralis*

PL: *pars lateralis*

PEP: principal eye primordia

RNAi: Ribonucleic Acid interference

scRNA-seq: single-cell RNA sequencing

SEP: secondary eye primordia

SNS: stomatogastric nervous system

WGD: Whole genome duplication

1. Chapter 1: General Introduction

1.1 The arthropod head

Arthropods are a species rich phylum defined by their segmented bodies and jointed limbs. There are three living clades of euarthropods (true arthropods): Chelicerata (spiders, scorpions, mites, ticks, and horseshoe crabs), Myriapoda (centipedes and millipedes), and Pancrustacea (crustaceans and insects) (Figure 1). The Chelicerata clade diverged from other mandibulate arthropods 500 million years ago (Figure 1). The success of arthropods has been attributed to their modular body plan, which allows individual segments to evolve independently to serve different functions (Chipman and Edgecombe, 2019).

Extant arthropod heads contain a conserved anterior region of three developmental units, each containing a ganglion that makes up the tripartite brain (Figure 1). The anterior-most segment of mandibulates is the antennal segment and of chelicerates the cheliceral segment, or cheliforal segment in pycnogonids (Jager *et al.*, 2006), which gives rise to the deutocerebrum (Figure 1). This segment is serially homologous to trunk segments (Posnien *et al.*, 2023). The segment posterior to the antennal/cheliceral segment gives rise to the tritocerebrum (Figure 1). In hexapods and myriapods, this segment is reduced and lacks an appendage, referred to as the intercalary segment (Figure 1). While in malacostracans and chelicerates, this segment is proper with a pair of appendages, the second antennae and pedipalps, respectively (Figure 1). In all arthropods, the Hox gene *labial* marks the tritocerebral segment while the protocerebral and deutocerebral segments free of Hox gene expression (Hughes and Kaufman, 2002) (Figure 1). The region anterior to the antennal/cheliceral segment is thought to be non-segmental, referred to as the pre-antennal/cheliceral region, the neuroectoderm of which gives rise to the protocerebrum (Posnien *et al.*, 2023) (Figure 1).

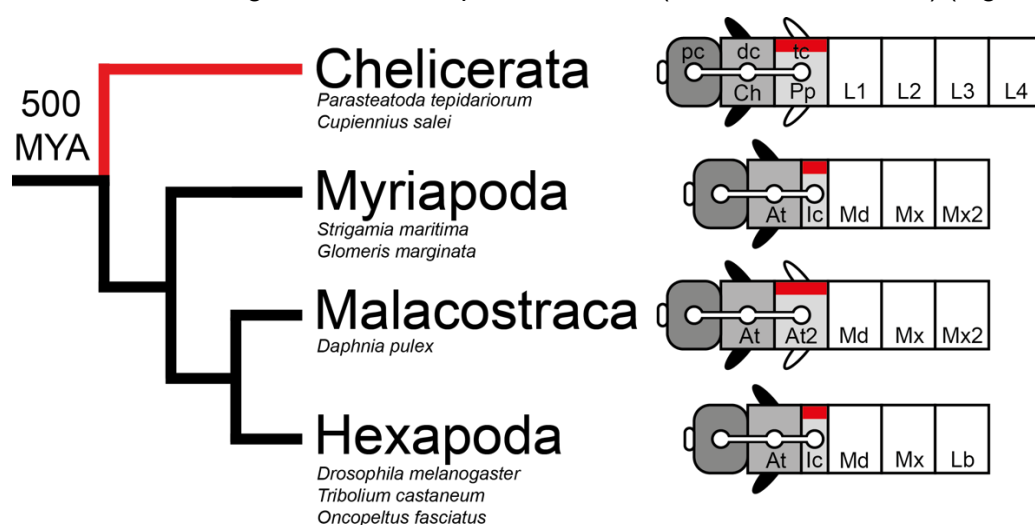


Figure 1. Head structure in Arthropoda.

Schematic of the prosoma of chelicerates and head of mandibulates (black branches) with the anterior to the left. Different shades of grey indicate their shared ancestry. The Hox gene *labial* (red) marks the

tritocerebral segment. Supraesophageal ganglion: protocerebrum (pc), deutocerebrum (dc), and tritocerebrum (tc). The labrum is shown as an anterior medial structure of the pre-antennal/chelicerai region. The deutocerebral appendage (black appendage) and tritocerebral appendage (white appendage). Other abbreviations: At: antennal segment; At2: second antennal segment; Ch: chelicerai segment; Ic: intercalary segment; Lb: labial segment; Md: mandibular segment; Mx: maxillary segment; Pp: pedipalpal segment; L1-4: walking leg segments 1-4.

1.1.1 Development of the pre-antennal, antennal, and intercalary segments in *D. melanogaster*

Most insights into arthropod segmentation are based on the fly *Drosophila melanogaster* (Clark *et al.*, 2019). In *D. melanogaster*, the pre-antennal, antennal, intercalary, and mandibular segment are patterned by the Bicoid-regulated head gap-like genes *orthodenticle* (*otd*), *empty spiracles* (*ems*), and *buttonhead* (*btd*) which are expressed early in embryogenesis in broad and overlapping domains (Cohen and Jürgens, 1990). Loss of function mutations of head gap-like genes results in the deletion of several contiguous head segments, roughly corresponding to their expression patterns.

In insects, pair-rule genes downstream of gap genes are not expressed in the pre-antennal, antennal, and intercalary during early development (Posnien *et al.*, 2010; Rogers and Kaufman, 1997). Therefore, the head gap-like genes activate the segment polarity genes directly or via second-order regulators, similar to trunk pair-rule genes. For example, *collier/knot* (*col*) expression in PS0 is under the control of *btd* and *ems* and is required for the expression of the segment polarity gene *hedgehog* (*hh*) in the posterior of the intercalary segment (Crozatier *et al.*, 1999; Ntini and Wimmer, 2011) (Figure 2). *col* is also required for the correct expression *cap 'n' collar* (*cnc*) (Crozatier *et al.*, 1999; Seecoomar *et al.*, 2000), which specifies mandibular gnathal segment identity (Mohler *et al.*, 1995) (Figure 2). Furthermore, *col* is repressed at its posterior limit by the trunk pair-rule gene *even-skipped* (*eve*) thus, taking input from both the head and the trunk patterning system (Crozatier *et al.*, 1999) (Figure 2).

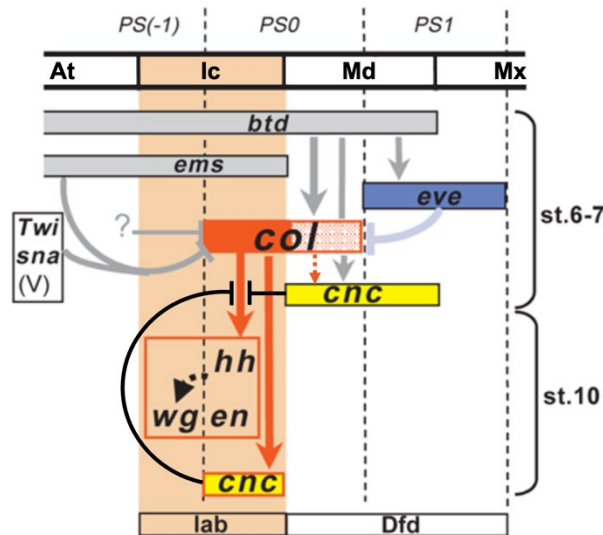


Figure 2. Regulatory cascade controlling formation of the intercalary segment in *D. melanogaster*.

Schematic representation of spatial expression domains (boxes) and repressing (bars) and activating (arrows) regulatory activities. *btd* activation of *eve* in PS1 creates a regulatory loop allowing for differential gene expression between PS0 and PS1. *col* activates primary *cnc* expression in the anterior Md (orange dashed arrow; Seecoomar et al. (2000)). At stage 10, *col* expression has become restricted to the anterior PS0 (orange box), where it activates *hh* and secondary *cnc* expression in the posterior Ic (orange arrows), and possibly *en*. Loss of *hh* activity leads to the non-cell-autonomous loss of *wg* expression in adjacent PS(-1) cells (black dashed arrow; Gallitano-Mendel and Finkelstein (1998)). Physical protein interaction detected between Col and CncB may be implicated in weakening the transcriptional activator function of Col on the Ic *hh* expression (black bars; Ntini and Wimmer (2011)), first to restrict expression to the posterior Ic and then diminish this expression at a later stage (late 10/11). Adapted from Crozatier et al. (1999). Other abbreviations: At: antennal segment; Ic: intercalary segment; Md; mandibular segment; Mx: maxillary segment; PS: parasegments; V: ventrally.

1.2 Genes involved in arthropod head patterning

In *D. melanogaster* head involution leads to a derived and reduced larval head morphology hampering mutant analysis (Posnien et al., 2010). To understand the evolution of genetic mechanisms controlling head development, orthologs of *D. melanogaster* head gap-like genes, second-order regulators, and segment polarity genes have been studied across arthropod representatives. In insects and myriapods, *otd*, *ems*, and *btd* orthologs are expressed during early head development and in the same anterior-to-posterior order as *D. melanogaster*, but in contrast to the fruit fly without significant overlap (Schinko et al., 2008; Hunnekuhl and Akam, 2017).

1.2.1 *otx/otd*

Across arthropods, *otd* often exhibits an early ring or arc-like pattern (Browne *et al.*, 2006; Pechmann *et al.*, 2009; Birkan *et al.*, 2011; Hunnekuhl and Akam, 2017), consistent with the hypothesis that the prebilaterian ancestor had a radially symmetric body plan (Bruce and Shankland, 1998). The head gap gene-like function of *otd* appears to be evolutionarily conserved across arthropods. In the red flour beetle *Tribolium castaneum*, *otd1* knockdown results in a range of phenotypes from deletion of the ocular and antennal segments (Schröder, 2003), similar to *otd* loss of function in *D. melanogaster* (Finkelstein and Perrimon, 1990), to deletion of the head and parts of or the entire thorax (Schinko *et al.*, 2008). Thus, it was suggested *T. castaneum otd1* has an early regionalisation role and later functions as a head gap-like gene (Schinko *et al.*, 2008). In a hemimetabolous insect, the milkweed bug *Oncopeltus fasciatus otd* knockdown results in reduction or loss of the pre-gnathal segments (PGS) as well as malformed mandibular and maxillary gnathal segments (Birkan *et al.*, 2011). In the spider *Parasteatoda tepidariorum*, *otd1* knockdown results in the deletion of pre-cheliceral region and cheliceral segments (Pechmann *et al.*, 2009).

1.2.2 *emx/ems*

In *D. melanogaster*, *ems* mutants lack the posterior portion of the ocular segment as well as the antennal and intercalary segments and the anterior portion of the mandibular segment (Dalton *et al.*, 1989; Cohen and Jürgens, 1990; Walldorf and Gehring, 1992). It is unclear whether the function of *ems* in *D. melanogaster* is evolutionarily conserved, even in other insects. In *T. castaneum*, *ems* knockdown has a minor role in patterning the antennal segment (Schinko *et al.*, 2008). In *Oncopeltus*, the role of *ems* could not be analysed due to the low penetrance of RNAi (Birkan *et al.*, 2011). Early embryonic expression of *ems* in the crustacean *Parhyale hawaiiensis* is inconsistent with a head gap gene role as *ems* expression in the anterior head occurs after the head lobes appear (Browne *et al.*). While in chelicerates early patterning roles of *ems* orthologs have not been studied (Simonnet *et al.*, 2006; Leite *et al.*, 2018).

1.2.3 *Sp5/btd*

In *D. melanogaster*, *btd* loss of function leads to the lack of antennal, intercalary, mandibular and the anterior portion of the maxillary segments (Wimmer *et al.*, 1996). Similarly in *T. castaneum*, *btd* knockdown results in the loss of the pre-antennal region to the mandibular segment (Jeon *et al.*, 2019). This contrasts previous knockdown experiments which found *btd* was not required for *T. castaneum* head development (Schinko *et al.*, 2008). While the role of *btd* as a head gap gene appears conserved in holometabolous insects no *Sp5/btd* orthologs

have been recovered from the genomes of *O. fasciatus* (Birkan *et al.*, 2011), *Parhyale*, *Parastetaoda* or other arachnids (Setton and Sharma, 2018).

1.2.4 *col* and *cnc*

col is expressed in the intercalary segment of hexapods and myriapods but not in the second antennal segment of crustaceans or pedipalps of chelicerates (Schaeper *et al.*, 2010; Janssen *et al.*, 2011; Hunnekuhl and Akam, 2017; Schomburg *et al.*, 2020).

At late embryonic stages, the expression of *cnc* orthologs in the labrum and mandibular segment is conserved among mandibulates (Janssen *et al.*, 2011; Sharma *et al.*, 2014; Hunnekuhl and Akam, 2017; Birkan *et al.*, 2011). In *T. castaneum*, *cnc* knockdown results in a deletion of the labrum and a homeotic transformation of the mandibular segment into a maxillary identity (Coulcher and Telford, 2012), similar to *D. melanogaster cnc* loss of function (Mohler *et al.*, 1995). While in chelicerates, *cnc* orthologs have ubiquitous expression throughout the embryo (Sharma *et al.*, 2014; Schomburg *et al.*, 2020).

1.2.5 *hh* stripe-splitting

In insects, myriapods, and chelicerates an early anterior *hh* domain splits once or twice to give rise to the three stripes each marking the border of the pre-antennal/cheliceral region, antennal/cheliceral segment and intercalary/pedipalpal segment (Figure 3). This stripe-splitting segmentation is understood in detail in *P. tepidariorum* (Pechmann *et al.*, 2009; Kanayama *et al.*, 2011; Akiyama-Oda and Oda, 2020) (Figure 3A). In this spider, an anterior *hh1* stripe travels posteriorly in the nascent germband, maintained at the posterior limit of the *otd1* stripe, and splits twice to produce first the pedipalpal and then cheliceral and pre-cheliceral stripes (Pechmann *et al.*, 2009; Kanayama *et al.*, 2011) (Figure 3A). In the centipede *Strigamia maritima*, it is unclear whether the anterior *hh1* stripe travels but it also splits twice, however, in contrast to the spider, the posterior of the first two *hh1* stripes undergoes the second split and only the initial *hh1* stripe overlaps with *otdB* expression (Hunnekuhl and Akam, 2017) (Figure 3B). In *Oncopeltus*, the anterior *hh* stripe splits once to give the pre-antennal/ocular and antennal stripes and a third stripe appears later marking the intercalary segment (Lev and Chipman, 2021) (Figure 3C). In other insects, including *T. castaneum* (Farzana and Brown, 2008) and the cricket *Gryllus bimaculatus* (Miyawaki *et al.*, 2004) as well as the millipede *Glomeris marginata* (Janssen, 2012), a process similar to that in *O. fasciatus* is observed. In *D. melanogaster* head segmentation, an early anterior *hh* domain is subdivided into three each marking a PGS (Lee *et al.*, 1992). This suggests across arthropods anterior segment generation involves a stripe-splitting mechanism, rather than a sequential addition of segments.

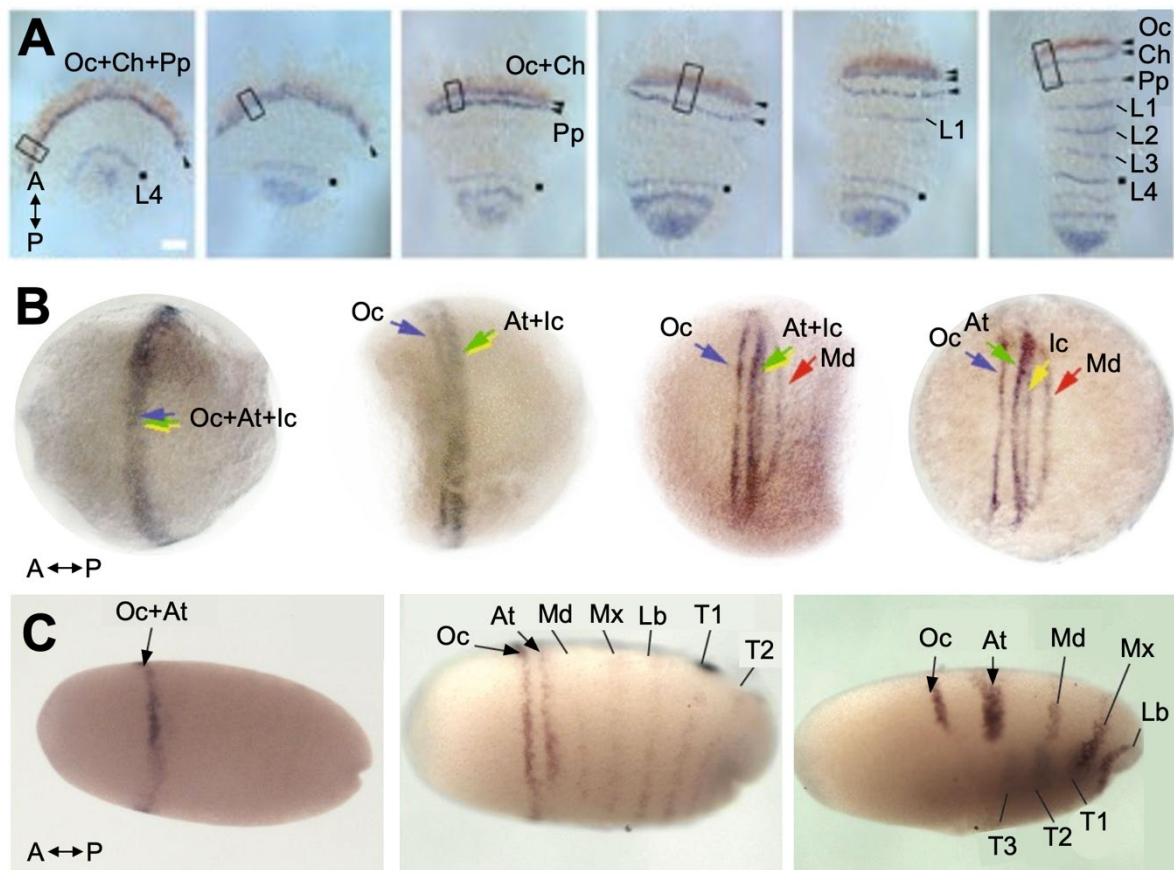


Figure 3. Examples of anterior segmentation through *hh* stripe-splitting in different arthropod taxa.

Formation of the ocular/pre-antennal/pre-chelicerai (Oc), antennal/chelicerai (At/Ch), and intercalary/pedipalpal (Ic/Pp) *hh* stripes in a representative of Chelicerata **(A)**, Myriapoda **(B)**, and Hexapoda **(C)**. **(A)** *Parasteatoda tepidariorum* embryos stained for *hh1* (purple) and *otd1* (red) transcripts. **(B)** *Strigamia maritima* embryos stained for *hh1* transcripts. **(C)** *Oncopeltus fasciatus* embryos stained for *hh* transcripts. Other abbreviations: Lb: labial; Md: mandibular; Mx: maxillary; T1-3: thoracic 1-3; L1-4; walking leg 1-4. Adapted from Kanayama et al. (2011), Hunnekuhl and Akam (2017), and Lev and Chipman (2021).

1.3 Arthropod visual system

In all arthropods, the pre-antennal/chelicerai region gives rise to the eyes, with most arthropods having two separate visual systems: the median and lateral eyes. Our understanding of arthropod eye development comes from *D. melanogaster*. The fly has a pair of lateral compound eyes, which are the main optical organs, and a group of three median ocelli.

1.3.1 Spider vision

Spiders have simple eyes, each with a single fixed cuticular lens (Land, 1985). Typically, spiders have one pair of median principal eyes, which are usually the largest and the main optical system, and three pairs of lateral secondary eyes. The principal and secondary eyes differ in retinal morphology, embryonic origin (Schomburg *et al.*, 2015; Samadi *et al.*, 2015), and functional capabilities. Furthermore, their visual inputs are processed in separate pathways with distinct neuropil regions (Figure 5A&B) (Long, 2021). The retinas of the principal eyes are everted, meaning the light-sensitive rhabdoms of the photoreceptors face incoming light (Homann, 1971). Moreover, the retinas of the principal eyes can be moved behind the fixed lenses by muscles attached to the retinal envelop. The sophistication of this musculature and degree of retinal movement is taxa-dependent (Homann, 1971). The secondary eyes possess a reflective tapetum behind the retinas that increases visual sensitivity.

Across spider families, there is variation in the number, placement, size, and function of spider eyes (Morehouse *et al.*, 2017). This diversity presents an opportunity to study the genetic bases of visual specialisations. However, our understanding of the patterning of the pre-cheliceral region and the molecular and genetic underpinnings of spider eye development is lacking.

1.3.2 Eye development in *P. tepidariorum*

Transcription factors of the *Drosophila* retinal determination gene (RDG) network are duplicated in *P. tepidariorum*, except for orthologs of *eyes absent* (*eya*) and *eyegone* (Schomburg *et al.*, 2015) (Figure 4A). *P. tepidariorum* possesses single-copy orthologs of *eyeless* (*ey*) and *twin of eyeless* (*toy*), *Pax6.1* and *Pax6.2*, respectively (Friedrich, 2017; Friedrich, 2022; Aase-Remedios *et al.*, 2023) (Figure 4A). The expression patterns of RDG paralogs have been studied during late embryonic head development (Schomburg *et al.*, 2015). The ancestral arachnospulmonate whole genome duplication (WGD) could have played a role in the evolution of spider visual systems, perhaps contributing to changes in function or developmental system drift.

In *P. tepidariorum*, the developing eye primordia can be visualised by *sine oculis-1* (*so1*) expression during late embryonic head development, from stage 10.2 to 14.1 (Figure 4B) (Schomburg *et al.*, 2015). The principal and secondary eye primordia (PEP and SEP) originate from separate groups of cells in the non-neurogenic ectoderm of stage 10.2 embryos, located at the anterior rim of the developing head lobes (Figure 4B). At stage 12 the PEP travels with the overgrowing edge of the non-neurogenic ectoderm towards the final position of the anterior median eye and the SEP split into three discernible domains, corresponding to

an individual of each secondary eye pair (Figure 4B). At stages 13.1 and 14.1 the secondary eye domains migrate towards their final position in the developing head (Figure 4B).

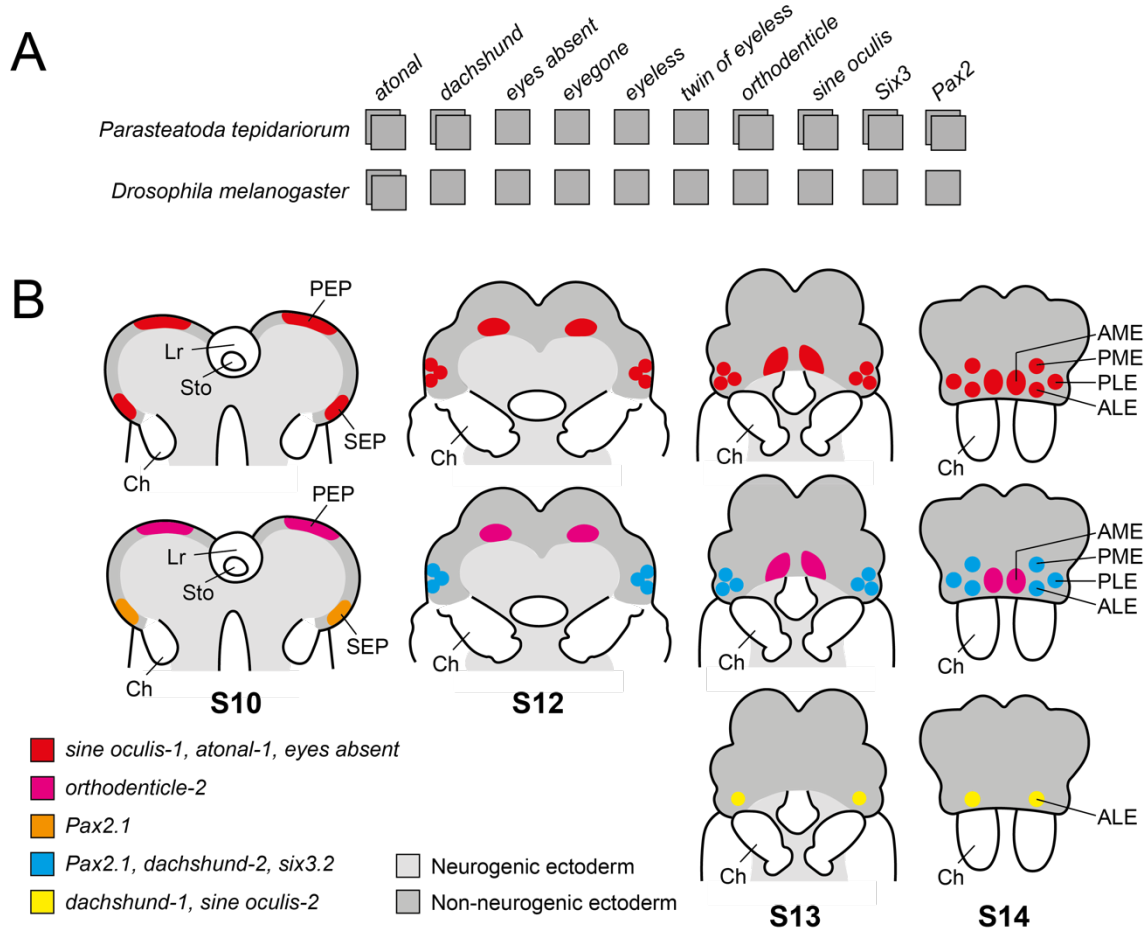


Figure 4. Dynamics of eye development in *Parasteatoda tepidariorum* and expression of retinal determination genes in the eye primordia.

(A) Copy number of RDG paralogs. (B) Expression of RDGN paralogs in *P. tepidariorum* based on *in situ* hybridisation data from Schomburg et al. (2015), Baudouin-Gonzalez et al. (2022), and Janeschik et al. (2022). The PEP and SEP are established in the non-neurogenic ectoderm (dark grey) adjacent to the anterior and lateral furrows (not shown), respectively. Other retinal determination genes mark the PEP (magenta), SEP (orange and blue), and ALE (yellow). Abbreviations: ALE: anterior lateral eyes; AME: anterior median eyes; Ch: chelicerae; Lr: labrum; PEP: principal eye primordia; PLE: posterior lateral eyes; PME: posterior median eyes; SEP: secondary eye primordia; Sto: stomodeum.

Not only do the two visual systems of spiders have a different embryonic origin, but there is also a molecular distinction between them (Schomburg et al., 2015) (Figure 4B). Interestingly the spider *Pax6* paralogs are not expressed in the developing eye primordia. Like *D. melanogaster* Wnt-signalling likely represses the development of spider eyes (Baudouin-Gonzalez et al., 2022). Furthermore, *so1* and *eyes absent* (*eya*) are expressed in all spider eyes along with the RDG downstream factor *atonal-1* (Schomburg et al., 2015; Baudouin-Gonzalez et al., 2022) (Figure 4B).

Parental RNAi has been used to knockdown RDG paralogs expressed in the eye primordia of *P. tepidariorum*, to assess their function during the development of spider eyes. Parental knockdown of *so1* resulted in a spectrum of eye defects that include reduced pigmentation, defects of the tapetum, or complete absence of the eye (Gainett *et al.*, 2020). As expected, all eye pairs were affected, with the anterior median eyes perturbed the most. While parental knockdown of unique markers for the PEP, *orthodenticle-2*, and SEP, *dachshund-2* and *six3.2*, did not result in any detectable eye phenotypic effects (Turetzek *et al.*, 2015; Gainett *et al.*, 2020; Schacht *et al.*, 2020) (Figure 4B).

1.4 Arthropod midline neuropils, the insect central complex and chelicerate arcuate body

Arthropods have an unpaired midline neuropil termed the central body in mandibulates and arcuate body in chelicerates (Strausfeld *et al.*, 1993; Homberg, 2008). In insects and malacostracans, the central body is embedded in the protocerebrum and connected to a second midline spanning neuropil, a superficial monolayered protocerebral bridge (Utting *et al.*, 2000). Together the two neuropils form the central complex. The central complex neuropils receive indirect visual input from the lateral eyes via higher-order visual neuropils (Homberg, 2008). The protocerebral bridge in insects and decapod crustaceans also has connections with the median eyes (Homberg, 2008). Functional studies in insects suggest the central complex plays a role in spatial orientation and object recognition (reviewed in Homberg (2008)).

The chelicerate arcuate body is a crescent-shaped neuropil, not embedded in the protocerebrum but occupying a superficial, dorsal posterior position in the adult brain (Figure 5A). Only in Acari (mites and ticks), where the central nervous system (CNS) is fused into a single compact synganglion the arcuate body is vestigial or absent (Homberg, 2008). The chelicerate arcuate body does not form complexes with satellite neuropils, unlike insects and malacostracans, and has a different internal neuroarchitecture (Homberg, 2008). Furthermore, the spider arcuate body receives direct input from the second-order visual neuropil (the medulla) of the median eyes (Strausfeld *et al.*, 1993) (Figure 5A). The retinal movements of the principal eyes support sustained evaluation of stationary objects and complex scenes, this suggests that the chelicerate arcuate body has a similar functional role to the insect central complex (Homberg, 2008). Across spider families, connections with the first-order visual neuropil (the lamina) of the lateral eyes are present (Long, 2021) (Figure 5B). Onychophora are considered the sister group the sister group of Arthropoda forming the clade Panarthropoda. The midline neuropil of onychophorans shares features of the chelicerate arcuate body (Homberg, 2008).

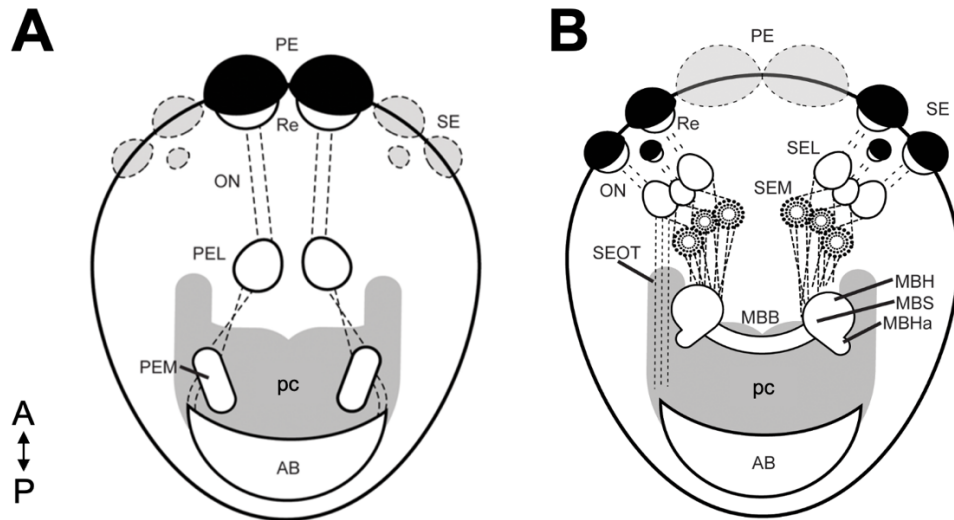


Figure 5. Generalised spider visual pathways.

Schematic of visual pathways of the principal eyes (PE) **(A)** and secondary eyes (SE) **(B)** with major nerve tracts (dashed lines) and examples of neuronal cell types based on *C. salei*. **(A)** The crescent-shaped neuropil arcuate body (AB) processes information from the PE. **(B)** The secondary eye lamina (SEL) is a first-order visual neuropil. The secondary eye medulla (SEM) is comprised of optic glomeruli. The mushroom body (MB) is a third-order visual neuropil composed of globuli cells. The input region is referred to as the MB head (MBH). The secondary eye optical tracts (SEOT) terminate on the ventral side of the AB. Other abbreviations: MBB, mushroom body bridge; MBHa, mushroom body haft; MBS, mushroom body shaft; ON, optic nerve; pc, protocerebrum; Re, retina. Adapted from Long (2021).

1.5 Mushroom bodies (MB)

Arthropods have paired lobed or dome-like neuropils referred to as mushroom bodies (MB) or dome-like hemiellipsoid bodies (HB) in malacostracans, respectively (Strausfeld *et al.*, 1998). Arthropod MB consist of several densely packed intrinsic neurones called globuli cells (Kenyon cells in insects). The globuli cells extend dendrites into the primary input region called the calyx. In insects, the calyx receives olfactory input from the antennal lobe glomeruli. The *D. melanogaster* MB, are important centres for olfactory learning and memory (reviewed in Heisenberg (2003) and Davis (2005)).

The Chelicerate MB can reach varying degrees of elaboration (Strausfeld *et al.*, 1998). For example, scorpion MB are relatively small while, amblypygid (whip-spider, not true-spider) MB are large and richly convoluted (Strausfeld *et al.*, 1998). Except for spiders (see below), globuli cells are supplied by afferents that relay from olfactory glomeruli situated outside the chelicerate brain, in contrast to insects. The olfactory glomeruli are situated in segmental neuromeres associated with specialised opisthosomal appendages or modified walking legs. For example, scorpions possess a pair of fan-like chemosensory organs called the pectines that extend from the third opisthosomal segment. In amblypygids, the first pair of legs is

modified into elongated and thin feelers called antenniform legs. Furthermore, in amblypygids the MB calyx was recently shown to receive visual input from the second-order visual neuropil (the medulla) of the secondary eyes (Sinakevitch *et al.*, 2021), similar to spiders (Strausfeld *et al.*, 1993) (Figure 5B).

In *C. salei*, the MB are third-order visual neuropils of the secondary eyes (Strausfeld *et al.*, 1993) (Figure 5B). Long (2021) investigated the secondary eye visual pathway across 19 different families with different eye morphology and found large variations in morphology and organisation of the neuropils (Long, 2021). Some spiders had greatly reduced and possibly absent MB (Long, 2021). Other spiders had complex systems similar to *C. salei*, with large laminae, medullae formed from optical glomeruli, and robust MB (Long, 2021) (Figure 5B).

The MB of other insects, myriapods and three arachnospulmonate-arachnids: scorpions, amblypygids, and thelyphonids (vinegaroons or whip scorpions) have been shown to express proteins essential for olfactory learning and memory in *D. melanogaster* (Wolff and Strausfeld, 2015): protein kinase A catalytic subunit alpha (DC0) (Skoulakis *et al.*, 1993); Leonardo (Leo), the ortholog of mammalian 14-3-3 (Skoulakis and Davis, 1996); and phosphorylated calcium/calmodulin-dependent protein kinase II (pCaMKII) (Wang *et al.*, 1998). In addition to the HB of malacostracans, insect-like MB of stomatopod crustaceans, and MB homologs in certain Lophotrochozoa including the nonganglionic brains of polychaete annelids, polyclad planarians, and nemerteans (Wolff and Strausfeld, 2015; Wolff *et al.*, 2017). This suggests an ancestral presence predating arthropod and lophotrochozoan origins, and a conserved function in learning and memory. In addition to challenging the doctrine that malacostracans HB are a derived trait and evolutionary convergent with insect MB (Strausfeld *et al.*, 2020).

I would expect the spider MB to express DC0, which Wolff and Strausfeld (2015), demonstrated is independent of sensory modality. However, Long (2021) reports from personal communication with Strausfeld, that spider MB do not express proteins MB functional proteins, and thus spider MB have been lost or may not serve the same function as other arachnospulmonate arachnid MB.

The spider's secondary eyes provide wider peripheral vision, primarily sensitive to motion, serving to guide reorientation of the visual system for subsequent evaluation by the principal eyes (Morehouse *et al.*, 2017). Thus, the spider MB have been suggested to have a role in movement detection. However, in some families of hunting spiders, including the wolf spiders (Lycosidae), the acuity and sensitivity of the secondary posterior median eye pair can exceed the principal eyes (Morehouse *et al.*, 2017).

1.6 Embryonic development of the insect central complex and MB

Gene expression has been used to homologize body regions across Bilateria. Molecular mapping of the neuroectoderm of insects separates the pre-antennal neuroectoderm into *six3* and *otd*-expressing protocerebrum, defined as the archi- and prosocerebrum, respectively (Posnien *et al.*, 2023). The insect central complex was assigned a mixed origin while the insect MB are thought to derive from the prosocerebrum (see Posnien *et al.* (2023) and references therein). The insect neuroendocrine centres, the *pars intercerebralis* (PI) and *pars* (PL), are derived from the archicerebrum (de Velasco *et al.*, 2007; Steinmetz *et al.*, 2010; Posnien *et al.*, 2011b). Furthermore, the arthropod protocerebrum is homologous to the vertebrate fore- and midbrain. The boundary between the pre-antennal region and antennal segment, termed the insect head boundary (IHB), corresponds to the vertebrate mid-hindbrain boundary (MHB) (Posnien *et al.*, 2023).

Comparable data from other arthropods, notably chelicerates, would test whether conclusions made by Posnien *et al.* (2023) were biased by holometabola-specific aspects of gene expression. Comparable data from spiders would aid in homologizing specific brain structures such as the insect central complex and MB with the chelicerate arcuate body and spider MB, respectively. In addition, neuroendocrine cell populations have yet to be characterised in a chelicerate. Tessmar-Raible (2007) note the difficulty in comparing possibly corresponding neurosecretory systems between the adult brains of animals. Therefore, suggests comparing early developmental stages, focusing on marker genes that help define the molecular identity of the developing brain regions (Tessmar-Raible, 2007). Thus, studying the expression of insect and centipede neuroendocrine in a chelicerate such as the spider will help elucidate the origin of these cell populations.

In spiders, invaginating neural precursors (INPs) arise sequentially and in a stereotyped pattern comparable to the formation of insect neuroblasts (Stollewerk *et al.*, 2001; Stollewerk *et al.*, 2003). The arcuate body is formed by large crescent-shaped grooves of the pre-chelicerar neuroectoderm, starting at the labrum, termed anterior furrows or semi-lunar/cerebral grooves (Doeffinger *et al.*, 2010; Samadi *et al.*, 2015). Within the walls of the anterior furrows are several INPs, for example, in *Cupiennius salei*, there are approximately 25 INPs in each anterior furrow (Doeffinger *et al.*, 2010). Distinct elevations of the pre-chelicerar neuroectoderm termed medial subdivisions have been shown to represent the neural primordium of the spider MB (Doeffinger *et al.*, 2010; Samadi *et al.*, 2015).

1.7 The spider model *Parasteatoda tepidariorum*

The spider *Parasteatoda tepidariorum* has emerged as one of the main chelicerate model species for evolutionary and developmental biology (Hilbrant *et al.*, 2012; Schwager *et al.*, 2017; Oda and Akiyama-Oda, 2020). Comparing spider development with mandibulates can be useful in identifying which modes and/or mechanisms of development are ancestral or derived with respect to the arthropod common ancestor. Furthermore, a whole-genome duplication (WGD) occurred in the arachnopulmonate ancestor approximately 450 MYA, the lineage includes spiders and scorpions but excludes other arachnids like mites, ticks, and harvestmen. In addition to duplicated Hox gene clusters, many other developmental genes have been retained in multiple copies in *P. tepidariorum* (Janssen *et al.*, 2010; Janssen *et al.*, 2015; Schwager *et al.*, 2017; Leite *et al.*, 2018; Aase-Remedios *et al.*, 2023). Therefore, the spider can be used to investigate the potential roles of WGD on arachnid diversification and evolution.

P. tepidariorum embryogenesis has been previously described, with a clear and detailed staging system (Mittmann and Wolff, 2012) (Figure 6). This spider species is easy to maintain under lab conditions, with females producing cocoons containing up to 300 embryos. Furthermore, general embryonic development is more synchronised in *P. tepidariorum* than in *C. salei* (Wolff and Hilbrant, 2011). A colourimetric *in situ* hybridisation (ISH) protocol allows the expression pattern analysis of any developmental gene of interest (Akiyama-Oda and Oda, 2003). Furthermore, double fluorescent ISH (dFISH) and multi-colour FISH of up to four different gene transcripts (Akiyama-Oda and Oda, 2016), enables simultaneous visualisation of gene expression. In addition, parental and embryonic RNAi (pRNAi or eRNAi) to disrupt gene expression in this spider (Akiyama-Oda and Oda, 2006; Pechmann *et al.*, 2011; Kanayama *et al.*, 2011).

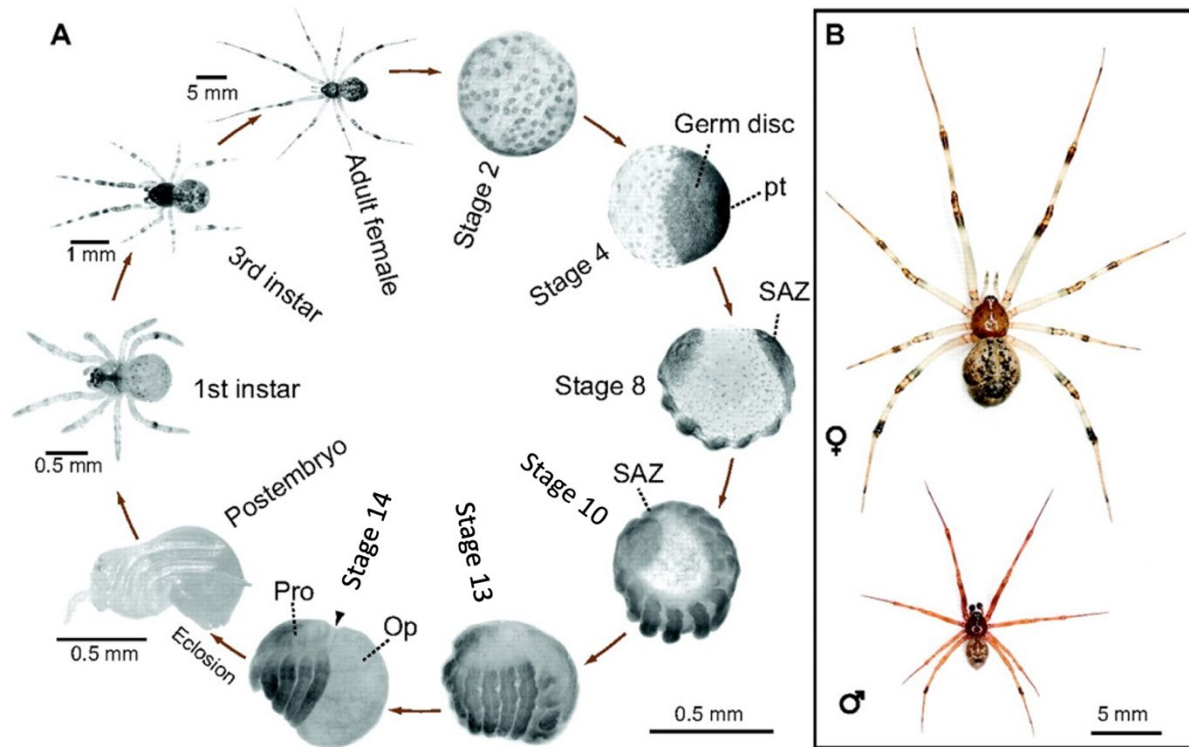


Figure 6. *Parasteatoda tepidariorum* embryogenesis, life cycle and morphology.

(A) Key stages of embryonic development and post-embryonic development: blastoderm (stage 2), germ disc (stage 4), early segmented germ band (stage 8), and leg and germ band elongation (stage 10). **(B)** Adult female (♀) and male (♂). Other abbreviations: Op; opithosoma; Pro: prosoma; pt: primary thickening; SAZ: segment addition zone. Adapted from Hilbrant et al. (2012).

1.8 Aims

My PhD project aimed to provide new insights into head and brain development in the spider *P. tepidariorum*. To address this overall aim, my experimental plan was split into the following three chapters:

- 1. Chapter Three Results:** Investigating cell types and patterning the spider pre-cheliceral region

This chapter aimed to understand the regulation of the development of the spider pre-cheliceral region, cheliceral and pedipalp segments of the spider embryo through analysis of marker genes in five putative anterior cell clusters from scRNA-seq analysis at three embryonic stages of *P. tepidariorum* (Leite et al., 2024).

- 2. Chapter Four Results:** Studying the development of the *pars intercerebralis* and *pars lateralis* in the spider brain

This chapter aimed to characterise the development of neuroendocrine cell populations in a chelicerate by analysing the expression of known marker genes from insects and centipedes during spider embryogenesis.

3. Chapter Five Results: Analysis of the retention of ohnologs in key developmental genes following whole-genome duplication in arachnopulmonates

This chapter aimed to survey Hox, Wnt, frizzled and head patterning genes in the transcriptomes or genomes of fourteen chelicerate species including spiders, scorpions, amblypygids and horseshoe crabs as well as other selected arthropods, and to determine the origin of duplicates by phylogenetic analysis.

2. Chapter Two: Methods and Materials

2.1 *P. tepidariorum* culture, embryo staging and fixation

The *P. tepidariorum* culture is maintained at 25°C in a 12h light/12h dark cycle. Embryos were staged according to (Mittmann and Wolff, 2012). Embryos were dechorionated with bleach (sodium hypochlorite, 5% active chlorine, Arcos) and tap water (1:1). Dechorionated embryos were fixed overnight at room temperature (RT) in a two-phase solution of 100% heptane and 37% formaldehyde in 2xPEM (PIPES, EGTA, and MgSO₄) (2:1:1). After fixation, embryos were washed with ice-cold 100% methanol, rocked for ≥30 minutes at room temperature (RT) and stored at -20°C. The vitelline membranes of fixed embryos were removed with Dumount 5 forceps in 100% methanol.

2.2 General molecular biology

2.2.1 RNA extraction and cDNA synthesis

P. tepidariorum embryos of stages 5-13 were pooled and total RNA was extracted using QIAzol Lysis Reagent [79306, Qiagen], following the manufacturer's guidelines, and stored at -80°C. Total RNA was used in cDNA synthesis using the QuantiTect Reverse Transcription Kit [205311, Qiagen], according to the manufacturer's guidelines.

2.2.2 Primer design

Primer3web (<http://primer3.ut.ee>; default settings) was used to design gene-specific primers that cover a sequence of the coding region. For ISH primers, T7 linker sequences were added to the 5' end of the forward primer (5'-GGCCGCGG-3') and reverse primer (5'-CCCGGGGC-3'). A complete list of primers can be found in Supplementary Table 2.

2.2.3 General PCR

Gene-specific cDNA fragments for *in situ* probe synthesis were amplified using the OneTaq 2x Master Mix [M0482, New England Biolabs/NEB]. PCR products were run on a 1% agarose gel to verify band size and purified using the NucleoSpin Gel and PCR Clean-up kit [12303368, Macherey-Nagel].

PCR program		35 cycles	50 µl reaction mix	
Initial Denaturation	94°C 30 s		OneTaq 2x MasterMix	25 µl
Denaturation	94°C 20 s		Forward primer (10 mM)	1 µl
Annealing	58-68°C 35 s		Reverse primer (10 mM)	1 µl
Extension	68°C 1min/kb		Template DNA	Variable
Final Extension	68°C 5 min		Autoclaved dH2O	Up to 50 µl
Final Hold	10°C			

2.2.4 *in situ* probe synthesis

Fragments of the coding region were amplified from cDNA by two rounds of PCR: the first PCR using gene-specific primers (Supplementary Table 2) and the second PCR using the gene-specific forward primer and 3'UTP-T7 reverse primer (AGGGATCCTAATACGACTCACTATAGGGCCCGGGGC). Digoxigenin (DIG) and Fluorescein (FITC) labelled RNA probes were synthesised using T7 RNA polymerase [10881775001, Roche] and DIG [11277073910, Roche] or FITC [11685619910, Roche] RNA Labelling Mix, respectively, according to the manufacturer's guidelines.

2.3 *in situ* hybridisation protocol

For colorimetric *in situ* hybridisation (ISH), minor modifications were made to the whole-mount ISH protocol described in Prpic *et al.* (2008). Post-fixation embryos were gradually moved from 100% methanol to Phosphate Buffered Saline with Tween-20 (0.02%) (PBS-T). The Proteinase K Digestion (steps 4-8) was replaced by two 10-minute washes in PBS-T (Prpic *et al.*, 2008). For steps 18 and 19, embryos were incubated with gentle agitation at RT for 30 minutes and 2 hours, respectively (Prpic *et al.*, 2008). When required, embryos were ethanol treated to decrease background: embryos were washed for 5 minutes 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 washed twice with PBS-T.

For double fluorescence ISH (dFISH), the protocol described by Clark and Akam (2016) was followed with some modifications. The post-fixation, pre-hybridisation, hybridisation, and probe removal steps follow steps 1-18 in the protocol described by Prpic *et al.* (2008) subject to the modifications previously stated. Embryos were incubated with 1:2000 Anti-DIG-alkaline phosphatase (AP) [11093274910, Roche] and 1:2000 Anti-FITC-horse-radishperoxidase (POD) [11426346910, Roche] for 2 hours at RT and washed at least four times for 15 minutes with PBS-T. Embryos were incubated with TSA Plus Biotin [NEL749A001KT, Perkin Elmer] for 5-10 minutes, followed by a 90-minute incubation in 1:500 Alexa Fluor 647 Streptavidin conjugate [S32357, ThermoFisher Scientific] to visualise the POD signal. The AP signal was visualised by a Fast Red reaction [4210, Kem En Tec Diagnostics].

2.4 Imaging and image processing

Embryos were counterstained with DAPI [10236276001, Roche] in PBS-T (1:2000), rocked for 20 minutes at RT and stored in PBS-T at 4 °C. Embryos were flat-mounted in PBS-T on poly-L-lysine-coated coverslips and mounted with 80% glycerol. Embryos were imaged using

the Zeiss AxioZoom V16 and Zeiss LSM800 Confocal with Airyscan. Images were processed using Image J (FIJI) and Adobe Photoshop.

2.4 Bioinformatics

2.4.1 Identification of candidate genes

To identify Hox, Wnt, and frizzled gene candidates across chelicerates, I performed BLAST searches (P-value 0.05) against ten existing genomic and transcriptomic resources and four new embryonic transcriptomes (Table 1). Query sequences used in my analysis included Hox peptide sequences from *Parasteatoda tepidariorum*, *Centruroides sculpturatus* (Schwager *et al.*, 2017), *Mesobuthus martensii* (Di *et al.*, 2015), *Phalangium opilio*, and *Ixodes scapularis* (Leite *et al.*, 2018); Wnt peptide sequences from *P. tepidariorum*, *Me. martensii*, and *I. scapularis* (Janssen *et al.*, 2010); and Frizzled peptide sequences from *P. tepidariorum*, *I. scapularis*, and *Me. martensii* (Janssen *et al.*, 2015).

Table 1. Summary of chelicerates resources surveyed.

Embryonic transcriptomes were provided by Matthias Pechmann ⁽¹⁾, Lauren Sumner-Rooney ⁽²⁾, Prashant Sharma ⁽³⁾, and Natascha Turetzek ⁽⁴⁾.

BLAST	Species	Resource
BLASTP	<i>Centruroides sculpturatus</i>	Genome (used NCBI proteome) [PRJNA422877]
	<i>Limulus polyphemus</i>	Genome (used NCBI proteome) [PRJNA238073]
	<i>Stegodyphus dumicola</i>	Genome (used NCBI proteome) [PRJNA642917]
	<i>Tetranychus urticae</i>	Genome (used NCBI proteome) [PRJNA315122]
TBLASTN	<i>Acanthoscurria geniculata</i>	Embryonic transcriptome ⁽¹⁾
	<i>Charinus acosta</i>	Embryonic transcriptome ⁽²⁾
	<i>Euphrynichus bacillifer</i>	Embryonic transcriptome ⁽²⁾
	<i>Marpissa muscosa</i>	Embryonic transcriptome ⁽²⁾
	<i>Mesobuthus martensii</i>	Genome (used CDS sequence file) (Cao <i>et al.</i> , 2013)
	<i>Pardosa amentata</i>	Embryonic transcriptome ⁽²⁾
	<i>Pardosa pseudoannulata</i>	Genome [PRJNA512163]
	<i>Phalangium opilio</i>	Embryonic transcriptome ⁽³⁾
	<i>Pholcus phalangioides</i>	Embryonic transcriptome ⁽⁴⁾
	<i>Phoxichilidium femoratum</i>	Embryonic transcriptome (Ballesteros <i>et al.</i> , 2021)

The peptide sequences were predicted using the Translate ExPASy online tool (<https://web.expasy.org/translate/>; default settings). Protein sequence identity was confirmed by reciprocal BLAST against the NCBI database and the construction of maximum-likelihood trees. Where more than one sequence was identified as a potential candidate for a single

gene, nucleotide and protein alignments were inspected to eliminate the possibility that they were isoforms, individual variants, or fragments of the same gene. Only the longest isoforms and gene fragments were selected for phylogenetic analysis. All Hox (see Harper et al. (2021)) and homeobox (Supplementary Table 4) sequences identified contain a complete or partial homeodomain. The TBLASTN search in *P. phalangioides* identified a *Ubx* gene not reported previously by Leite et al. (2018). Our BLAST searches did not recover *Cs-ftz-a* (Schwager et al., 2017), *Tu-Antp-2* (Grbic et al., 2011), or *Is-Wnt8* (Janssen et al., 2010) but I included the previously identified sequences in our subsequent phylogenetic analysis. The Trinity transcript accession numbers of all identified sequences, NCBI protein accession and other sequence identifiers used in the subsequent phylogenetic analysis are found in Supplementary Table 4 and Harper et al. (2021).

2.5 Phylogenetic analysis

Hox, Wnt, Frizzled, and sensory/neurosecretory TFs (*nkx2.1/scro*, *otp*, *rax*, *vax*, *vsx/chx*) protein predictions were retrieved from NCBI for the insects' *D. melanogaster*, *Bombyx mori*, *T. castaneum*; the crustacean *Daphnia pulex*; and the onychophoran *Euperipatoides kanangrensis* (see Supplementary Table 4 and Harper et al. (2021)). The Hox, Wnt, Frizzled, and sensory/neurosecretory peptide sequences for the myriapod *Strigamia maritima* were retrieved from Chipman et al. (2014) (see Supplementary Table 4 and Harper et al. (2021)). Alignments were performed in Clustal Omega using default parameters (Sievers et al., 2011), except for the full Hox peptide sequences, which were aligned using COBALT (Papadopoulos and Agarwala, 2007). Maximum-likelihood trees were generated from whole-sequence alignments to assign sequences to families and study the relationship between candidate duplicates. Phylogenetic analyses were performed in IQ-Tree v2.0.3 (Nguyen et al., 2015) using ModelFinder to identify optimal substitution models (JTT+F+R9 for full Hox peptide sequences, LG+G4 for Hox homeodomain sequences, GTR+F+I+G4 for Hox homeobox sequences, LG+R8 for Wnt, JTT+R8 for Frizzled, JTT+F+R6 for Otp, Rax, and Vax) (Kalyaanamoorthy et al., 2017) and 100,000 bootstrap replicates. Trees were visualised in FigTree v.1.4.4 (<http://tree.bio.ed.ac.uk/software/figtree/>) and tidied in Adobe Illustrator. Alignments are provided in Supplementary Figure 44 and Harper et al. (2021).

3. Chapter Three Results: Investigating cell types and patterning of the spider pre-cheliceral region

3.1 Background

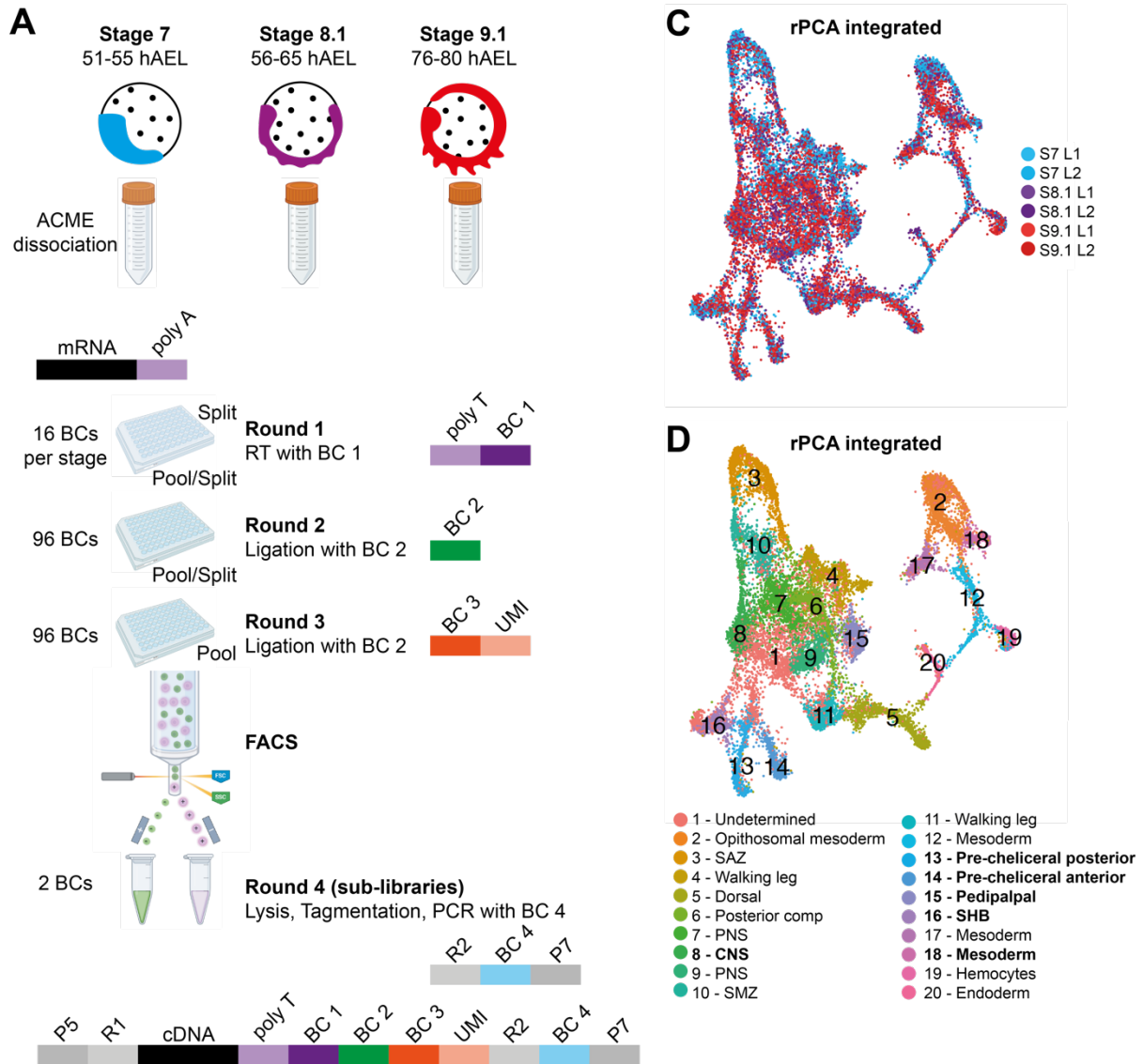
3.1.1 Forming the spider head: *hh* stripe-splitting segmentation

The non-segmental pre-cheliceral region and the appendage-bearing cheliceral and pedipalpal segments are formed by *hh* stripe-splitting segmentation during embryonic stages 5 to 8.2 (Pechmann *et al.*, 2009; Kanayama *et al.*, 2011; Akiyama-Oda and Oda, 2020). At stage 5, *otd1* and *hh1* are co-expressed at the rim of the germ disc, which represents the future anterior of the germband. At stage 6, *otd1* expression remains at the anterior margin and *hh1* expression moves to the posterior limit of the *otd1* expression domain (Figure 3). At stage 7, *otd1* moves away from the anterior margin and the anterior *hh1* stripe broadens and simultaneously or subsequently, splits into two the pre-cheliceral/cheliceral and pedipalpal *hh1* stripes (Figure 3). The anterior of the two resulting *hh1* stripes remains *otd1* positive and undergoes a second split to form the pre-cheliceral and cheliceral *hh1* stripes (Figure 3). At stage 8.2, the three stripes of *hh1* expression in the head ectoderm correspond to the posterior compartments of the pre-cheliceral, cheliceral, and pedipalpal segments. Knockdown of *otd1* results in the deletion of the pre-cheliceral and cheliceral segments, as *hh1* remains localised at the anterior rim of the germband and fails to split (Pechmann *et al.*, 2009). Furthermore, *otd1* is required to maintain *hh1* expression during the travelling process and splitting events (Kanayama *et al.*, 2011). The two splitting events are preceded by the onset of *msx1* expression, which expands between the splitting *hh1* stripe and where it is required to downregulate *hh1* (Akiyama-Oda and Oda, 2020). An ortholog of the *D. melanogaster* pair-rule gene *odd-paired* (*opa*) is consistently expressed at and near the anterior *otd1*-expressing *hh1* stripe and is involved in promoting the splitting events (Kanayama *et al.*, 2011). While *hh* and *otd1* are involved in setting up the pre-cheliceral region it is unclear how it is thereafter patterned.

3.1.2 Single-cell sequencing of three stages of spider embryogenesis

Recently, Leite *et al.* (2024) carried out single-cell RNA sequencing (scRNA-seq) on three embryonic stages (stages 7, 8.1, and 9.1) of *P. tepidariorum* (Figure 7A) because they mark the onset and/or continued progress of key developmental processes, including stripe-splitting segmentation. Leite *et al.* (2024) combined ACetic-MEthanol (ACME) dissociation (García-Castro *et al.*, 2021) and Split-Pool Ligation-based Transcriptome sequencing (SPLiT-seq) scRNA-seq (Rosenberg *et al.*, 2018) technology (Figure 7A). SPLiT-seq allows for sample multiplexing and consists of four rounds of barcoding: the first is carried out by in-cell reverse transcription (RT), the second and third by oligo ligation, and the fourth in the sub-library

amplification step (Figure 7A). The integrated dataset helped merge stages without considerable loss of stage-specific information and was comprised of 20 cell clusters, from 18,516 cells (Leite *et al.*, 2024) (Figure 7B-D).



B

	Stage 7	Stage 8.1	Stage 9.1	rPCA integrated
Cells	4824	4833	8859	18516
Median UMI counts per cell	1465	1656	1343	/
Median gene counts per cell	674	713	563	/
Markers	199	172	339	574
PCAs	7	8	12	19

Figure 7. Single-cell RNA sequencing of three embryonic stages of *P. tepidariorum*.

(A) Experimental workflow. Schematic of stage 7, 8.1 and 9.1 embryos. Independent ACME dissociations were carried out for each stage. In the first round of SPLiT-seq barcoding, ACME fixed cells for each stage were split into 16 wells, and transcripts were labelled with well-specific BCs. Thus, BC 1 was used for de-multiplexing stages. In the following split-pool rounds fixed cells were randomly distributed into new 96 well-plates. Fluorescence-Activated Cell Sorting (FACS) removed cellular debris

and sorted cells into two sub-libraries. The bottom schematic shows the final barcoded cDNA molecule. **(B)** Stages 7 and 8.1 are comparable, whereas stage 9.1 had fewer median UMI and gene counts per cell, but more cells overall. Marker genes were predicted with an in-cluster-versus-all-others approach, including genes that were expressed in $\geq 25\%$ of cells in their respective cluster and an adjusted p-value return threshold of $1e-5$. As expected, the transcriptomic complexity, as determined by PCA, increased as development progressed, with merged containing the most. **(C&D)** UMAP visualisation for all stages merged with reciprocal PCA (rPCA) integration with 1.3 threshold for variable genes and 40 anchors, coloured by the library **(C)** and cluster identity **(D)**, with those I analysed (bold). Other abbreviations: BC: barcode; CNS: central nervous system; hAEL: hours after egg laying; PNS: peripheral nervous system; SAZ: segment addition zone; SHB: spider head boundary; SMZ: segment maturation zone; UMI: unique molecular identifier.

The Hox gene *labial* (*lab*) marks the tritocerebral segment in all arthropods (Hughes and Kaufman, 2002). In *P. tepidariorum*, the *lab-A* ohnolog is the earliest Hox gene to be expressed in the pedipalpal segment and *lab-A* knockdown results in loss of the pedipalps and increased cell death in the remaining tissue of the pedipalpal segment (Pechmann *et al.*, 2015). In insects, loss of *lab* orthologs also results in increased cell death in the intercalary segment as well as adjacent head segments. While the general role of *lab* orthologs is similar in spiders and insects, the tritocerebral segment of insects is reduced and appendage-less. Divergent co-factors of *lab* orthologs or regulatory targets downstream of *lab* orthologs are likely responsible for differences in segment morphology (Schomburg *et al.*, 2020). To identify interaction partners of *lab-A*, Schomburg *et al.* (2020) used a candidate gene screen in *P. tepidariorum* with *D. melanogaster* genes known or predicted to interact with *lab* or to be expressed in the intercalary segment. However, this approach failed to identify large numbers of factors implicated in the specification and formation of the pedipalps. In the integrated dataset, clusters 15 and 18 were marked by *lab-A*, suggesting cells in these clusters relate to the pedipalp segment.

In arthropods, the deutocerebral segment and the non-segmental anterior head tissue lack Hox gene expression. In the integrated dataset, six clusters (12, 13, 14, 16, 19, and 20) did not have Hox marker genes or express any Hox gene, suggesting that there are populations of cells in *P. tepidariorum* embryos that are not patterned by Hox genes at stages 7 to 9.1 (Leite *et al.*, 2024). Clusters 12, 19, and 20 were analysed by collaborators and based on known markers and new candidate genes studied by ISH likely represent mesoderm, haemocyte, and endoderm cell populations, respectively (Leite *et al.*, 2024) (Figure 7D). The remaining three clusters (13, 14 and 16) were marked by known anteriorly expressed genes suggesting they relate to the non-segmental pre-cheliceral region and/or cheliceral segment. scRNA-seq transcriptomics can provide an unbiased identification of key marker genes

involved in pre-chelicer patterning, chelicer and pedipalp development, stripe-splitting segmentation, and the initiation of eye development.

3.1.3 Chapter three aims

This chapter aimed to verify the identity of five putative anterior clusters in scRNA-seq integrated dataset (Leite *et al.*, 2024) using known markers and interrogate these clusters for new candidate genes involved in spider head development in *P. tepidariorum*. I expect to identify clusters corresponding to the pre-chelicer region as well as chelicer and pedipalpal segments.

- Examine known markers and the spatial expression of new candidate genes for putative pedipalpal clusters (15 and 18) by colourimetric ISH.
- Examine known markers and the spatial and relative expression of new candidate genes for putative pre-chelicer and/or chelicer clusters (13, 14 and 16) by colourimetric ISH and dFISH, respectively.
- Identify genes significantly upregulated in *so1*-expressing cells.

Data in this chapter was published in Leite *et al.* (2024) except for *croc* and *tup* ISH, and dFISH of *Pax6.1* and *six3.1* which are unpublished. *otd2*, *so1*, and *Pax2.1* ISH expression and dFISH of *otd2* and *Pax6.2* were published in Baudouin Gonzalez *et al.* (2024). Interpretations of the clusters may differ slightly from Leite *et al.* (2024) due to the analysis of additional cluster markers.

3.2 Results: Clusters 15 and 18: Pedipalpal segment

Clusters 15 and 18 were marked by 43 and 73 genes, respectively (Leite *et al.*, 2024). In *P. tepidariorum*, five Hox genes are expressed in the pedipalpal segment (Figure 8A) and were markers for clusters 15 and/or 18, suggesting cells in this cluster relate to this segment (Figure 8B). Therefore, to further characterise clusters 15 and 18, I analysed the early embryonic expression of one of the top cluster 15 markers *latrophilin cir1* [g14495], and the top two cluster 18 markers, *integrin-alpha-8* [g23098] and *neuroactin* [g29023], and another cluster 18 marker *FoxC/crocodile (croc)* [g24984] from stages 7 to 9.

Some aspects of *croc* expression were described by Schomburg *et al.* (2020) and Janssen *et al.* (2022) however, the role of *croc* in anterior head patterning in the spider was unknown. The expression of the other three markers had not been previously characterised in this spider or any other chelicerate.

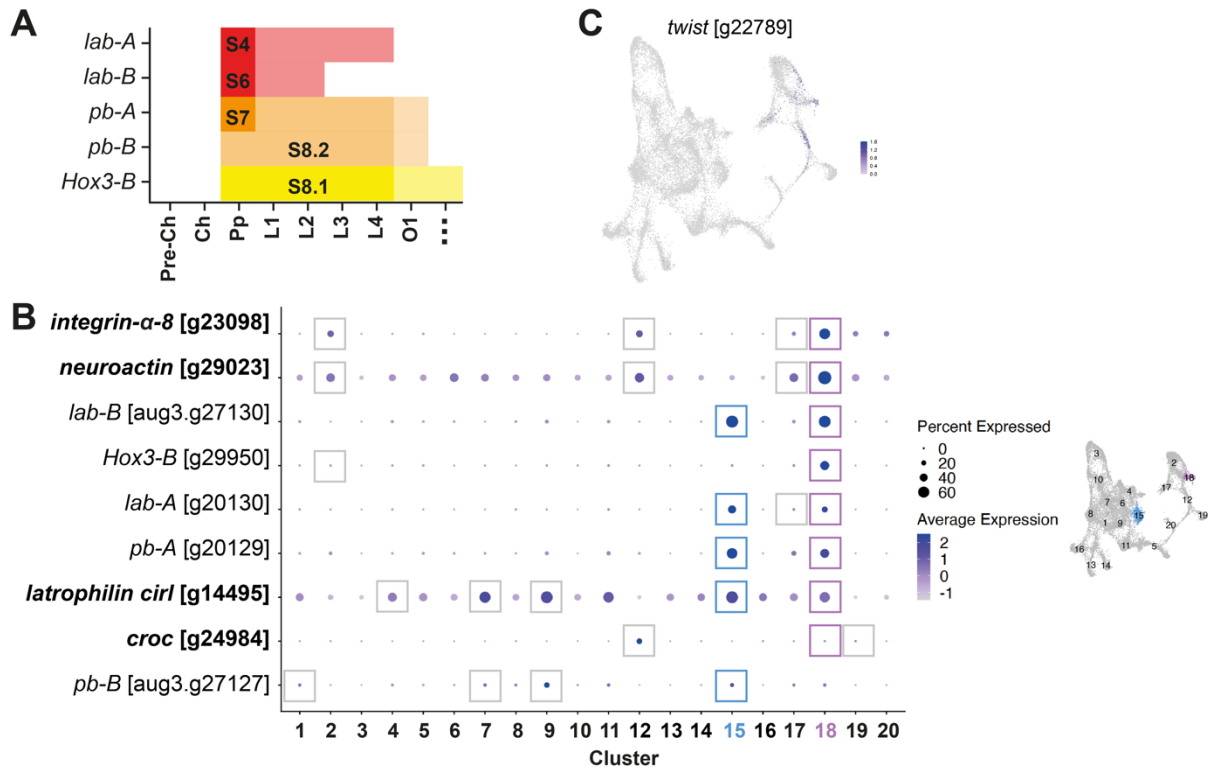


Figure 8. Marker genes for clusters 15 and 18.

(A) Expression of anterior Hox genes in *P. tepidariorum*. For each gene, the initial and strongest expression domain (darkest colour), expanded domain (medium colour), and further late expansions (lightest colour) are shown. Adapted from Schwager et al. (2017). **(B)** Dotplot of Hox markers and those I analysed by ISH (bold) for clusters 12 (green), 15 (blue box) and 18 (pink box). In order of cluster 18. Clusters markers were also markers for one or more other clusters (grey box): 1 [Undetermined], 2 [Opisthosomal mesoderm], 4 [Leg], 7 [PNS], 9 [PNS], 12 [Pc and Ch mesoderm], 17 [Leg mesoderm], and 19 [Hemocyte]. **(C)** UMAP of *twist*.

latrophilin cirl was expressed at stage 7 throughout the prosoma, but most strongly in the nascent L4 segment (Figure 9A). At stage 8, *latrophilin cirl* was expressed in the nascent pedipalp, L1, L2 and L4 segments (Figure 9B). At stage 9, it was weakly expressed in cells on either side of the ventral midline that likely become part of the nervous system (Figure 9C).

integrin-alpha-8 was expressed at stage 7 in the SAZ and nascent walking leg segments (Figure 9D). At stage 8, *integrin-alpha-8* was additionally expressed in the mesoderm of and weakly anterior to the nascent pedipalpal segment, however, at this stage expression was lost from the SAZ (Figure 9E). At stage 9, *integrin-alpha-8* was no longer detectable in the prosomal and was only weakly expressed in the opisthosomal segments (Figure 9F).

neuroactin was expressed at stage 7, in a broad medial stripe and then at stage 8 this expression domain split into three stripes representing the nascent L1, L2 and L3 segments (Figure 9G&H). *neuroactin* was additionally weakly expressed in and anterior to the nascent

pedipalpal segment (Figure 9H). At stage 9, *neuroactin* was expressed in the prosoma and opisthosoma (Figure 9I).

At stage 7, *croc* was expressed in a broad stripe at the anterior rim of the germ band (Figure 9J). By stage 8.1, *croc* was expressed in the pre-cheliceral lobes, most strongly in an anterior medial axe-shaped domain, as well as the cheliceral and pedipalpal segments (Figure 9K). At stages 7 and 8.1, *croc* was expressed in an evenly dispersed “salt-and-pepper” pattern, suggesting that they may represent mesodermal precursor cells, which has not been reported previously by Schomburg et al. (2020) or Jannsen et al. (2022). Note that in the beetle *T. castaneum*, *croc* marks the anterior mesoderm anlagen at early germ rudiment stages (Economou and Telford, 2009). At stage 9, *croc* was expressed in domains on either side of the stomodeum primordium and towards the posterior boundary of the pre-cheliceral region (Figure 9L).

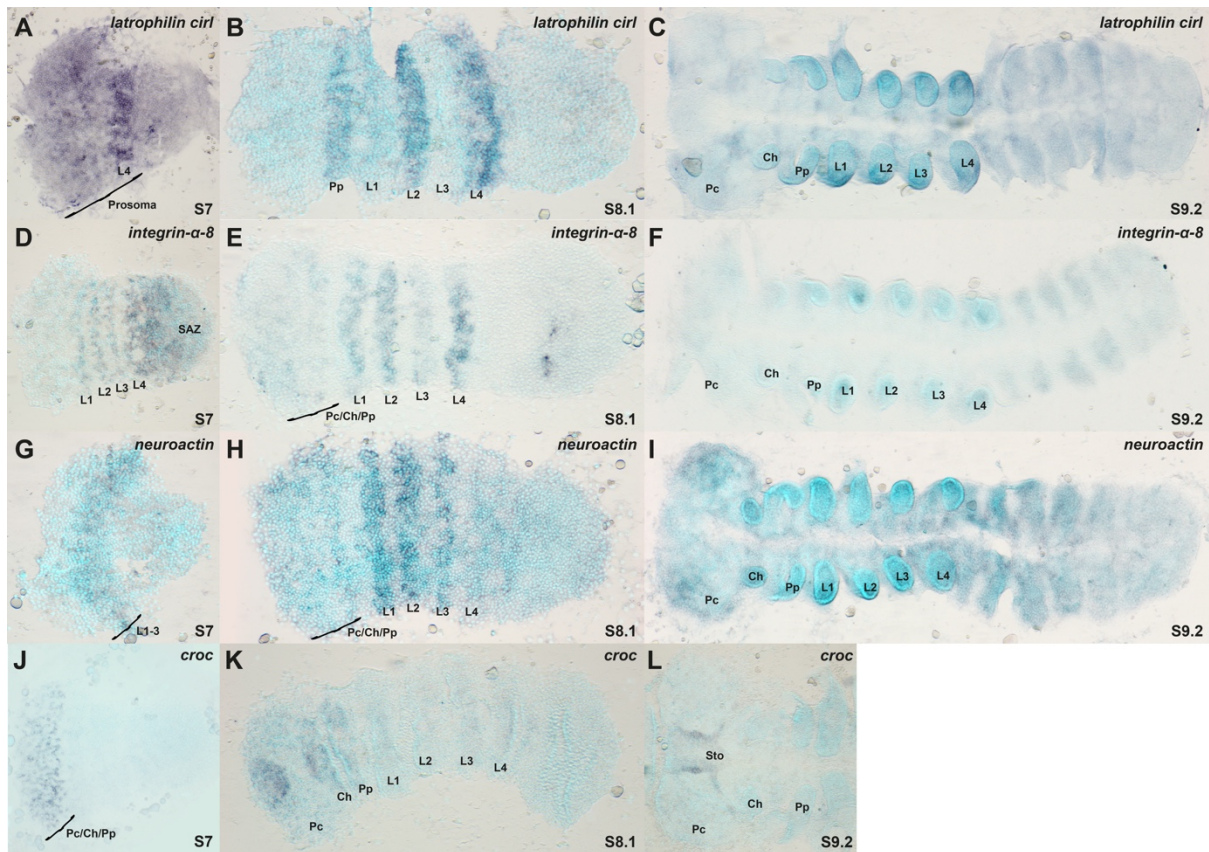


Figure 9. Expression of cluster 15 and 18 marker genes in *P. tepidariorum* embryos.

Embryos were flat mounted with the anterior to the left. Stages prior to the onset of expression were omitted. Embryos stained for *latrophilin cirl* transcripts at stages 7 (A), 8.1 (B), and 9.2 (C); *integrin-α-8* transcripts at stages 7 (D), 8.1 (E), and 9.2 (F); *neuroactin* transcripts at stages 7 (G), 8.1 (H) and 9.2 (I); *croc* transcripts at stages 7 (J), 8.1 (K) and 9.2 (L). Other abbreviations: Ch: chelicerae; Pp: pedipalp; Pc: pre-cheliceral; SAZ: segment addition zone; Sto: stomodeum; L1-4: walking legs 1-4.

Interestingly *twist* [g22789], a mesodermal marker (Yamazaki *et al.*, 2005), was not a marker gene in the integrated dataset. However, it was expressed in cells of clusters 2, 12, 17 and 18 (Figure 8C), indicating that at least some of these cells are mesodermal (Leite *et al.*, 2024). Overall, my analysis of the spatial expression of cluster 15 and 18 markers suggests they likely contain cells that correspond to non-mesodermal and partially mesodermal cells of the pedipalpal segment, respectively.

3.3 Cluster 13: Pre-cheliceran posterior

Cluster 13 had 55 marker genes (Leite *et al.*, 2024) including four out of five Wnt genes known to be expressed in the pre-cheliceran region in *P. tepidariorum* (Janssen *et al.*, 2021): *Wnt5* [g2035], *Wnt8* [g19404], *Wnt7.2* [g2043] and *Wnt16* [g26171] (Figure 10A&B). The fifth Wnt gene expressed in the head, *Wnt2* [g17472], was not a marker gene in the integrated dataset but was expressed in cells of cluster 13 (Figure 10C). To better characterise cluster 13, I first analysed the early embryonic expression of the top three markers: *Pax6.1* [g12873], *Pax6.2* [g12868], and *optomotor-blind-3* (*omb3*) [aug3.g3790] as well as *orthodenticle-2* (*otd2*) [g5047] and *Tbx2/3* [aug3.g3745] from stages 7 to 9 (Figure 10A&B).

Expression of the *P. tepidariorum* *Pax6* paralogs and *otd2* was previously characterised during late embryonic head development, from stage 10.2 to 14, to investigate their role in eye development (Schomburg *et al.*, 2015). Although both *Pax6* paralogs were expressed in the neurogenic ectoderm of the pre-cheliceran lobes, the pattern was difficult to resolve, and neither was expressed in the developing eye primordia (Schomburg *et al.*, 2015). *otd2* was expressed in the PEP, pre-cheliceran neurogenic ectoderm, and along the ventral midline (Schomburg *et al.*, 2015). Earlier expression of the *Pax6* paralogs and *otd2* was not previously characterised, so their role in anterior head patterning in the spider was unknown. The expression of *omb3* and *Tbx2/3* had not been previously characterised in this spider or any other chelicerate.

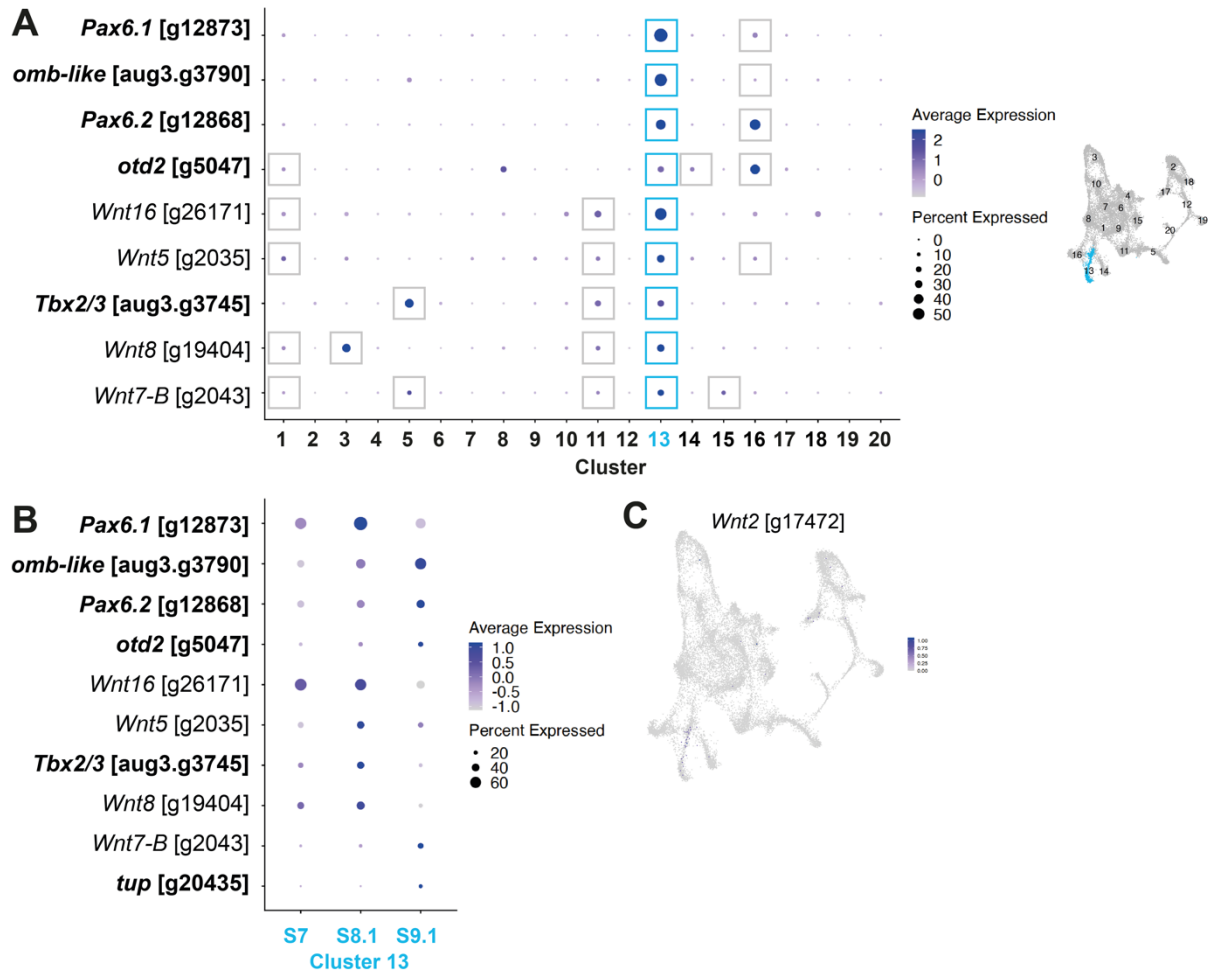


Figure 10. Cluster 13 markers.

Dotplot (A) and temporal expression (B) of Wnt markers and those I analysed by ISH (bold) for cluster 13 (blue box), in order. Cluster 13 markers were also markers for other clusters (grey box): clusters 1 [Undefined], 3 [SAZ], 5 [Dorsal], 11 [Leg], 14 [Pc anterior], 15 [Pedipalp], and 16 [Pc posterior]. (C) UMAP of *Wnt2*.

I found *Pax6.2* was first expressed at stage 6 followed by *Pax6.1* at stage 7, in stripes at the anterior rim of the germ band (Figure 11A&D). Expression of the *Pax6* paralogs shifted to a more posterior position in the nascent germ band and split down the midline into two complex medial lateral domains in the pre-cheliceral lobes (Figure 11B, C, E, & F). At stage 9, the *Pax6* paralogs were expressed in segmental clusters along the CNS (Figure 11C&F).

omb3 was expressed from early stage 8, in two neurogenic posterior lateral patches in the pre-cheliceral lobes (Figure 11G-I). Later *omb3* was expressed in the stomodeum primordium and all appendages, including the opisthosomal limb buds (Figure 11I).

otd2 was first expressed from stage 8.2 in a stripe in the posterior of the pre-cheliceral region (Figure 11J). At stage 9, this domain transformed into two neurogenic domains at the rim of the pre-cheliceral lobes and a posterior medial neurogenic domain, connected to *de novo* expression along the ventral midline (Figure 11K).

Tbx2/3 was weakly expressed at stage 7 in two lateral stripes about four cell rows from the rim of the germ band (Figure 11L). At stage 8, this expression domain expanded to two posterior lateral neurogenic patches, and *Tbx2/3* was also expressed along the ventral midline and dorsal regions of the prosomal appendages (Figure 11M). At stage 9, *Tbx2/3* was also expressed in the opisthosomal appendages but the pre-cheliceral domain became reduced (Figure 10B; Figure 11N).

My analysis of marker genes indicated that cluster 13 represents a neurogenic cell population in the posterior lateral pre-cheliceral region, likely contributing to the central protocerebrum.

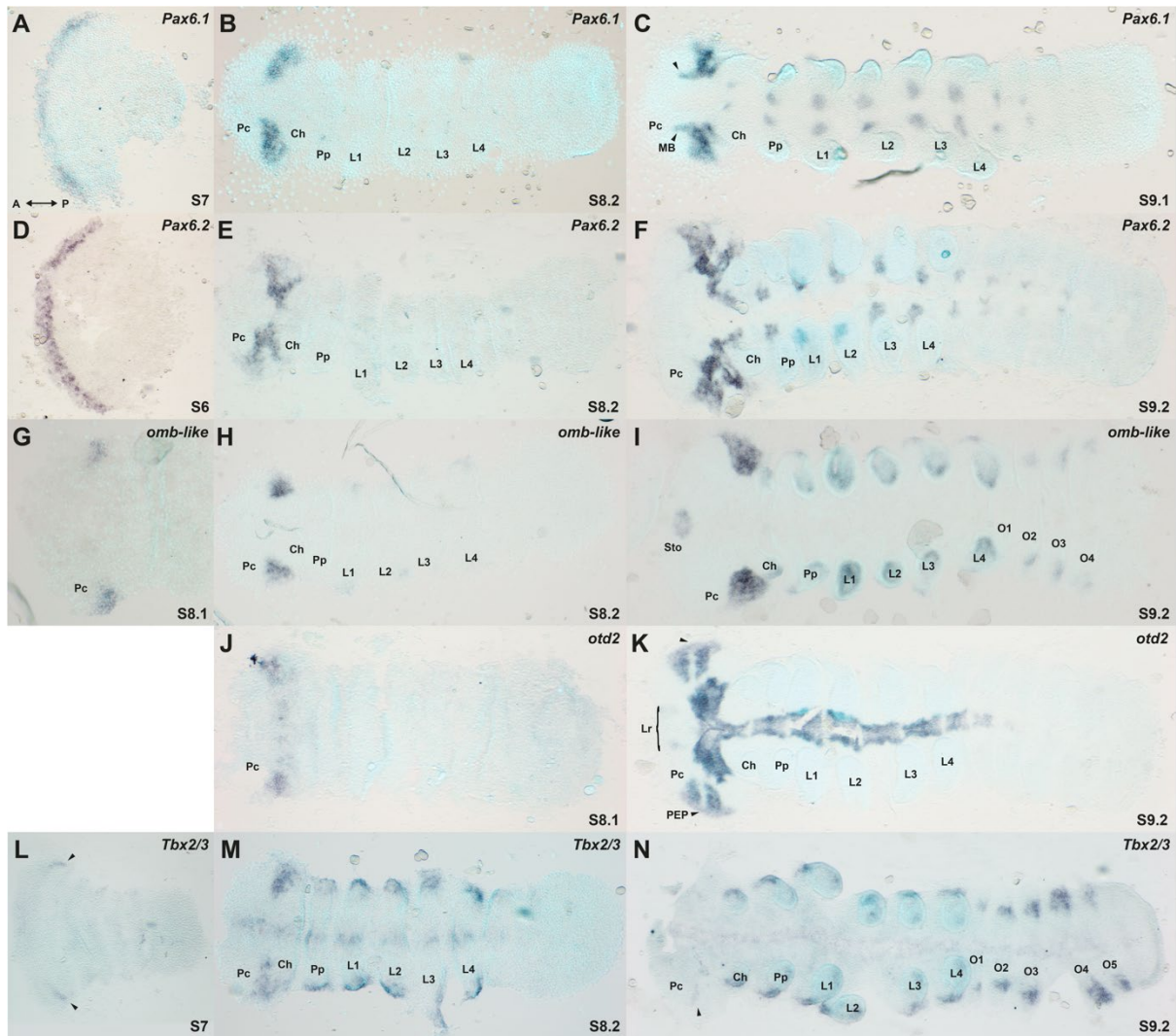


Figure 11. Expression of cluster 13 marker genes in *P. tepidariorum* embryos.

Embryos were flat mounted with the anterior to the left. Stages prior to the onset of expression were omitted. Embryos stained for *Pax6.1* transcripts at stages 7 (A), 8.2 (B), and 9.2 (C); *Pax6.2* transcripts at stages 6 (D), 8.2 (E), and 9.2 (F); *omb3* transcripts at stages 8.1 (G), 8.2 (H) and 9.2 (I); *otd2* transcripts at stages 8.1 (J) and 9.2 (K); *Tbx2/3* transcripts at stages late 7 (L), 8.1 (M) and 9.2 (N). Other abbreviations: Ch: chelicerae; Lr: labrum; MB: mushroom body; Pp: pedipalp; Pc: pre-cheliceral; PEP: principal eye primordia; Sto: stomodeum; L1-4: walking legs 1-4.

3.3.1 Pax6 paralogs mark the spider ocular segment

In the *T. castaneum* early blastoderm, the *Pax6* orthologs *eyeless* (*ey*) and *twin of eyeless* (*toy*) are co-expressed in the ocular segment of the pre-antennal region. This ocular segment gives rise to significant parts of the protocerebrum including the MB and the visual system (Luan *et al.*, 2014). To determine whether *P. tepidariorum* has an ocular segment, I next looked at the relative expression of *Pax6.1* and *Pax6.2* from stages 7 to 9.

Consistent with the single ISH (Figure 11A&D), I found *Pax6.2* comes on before the onset of *Pax6.1* (Figure 12A). From the onset of *Pax6.1* expression to early stage 8, *Pax6.1* was expressed at the anterior limit of the *Pax6.2* expression domain (Figure 12B). At late stage 8, the *Pax6* paralogs were expressed in different cells of the pre-chelicerai region with only minimal overlap in the medial region (Figure 12C). This suggests at the gene expression level, spiders do have an ocular segment.

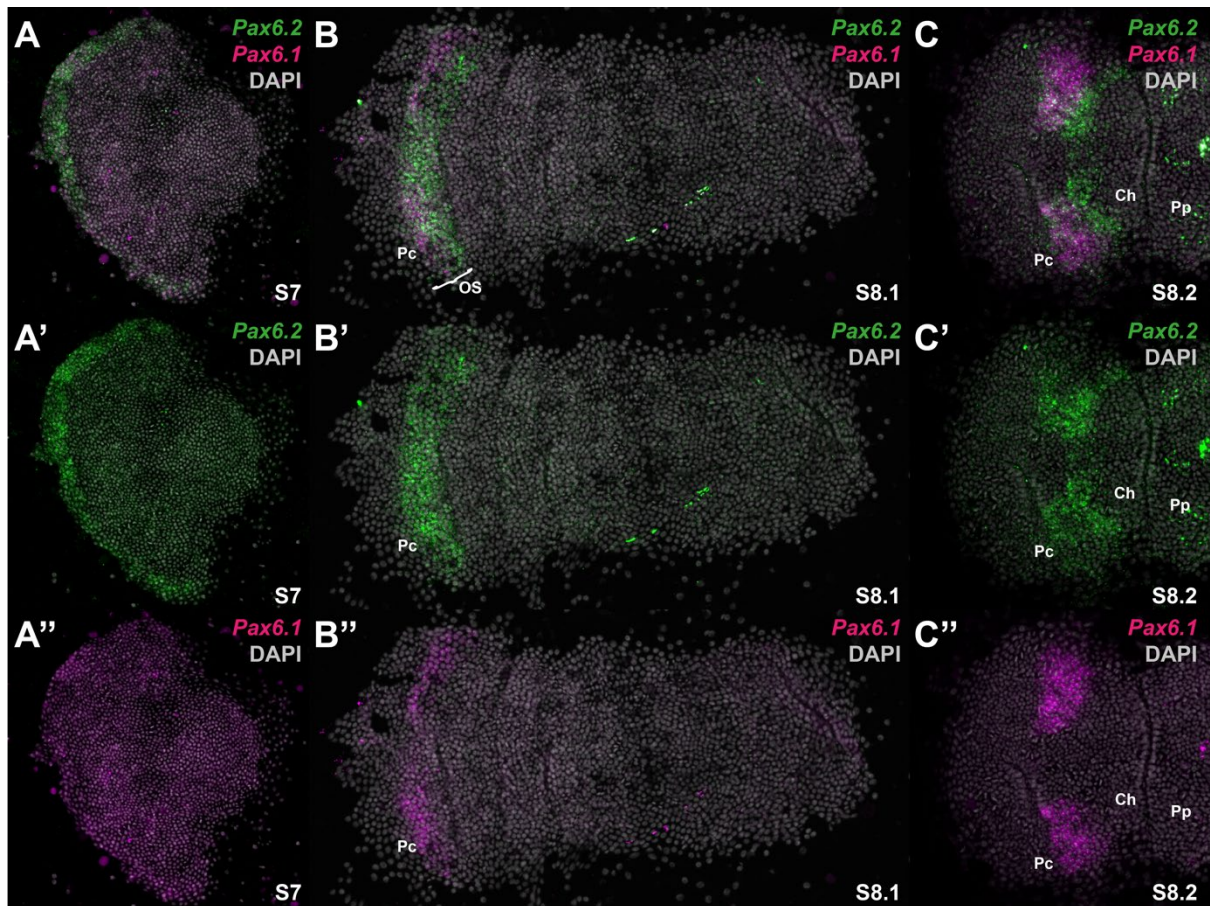


Figure 12. Expression of *Pax6.1* and *Pax6.2* in the developing head of *P. tepidariorum* embryos. Embryos were flat mounted with the anterior to the left. Embryos stained for *Pax6.2* (green) and *Pax6.1* (magenta) transcripts and nuclei/DNA (DAPI: grey) at stages 7 (**A**), 8.1 (**B**), and 8.2 (**C**). Other abbreviations: Ch: chelicerai; Pp: pedipalp; Pc: pre-chelicerai; OS: ocular segment.

3.3.2 *Pax6.1* is not expressed in the PEP

To determine whether the spider ocular segment gives rise to the PEP, I looked at the relative expression of the *Pax6.2* and the PEP marker *otd2* (Schomburg *et al.*, 2015) from stages 8 to 10.1.

I found *Pax6.2* and *otd2* were expressed in different cells in the pre-cheliceral lobes with only minimal overlap in the neurogenic ectoderm of the developing brain (Figure 13A-C), this is consistent with fewer cells expressing *otd2* in cluster 13 (Figure 10A&B).

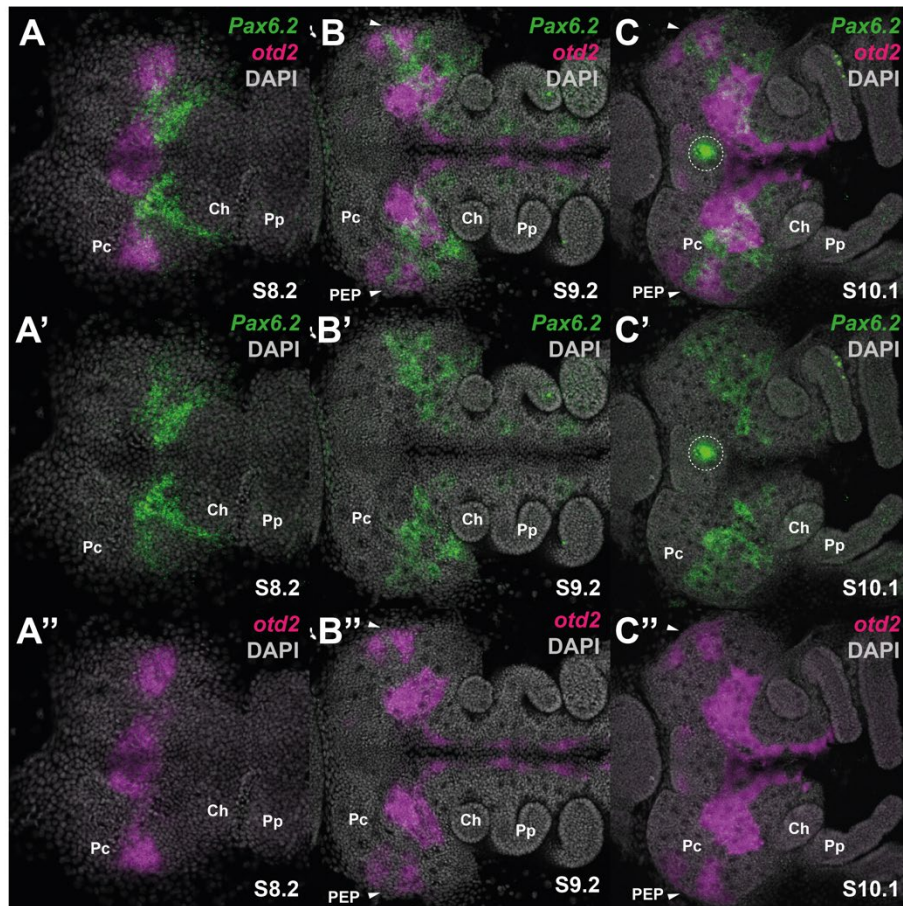


Figure 13. Expression of *Pax6.2* and *otd2* in the developing head of *P. tepidariorum* embryos.

Embryos were flat mounted with the anterior to the left. Background staining (dotted line). Embryos stained for *Pax6.2* (green) and *otd2* (magenta) transcripts and nuclei/DNA (DAPI: grey) at stages 8.2 (A), 9.2 (B), and 10.1 (C). Other abbreviations: Ch: chelicerae; Pp: pedipalps; PEP: principal eye primordia; Pc: pre-cheliceral.

3.3.3 *Pax6* paralogs are not involved in the formation of the cheliceral and pedipalpal segments

To better understand the relationship and phasing of cluster 13 and 16 markers and determine whether *Pax6* paralogs are involved in stripe-splitting segmentation, I looked at the relative expression of most posteriorly expressed *Pax6* paralog *Pax6.2* (cluster 13 and 16) with respect to the spider head segmentation gene *hh1* (cluster 16) from stages 7 to 9.

I found *Pax6.2* was expressed at the anterior rim of the germ band after the anterior *hh1* stripe had travelled away from the anterior rim of the germ band (Figure 14A&B). From the onset of its expression to stage 8, *Pax6.2* remained anterior to and abutting the travelling and splitting anterior *hh1* stripe (Figure 14B-D). From stage 6, *hh1* is expressed at the posterior boundary of *otd1* (cluster 16) stripe (Figure 3) (Kanayama *et al.*, 2011). Thus, *Pax6.2* expression overlaps with *otd1* expression from stage 6 to stage 8 and explains why *Pax6.2* was also a marker for cluster 16 (Figure 10A). However, unlike *otd1*, which is co-expressed with *hh1* from stage 5 (Pechmann *et al.*, 2009; Kanayama *et al.*, 2011), *Pax6.2* is not involved in stripe-splitting segmentation. At late stage 8, the *Pax6.2* overlaps with *hh1* expression marking the lateral posterior boundary of the pre-chelicerous region (Figure 14D).

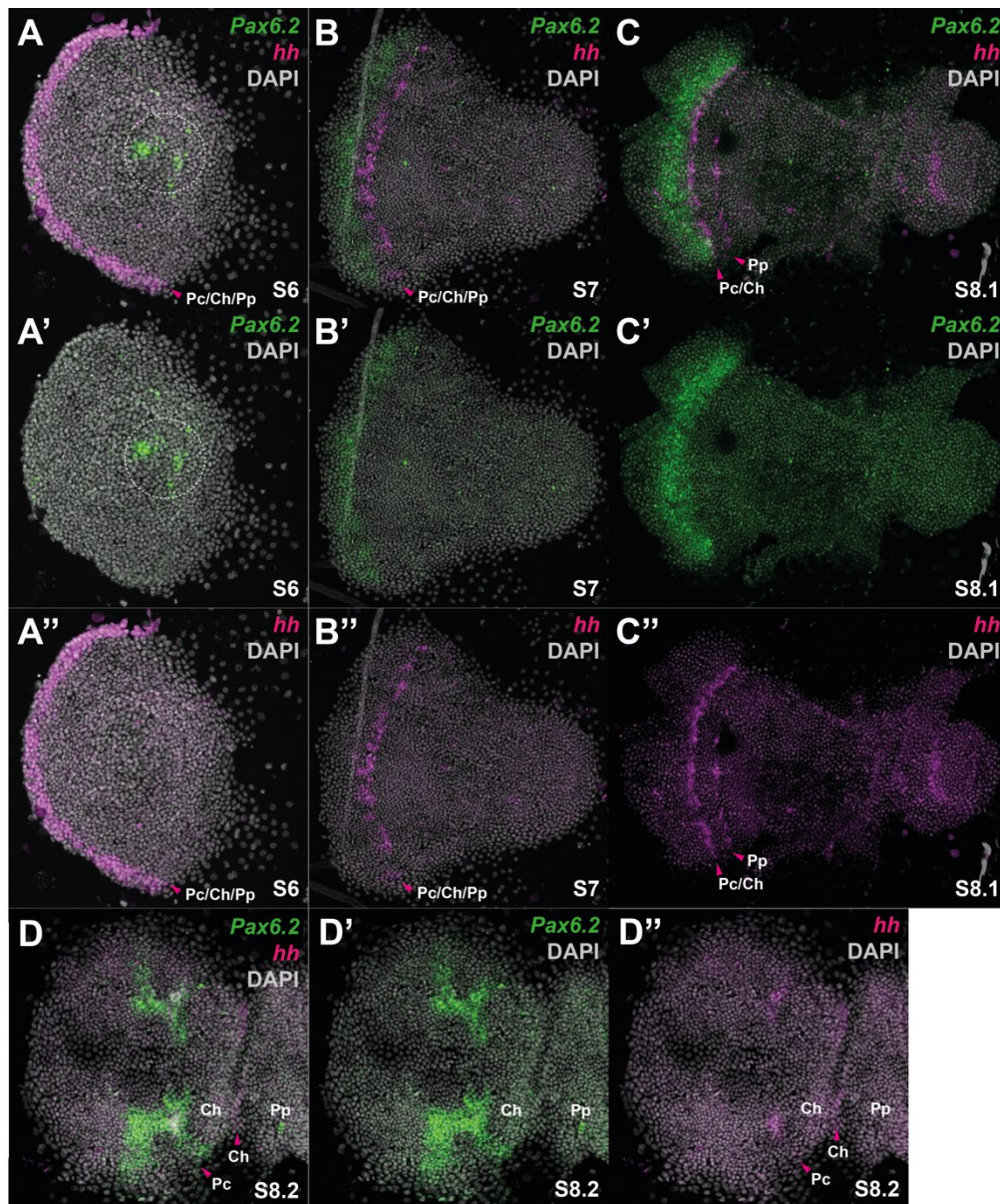


Figure 14. Expression of *Pax6.2* and *hh1* in the developing head of *P. tepidariorum* embryos. Embryos were flat mounted with the anterior to the left. Anterior *hh1* stripes (magenta arrowheads) and background staining (dotted line). Embryos stained for *Pax6.2* (green) and *hh1* (magenta) transcripts

and nuclei/DNA (DAPI: grey) at stages 6 **(A)**, 7 **(B)**, 8.1 **(C)**, and 8.2 **(D)**. Other abbreviations: Ch: chelicerae; Pp: pedipalp; Pc: pre-chelicerar.

3.4 Cluster 14: Pre-chelicerar anterior

Cluster 14 had 61 marker genes (Leite *et al.*, 2024), including *FoxA1/forkhead1* (*FoxA1/fkh1*) [g3001] which is involved in stomodeum development in the spider (Feitosa *et al.*, 2017) (Figure 15A&B). The cluster 15 and 18 marker *otd2* was also marker for cluster 14, and was expressed in the labrum from stage 9 (Figure 15B). *foxQ2* [g224] which is involved in labrum development in the spider (Schacht *et al.*, 2020), was not a marker in the integrated dataset but was expressed in cells of cluster 14 (Figure 15C). To better characterise cluster 14, I first analysed the early embryonic expression of the three top markers: *six3.1* [g1245], *six3.2* [g25543], and *visual system homeobox* (*vsx*)/*chx* [aug3.g27186] as well as *islet/tailup* (*tup*) [g20435] from stages 7 to 9 (Figure 15A&B).

six3.1 and *six3.2* expression was previously characterised during late embryonic head development, from stage 10.2 to 14, to investigate their role in eye development (Schomburg *et al.*, 2015). *six3.1* is not expressed in the primordia of the eyes while *six3.2* is expressed from stage 12 in all primordia of the secondary eyes (Schomburg *et al.*, 2015). Both *six3* paralogs are expressed in the labrum. Earlier expression of *six3.1* and *six3.2* was not previously characterised, so their role in anterior head patterning in the spider was unknown. The expression of *vsx/chx* and *tup* had not been previously characterised in this spider or any other chelicerate.

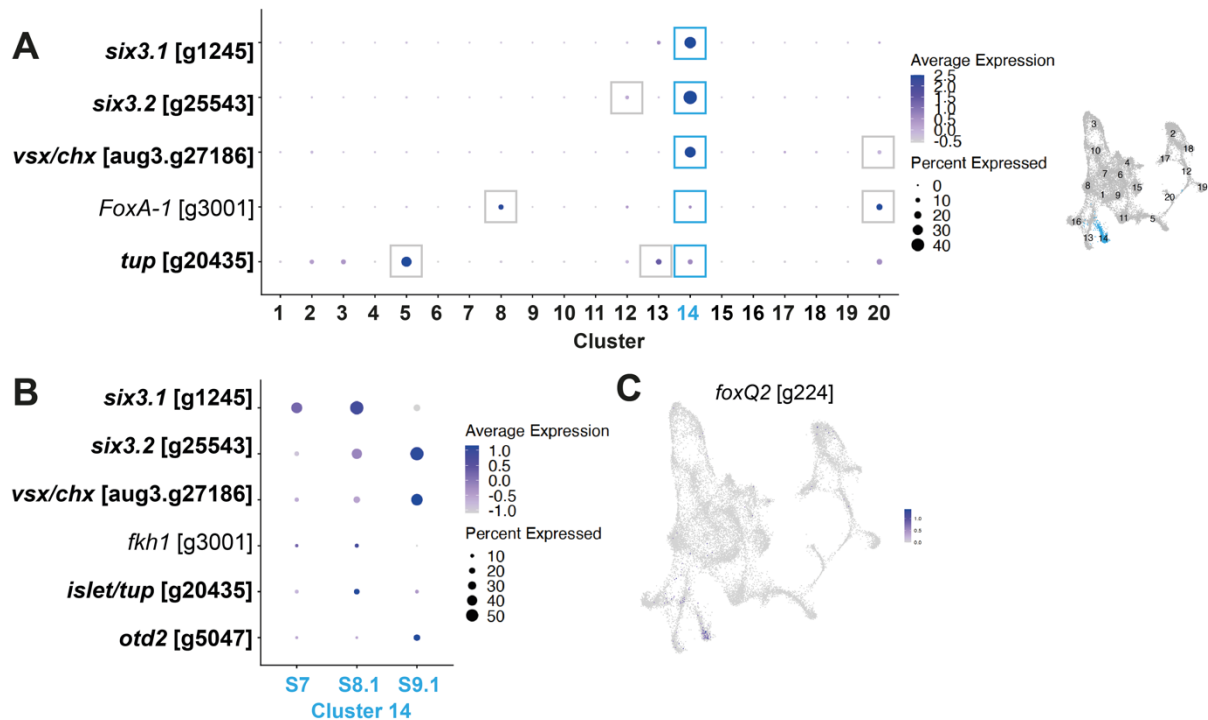


Figure 15. Cluster 14 markers.

Dotplot **(A)** and temporal expression **(B)** of known markers and those I analysed by ISH (bold) for cluster 14 (blue box), in order. Cluster 14 markers were also markers for other clusters (grey box): clusters 5 [Dorsal], 8 [CNS], 12 [Pc and Ch mesoderm], 13 [Pc posterior], and 20 [Endoderm]. **(C)** UMAP for *foxQ2*.

I found that *six3.1* was first expressed at stage 7 in a stripe along the anterior rim of the germ band (Figure 16A). This stripe broadened and then at stage 9, *six3.1* was expressed in the labrum, at the anterior rim of the pre-chelicerall lobes, and a posterior lateral spot (Figure 16B-D). This suggests *six3.1* has an early role in patterning the anterior pre-chelicerall region.

six3.2 was first expressed at early stage 8 in two antero-lateral domains (Figure 16E). The *six3.2* expression domain expanded, and from stage 9 onwards was detected in a bat-shaped domain in the anterior pre-chelicerall region, encompassing the labrum (Figure 16F&G).

vsx/chx was first weakly expressed at stage 8 in segmental patches along the ventral midline (Figure 16H). At stage 9, *vsx/chx* was also expressed in bilateral circular domains in the neurogenic ectoderm at the anterior rim of the germ band joined by weaker labral C-shaped expression (Figure 16I&J).

islet/tup expression was first seen at stage 8.1, in a circular domain at the anterior medial rim of the germ band and two medial-lateral neurogenic stripes in the pre-chelicerall lobes (Figure 16K). At stage 9, the circular domain refined into two domains, one at the anterior rim of the developing labrum and one surrounding the stomodeum primordia, and the neurogenic domain expands into a triangular shape (Figure 16L). *islet/tup* was also a marker for cluster 13 (Figure 15A), suggesting its neurogenic domain overlaps with cluster 13 markers. Additional expression of *islet/tup* around the germ band of the opisthosoma segments (Figure 16K&L). Based on the early onset of expression, I suggest *islet/tup* acts high in the hierarchy of labrum development in *P. tepidariorum*.

My analysis of marker genes indicated that cluster 14 represents anterior pre-chelicerall cell populations that include the developing stomodeum, labrum, and anterior neurogenic ectoderm.

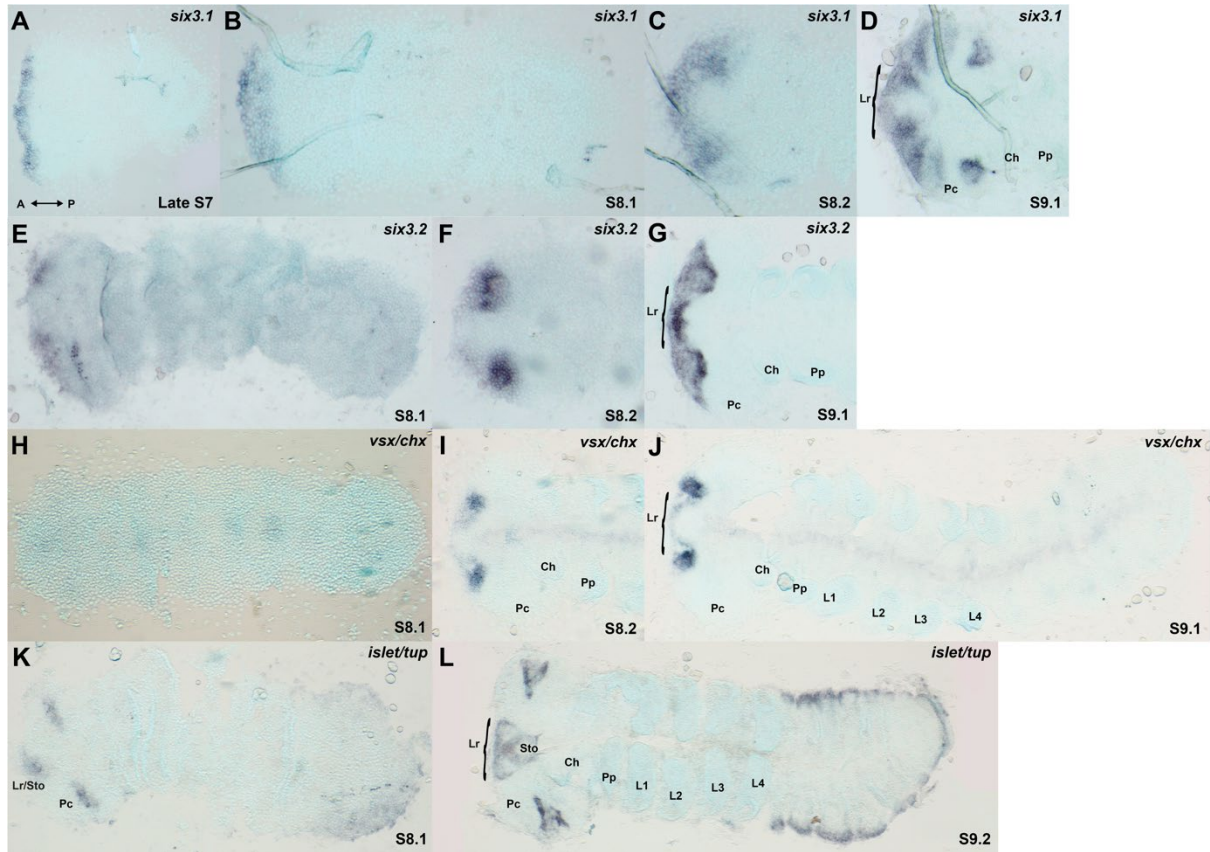


Figure 16. Expression of Cluster 14 marker genes in *P. tepidariorum* embryos.

Embryos were flat mounted with the anterior to the left. Stages prior to the onset of expression were omitted. Embryos stained for *six3.1* transcripts at stages 7 (A), 8.1 (B), 8.2 (C), and 9.1 (D); *six3.2* transcripts at stages 8.1 (E), 8.2 (F), and 9.1 (G); *vsx/chx* transcripts at stages 8.1 (H), 9.1 (I), and 9.2 (J); *tup* transcripts at stages 8.1 (K) and 9.2 (L). Other abbreviations: Ch: chelicerae; Lr: labrum; Pp: pedipalp; Pc: pre-chelicerar; Sto: stomodeum; L1-4: walking legs 1-4.

3.4.1 The *Pax6*-expressing ocular segment and the *six3*-expressing anterior median region represent early independent units in the pre-chelicerar region

To better understand the relationship and phasing of cluster 13 and 14 markers, I looked at the relative expression of the most anteriorly expressed *Pax6* paralog, *Pax6.1* (cluster 13), and the earliest expressed *six3* paralog, *six3.1* (cluster 14).

six3.1 was expressed at the anterior rim of the germ band once the *Pax6.1* travelled away from the anterior rim of the germ band (Figure 17A). *six3.1* and *Pax6.1* were expressed in different cells in the pre-chelicerar lobes (Figure 17A-D). *Pax6.1* is expressed

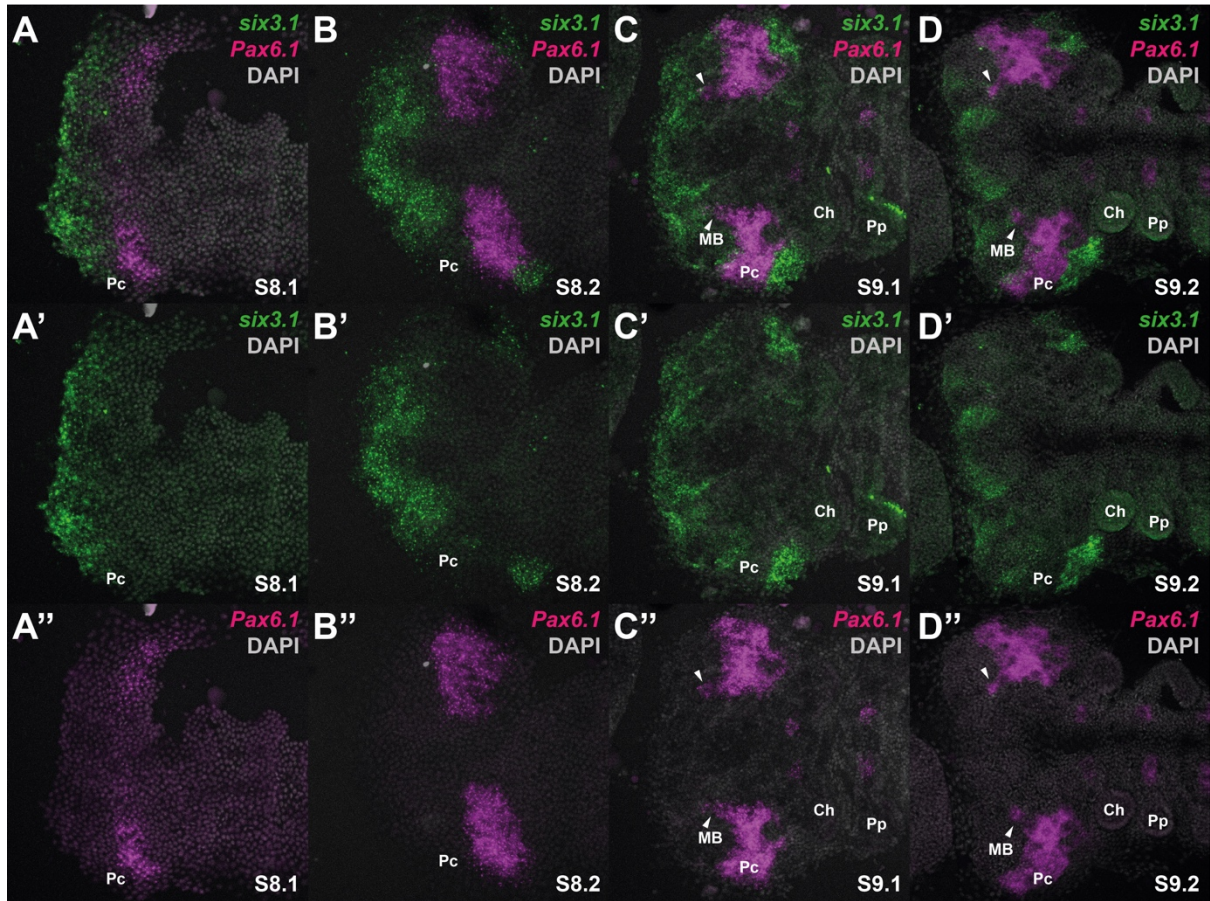


Figure 17. Expression of *six3.1* and *Pax6.1* in the developing head of *P. tepidariorum* embryos. Embryos were flat mounted with the anterior to the left. Embryos stained for *six3.1* (green) and *Pax6.1* (magenta) transcripts and nuclei/DNA (DAPI: grey) at stages 8.1 (A), 8.2 (B), 9.1 (C), and 9.2 (D). Other abbreviations: Ch: chelicer; MB: mushroom body; Pp: pedipalp; Pc: pre-chelicer.

3.5 Cluster 16: Pre-chelicer/chelicer boundary

Cluster 16 had 45 marker genes (Leite *et al.*, 2024) including known markers involved in stripe-splitting segmentation: *otd1* [g9628], *hh1* [g23071], *msx1* [g11815] and *opa* [g16937] (Pechmann *et al.*, 2009; Kanayama *et al.*, 2011; Akiyama-Oda and Oda, 2020) as well as the cluster 13 markers *Pax6.1*, *Pax6.2*, *omb3*, *otd2*, and *Wnt5* (Figure 18A&B). To better characterise cluster 16, I analysed the early embryonic expression of a top marker *lim1a* [g12191] from stages 7 to 9 (Figure 18A&B). The expression of *lim1a* had not been previously characterised in this spider or any other chelicerate.

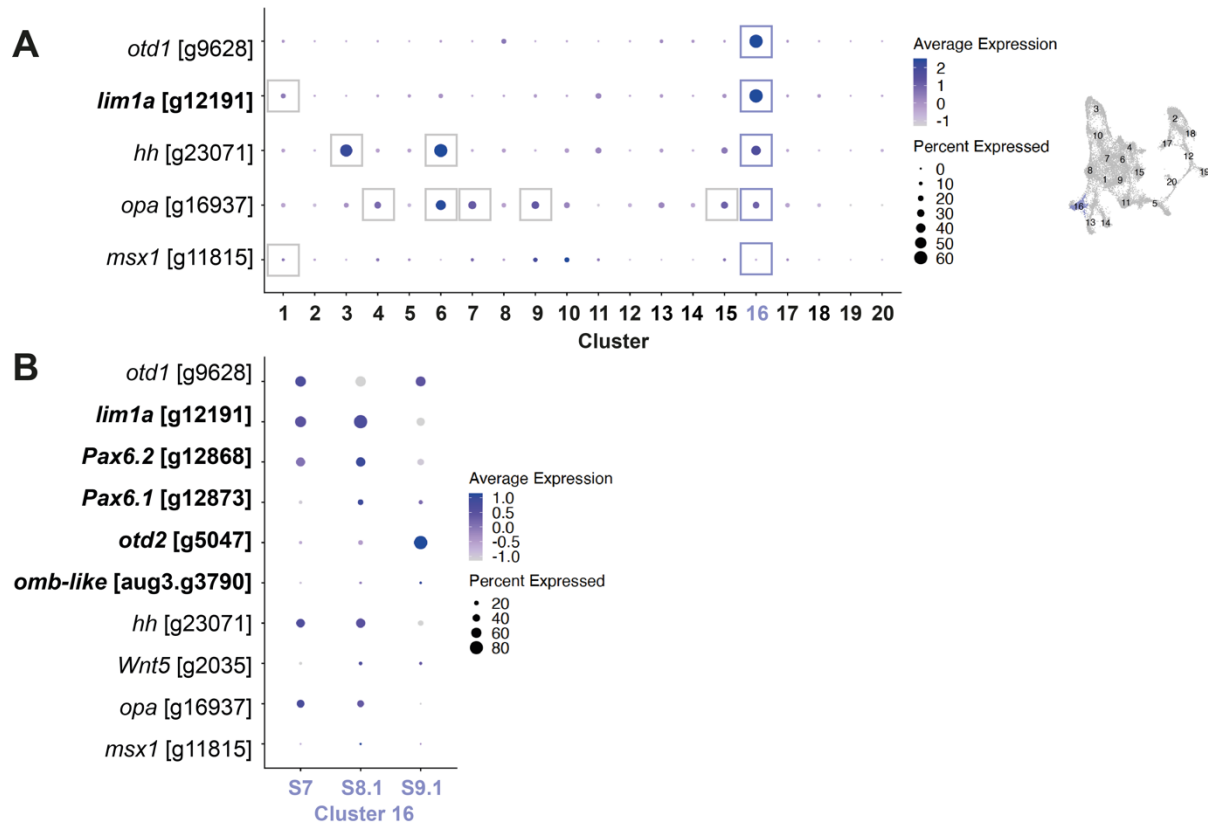


Figure 18. Cluster 16 markers.

Dotplot **(A)** and temporal expression **(B)** of stripe-splitting markers and those I analysed by ISH (bold) for cluster 16 (lilac box), in order. Cluster 16 markers were also markers for other clusters (grey box): 1 [Undetermined], 3 [SAZ], 4 [Leg], 6 [Posterior comp], 7 [PNS], 9 [PNS], and 15 [Pedipalps].

I found *lim1a* was first expressed at stage 7 about five cell rows from the rim of the germ band (Figure 19A). The *lim1a* stripe then travelled to a more posterior position, broadened, and subsequently encompassed a *lim1a*-negative domain, exhibiting a pattern similar to an open mouth (Figure 19B&C). This domain simultaneously split down the midline into two lateral stripes in the anterior cheliceral segment (Figure 19C&D). At stage 9, *lim1a* was also expressed in rings at the base of the developing appendages (Figure 19D).

My analysis of *lim1a* and temporal expression of spider stripe-splitting segmentation genes in cluster 16 (Figure 18B) suggests this cluster represents pre-cheliceral/cheliceral boundary.

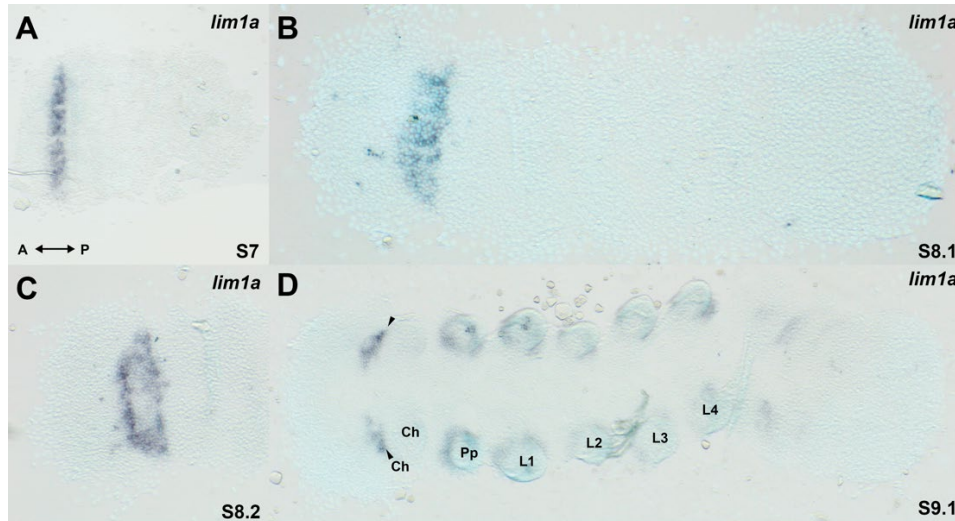


Figure 19. Expression of Cluster 16 marker genes in *P. tepidariorum* embryos.

Embryos were flat mounted with the anterior to the left. Stages prior to the onset of expression were omitted. Embryos stained for *lim1a* transcripts at stages 7 (A), 8.1 (B), 8.2 (C), and 9.1 (D). Other abbreviations: Ch: chelicerae; Pp: pedipalp; Pc: pre-chelicer; L1-4: walking legs 1-4.

3.5.1 *lim1a* is likely involved in the formation of the chelicer segment

To determine whether *lim1a* was involved in stripe-splitting segmentation, I looked at the relative expression of *lim1a* and the spider head segmentation gene *hh1* from stages 7 to 9.

At stage 7, *lim1a* was expressed anterior to and abutting the anterior *hh1* stripe before the first splitting event (Figure 20A). Thus, *lim1a* expression overlaps with *otd1* and *Pax6.2* (cluster 13) expression at stage 7 (Figure 18B). At late stage 8, *lim1a* was expressed between the splitting pre-chelicer and chelicer *hh1* stripes (Figure 20B). This suggests *lim1a* may downregulate *hh1*, similar to *msx1*, during stripe-splitting segmentation (Akiyama-Oda and Oda, 2020). At stage 9, the *lim1a* expression marks the anterior portion of the chelicer segment and partially overlaps with *hh1* expression marking the lateral posterior boundary of the pre-chelicer region (Figure 20C).

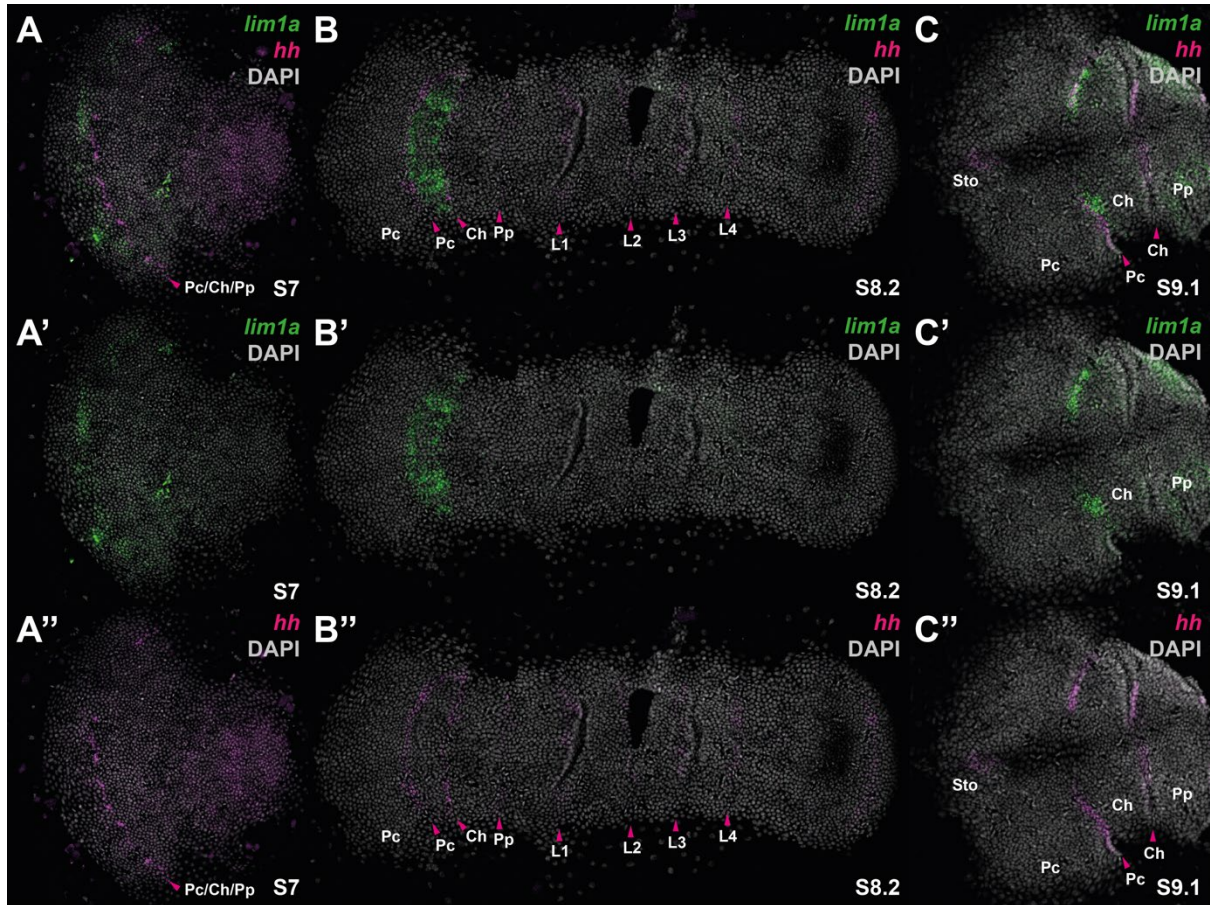


Figure 20. Expression of *lim1a* and *hh1* in the developing head of *P. tepidariorum* embryos.

Embryos were flat mounted with the anterior to the left. Stages prior to the onset of expression were omitted. *hh1* stripes (magenta arrowheads). Embryos stained for *lim1a* (green) and *hh1* (magenta) transcripts and nuclei/DNA (DAPI: grey) at stages 7 (**A**), 8.2 (**B**), and 9.1 (**C**). Other abbreviations: Ch: chelicerae; Pp: pedipalps; Pc: pre-chelicerar; Sto: stomodeum; L1-4: walking legs 1-4.

3.6 Differentially expressed genes between *so1*⁺ cells VS *so1*⁻ cells

While neurogenic clusters were identified, we were unable to definitively relate clusters to the non-neurogenic ectoderm that also forms at the anterior and lateral rim of the pre-chelicerar lobes (Leite *et al.*, 2024). Identifying these cells could give further insights into the differentiation of the non-neurogenic ectoderm including eye development.

In *P. tepidariorum*, the developing eye primordia can be visualised by *sine oculis 1* (*so1*) [g6476] expression from stage 10.2 to 14 (Schomburg *et al.*, 2015) and is functionally required for the development of all eyes (Gainett *et al.*, 2020). *so1* was not a marker gene in the integrated dataset and was only expressed in 702 cells (4%) (Figure 21A). I found *so1* was expressed at stage 9, at the perimeter of the pre-chelicerar lobes (Figure 21B&C), which is earlier than previously reported (Schomburg *et al.*, 2015). Later this expression domain split into two, corresponding to the PEP and SEP.

To identify new candidate genes for eye development, I used the FindMarkers() function to identify differentially expressed genes between *so1*-positive cells and *so1*-negative cells across all clusters and all stages. Including genes that were expressed in at least 25% of cells. I identified two candidates, *Arrowhead (Awh)* [g21739] and *Irregular chiasmC-roughest (IrreC-rst)* [g992] (Supplementary Table 3; Figure 21D). *Awh* was a marker for cluster 9, the PNS cluster, and overlaps with *so1* expression at the base of each appendage (Leite *et al.*, 2024). Expression of *IrreC-rst* had not been previously characterised in this spider or any other chelicerate. To determine whether *IrreC-rst* has a potential role in eye development in *P. tepidariorum*, I analysed the embryonic expression of *IrreC-rst* from stage 7 onwards.

I found *IrreC-rst* was first expressed at stage 9 at the rim of the developing head lobes, where the PEP and SEP are established (Figure 21E&F). However, at later embryonic stages *IrreC-rst* expression is difficult to resolve (Figure 21G&H) therefore, I am unable to make any meaningful conclusions on its role in eye development.

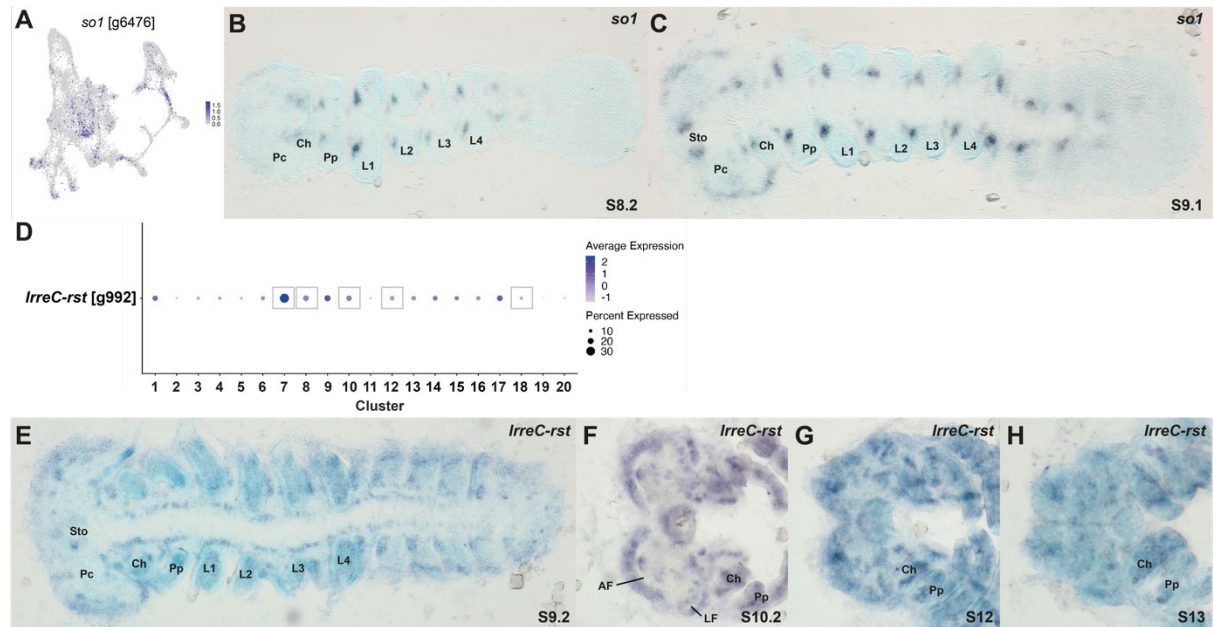


Figure 21. Expression of *IrreC-rst* in *P. tepidariorum* embryos.

(A) UMAP of *so1*. (D) Dotplot of *IrreC-rst*. *IrreC-rst* was a marker for several clusters (grey box): 7 [PNS]; 8 [CNS]; 10 [SMZ]; 12 [Pc and Ch mesoderm]; and 18 [Pedipalps]. (B,C,E-H) Embryos were flat mounted with the anterior to the left. Stages prior to the onset of expression were omitted. Embryos stained for *so1* transcripts at stages 8.2 (B) and 9.1 (C) and *IrreC-rst* transcripts at stages 9.2 (E), 10.2 (F), 12 (G), and 13 (H). Other abbreviations: AF: anterior furrow; Ch: chelicerous; LF: lateral furrow; Pp: pedipalp; Pc: pre-chelicerous; Sto: stomodeum; L1-4: walking legs 1-4.

3.7 Chapter Three Discussion

In this chapter, I verified the identity of five putative anterior clusters from scRNA-seq analysis on three embryonic stages 7, 8.1, and 9.1 of *P. tepidariorum* (Leite *et al.*, 2024) using known

markers and new candidate genes. The anterior neuroectoderm of many if not all bilaterian animals is partitioned into an anterior *six3*-positive tissue and a more posterior *otd/otx* marked region (Steinmetz *et al.*, 2010). Unsurprisingly, *six3*- and *otd*-expressing cells fell into separate clusters. *six3.1*- and *six3.2*-expressing cells fell into cluster 14, marking the anterior pre-chelicerar region. While *otd2*- and *otd1*-expressing cells expressing cells fell into clusters 13 and 16, respectively. Cluster 13 was further marked by the *Pax6* paralogs and Wnt genes, suggesting it marks the posterior pre-chelicerar region. Cluster 16 was also marked by *hh1* morphogen, suggesting it marks the pre-chelicerar/chelicerar boundary, which I suggest calling the spider head boundary (SHB) (see General Discussion Figure 41). Two clusters, 15 and 18, corresponded to the developing pedipalpal segment. To my surprise, I did not identify a cluster specifically corresponding to the chelicerar segment. This work provided new insights into pre-chelicerar patterning and the development of neuropils involved in processing visual information in the spider. In addition to identifying new candidate genes labrum development and stripe-splitting segmentation.

3.7.1 Cluster 15 and 18: Pedipalpal segment

I did not identify any genes specific to the pedipalp segment. *pb-A* was a marker for clusters 15 and 18, the top marker of the former, and was previously found to be differentially expressed between the pedipalps and walking legs (Schomburg *et al.*, 2020). It may be worth further expression analysis of cluster 15 markers to identify further interaction partners of *lab-A*.

Cluster 18 corresponded to a partially mesodermal cell population of the pedipalp segment. In a previous study, one of the top cluster 18 markers, *integrin-alpha-8*, was also identified as a marker for a cluster at stage 5 and was expressed in cells that form a mesodermal cell lineage at the germ disc periphery (Akiyama-Oda *et al.*, 2022). Another cluster 18 marker *croc* was expressed in a similar pattern to the *T. castaneum* *croc* ortholog (Economou and Telford, 2009; Kittelmann *et al.*, 2013), initially marking the anterior mesoderm primordium and later expressed in two domains on either side of the stomodeum primordium (Schomburg *et al.*, 2020; Janssen *et al.*, 2022). In the spider, *croc* may act upstream of *FoxA1/fkh1* in stomodeum development (Feitosa *et al.*, 2017), similar to *T. castaneum* (Kittelmann *et al.*, 2013) however, this would require functional confirmation.

3.7.2 Cluster 14: Pre-chelicerar anterior, insights into arcuate body development, *pars intercerebralis*, and labrum development

Homology between the insect central body and chelicerate arcuate body, midline neuropils of the protocerebrum, has been challenged because of differences in the positions in the brain and internal neuroarchitecture (Homborg, 2008). In *T. castaneum*, *six3* knockdown results in

the loss of the central body in an otherwise normal brain (Posnien *et al.*, 2011b). This suggests *six3* is required in the neuroectoderm for specifying the identity of central body neuroblasts however, tools to genetically trace the offspring of these neuroblasts are required (Posnien *et al.*, 2011b). I found that *P. tepidariorum* *six3.1* and then *six3.2* were expressed at the anterior germband during early embryogenesis and before the morphogenic movements in the pre-chelicerel region that form the anterior furrows, the neural primordium of the arcuate body (Doeffinger *et al.*, 2010). In *P. tepidariorum*, *six3.2* is later expressed in and surrounding the anterior furrow (Schomburg *et al.*, 2015). This provides molecular evidence that the spider arcuate body, like the insect central complex, is derived from *six3*-expressing neuroectoderm. Previously, in *P. tepidariorum*, parental RNAi against each *six3* ohnolog was performed to investigate their role in labrum development (Schacht *et al.*, 2020). In the spider, neither *six3* ohnolog had a deleterious effect on the AMR (Schacht *et al.*, 2020), in contrast to *six3* knockdown in *T. castaneum* (Posnien *et al.*, 2011b). This was suggested to be the result of functional redundancy thus, an observable phenotypic effect on the AMR may require a combinatorial knockdown of both *six3* ohnologs (Schacht *et al.*, 2020).

In the insects' *D. melanogaster* and *T. castaneum*, and the centipede *S. maritima* the PI is derived from bilateral *vsx/chx*-expressing placodes in the *six3*-expressing neuroectoderm (de Velasco *et al.*, 2007; Posnien *et al.*, 2011b; Hunnekuhl and Akam, 2014). I identified bilateral *vsx/chx*-expressing domains in the *six3*-expressing neuroectoderm, likely corresponding to the spider PI. This is the first indication such a cell population exists in a chelicerate. I further characterised the molecular signature of the spider PI in Chapter Four. Cluster 14 likely corresponds in part to the insect anterior archicerebrum from which the central complex and neuroendocrine centers develop (Posnien *et al.*, 2023).

Here I show that the *P. tepidariorum* *tup/islet* was expressed in a similar pattern to the *T. castaneum* ortholog (Kittelmann, 2012), at the anterior rim of the developing labrum and surrounding the stomodeum primordia. In the beetle, *tup/islet* knockdown abolishes the development of labral structures (Kittelmann, 2012), similar to the knockdown of *cnc* (Coulcher and Telford, 2012). Although it remains to be tested, *tup/islet* was suggested to function upstream of *cnc* in the beetle by being necessary for the formation of the anterior fold, which separates the ectoderm of the head lobes during early embryogenesis (Kittelmann, 2012). This suggests *tup/islet* may act downstream of *foxQ2* (Schacht *et al.*, 2020) in labrum development in the spider.

Future work in *P. tepidariorum*, would involve parental RNAi against *tup/islet*. In *tup/islet* knockdown embryos, I would expect the labrum to be strongly reduced or lost, similar to *foxQ2* knockdown in the spider (Schacht *et al.*, 2020). Furthermore, I would expect a reduction or loss of gene expression domains that mark the developing labrum. For example, *vsx/chx* which I found to be expressed in the developing labrum and is downstream of *cnc* in

labrum development in the beetle (Kittelmann *et al.*, 2013). The expression of spider stomodeal markers, for example, *FoxA1/fkh1* and *hh1* (Feitosa *et al.*, 2017), may also be reduced or lost.

tup/islet and *cnc* orthologs are retained in a single copy in spiders and the scorpion *Centruroides sculpturatus* (Sharma *et al.*, 2014; Aase-Remedios *et al.*, 2023) and likely reflect an early loss of the duplicate in the arachnoplumonate ancestor following WGD. However, *tup/islet* and *cnc* orthologs have not been surveyed across other arachnoplumonate lineages such as amblypygids, which have retained high numbers ohnologs of key developmental gene families (Harper *et al.*, 2021). Studying the expression of *cnc* and *tup/islet* orthologs in other chelicerates and arthropod outgroups, such as Onychophora, will shed light on labrum development in the arthropod common ancestor.

3.7.3 Cluster 13: Pre-cheliceral posterior and insights into the initiation of eye development and MB development

In *T. castaneum*, *ey* and *toy* co-expressed in the ocular segment of the pre-antennal region, which gives rise to the visual system and significant parts of the protocerebrum including the MB (Luan *et al.*, 2014). The ocular segment does not represent a serially homologous segment, the main argument for this is that it lacks an anterior parasegment boundary that would separate it from the non-segmental anterior head tissue (Posnien *et al.*, 2023).

I found that the *P. tepidariorum* *Pax6* paralogs are co-expressed in a stripe in the early pre-cheliceral region however, the extent of co-expression is subtle compared to broad co-expression of *ey* and *toy* in insects. Identifying a spider ocular segment suggests cells in the emerging eye primordia experienced earlier expression of the *Pax6* paralogs and, therefore, possible gene regulatory cell fate competence priming by the two factors (Friedrich, 2022). This idea is supported by the fact that *ey* and *toy* act as top-level field selector genes in the insect ocular segment rather than visual system primordia-specific selector genes (Luan *et al.*, 2014).

I have shown that *so1* and *eya* are peripherally expressed in the pre-cheliceral lobes (Baudouin Gonzalez *et al.*, 2024). In *D. melanogaster*, *so* and *eya* are part of a gene regulatory module that suppresses *ey* after the initiation of retinal differentiation (Atkins *et al.*, 2013). Thus, the absence of spider *Pax6* paralogs in the early non-neurogenic precursor cell population could be the outcome of repression by peripherally expressed transcription factors such as *eya* as *so1* as well as *otd2* shown here.

In *P. tepidariorum*, *Pax6.1* expression (Baudouin-Gonzalez *et al.*, 2022) is later expressed in distinct elevations in the developing brain, the medial subdivisions which represent the neural primordium of the spider MB (Doeffinger *et al.*, 2010; Samadi *et al.*, 2015). Here I show that this domain is free of *six3*-expression. Similarly, *D. melanogaster* MB

neuroblasts (MBNBs) in seem to lie outside of the *six3*-expressing protocerebrum (Posnien *et al.*, 2023; Kunz *et al.*, 2012). Furthermore, the fly's MBNBs originate from adjacent neuroectodermal progenitors within the same proneural domain (Kunz *et al.*, 2012), this mode of neurogenesis parallels observations of MB development in spiders (Doeffinger *et al.*, 2010). Moreover, *six3* knockdown in *T. castaneum* does not have much an effect on MB development (Posnien *et al.*, 2011b).

In contrast to other chelicerates, the spider MB are visual neuropils of the secondary eyes (Long, 2021). In spiders, the PEP migrate medially with the overgrowing non-neurogenic ectoderm eventually ending up on top of the visual neuropil primordia (Schomburg *et al.*, 2015). At stage 13.2, *P. tepidariorum* *Pax6.1* expression looks to be directly beneath *so1* expression marking the PEP (Baudouin-Gonzalez *et al.*, 2022; Baudouin Gonzalez *et al.*, 2024). Thus, the spider ocular segment likely contributes to the development of visual neuropils associated with the secondary and principal eyes. So, while neither *Pax6* paralog is expressed in the eye primordia, other components of the spider visual system likely originate from the same early embryonic field of *Pax6*-expressing cells as in other arthropods.

Future work in spiders would involve confirming the role of the spider ocular segment in head, eye, and MB development. My preliminary data in *P. tepidariorum* (not shown) suggests that to see an observable phenotypic effect on pre-chelicerar head development likely requires a combinatorial knockdown of both *Pax6* paralogs. Furthermore, the MB of *P. tepidariorum* are reduced (Long, 2021) and thus, not the best species to study the development of spider MB.

Spiders have two *Pax2* paralogs, which are likely ohnologs retained from the arachnoplumonate WGD event (Janeschik *et al.*, 2022). In *P. tepidariorum* and the more distantly related haplogyne spider *Pholcus phalangioides* *Pax2.1* is expressed in the SEP before and after splitting (Janeschik *et al.*, 2022). *Pax2.1* [g1094] was also a marker for cluster 13 in the integrated dataset, however, was expressed at low levels (Supplementary Figure 43). Unlike the *Pax6* paralogs, neither *Pax2* paralog was expressed in the spider ocular segment (Supplementary Figure 43). Furthermore, unlike *so1*, *Pax2.1* was not expressed at the periphery of the pre-chelicerar lobes (Supplementary Figure 43). Thus, *Pax2.1* does not appear to be involved in the initiation of secondary eye development as suggested by Janeschik *et al.* (2022). Instead *Pax2.1* looks to have an early role in patterning the nascent chelicerar, pedipalpal, walking leg, and first opisthosoma segments (Supplementary Figure 43), not previously reported by Janeschik *et al.* (2022).

Medina-Jiménez *et al.* (2024a) recently performed live-cell dissociation and droplet-based scRNA-seq on pooled stage 10-12 embryos of *P. tepidariorum*, with the processed dataset only containing 4103 cells. At stage 10, the PEP and SEP have separated and at stage 12 the SEP split into three discernible domains which correspond to each lateral eye

pair (Schomburg *et al.*, 2015). Recently, this scRNA-seq dataset was re-analysed and a small cluster (C32) of 39 cells and 25 marker genes (including *so1*, *eya*, *six3.2*, *dac2*, and *Pax2.1*) was identified corresponding to the developing secondary eyes (Medina-Jiménez *et al.*, 2024b). The other top cluster 13 marker in this study, *omb3*, was also a marker for C32 and was expressed in the developing secondary eyes after splitting (Medina-Jiménez *et al.*, 2024b). The top C32 marker *hh2* was found to be expressed in the SEP before and after splitting (Medina-Jiménez *et al.*, 2024b). Embryonic expression of *hh2* across spider families confirmed Hh signaling is involved in eye development in spiders and insects (Medina-Jiménez *et al.*, 2024b). While new candidates were identified for spider eye development, no cluster corresponding to the developing principal eyes was identified nor were individual clusters corresponding to each secondary eye pair.

3.7.4 Cluster 16: Pre-cheliceral/cheliceral boundary and insights into stripe-splitting segmentation

The SHB, also the head generative zone (Kanayama *et al.*, 2011), separates the ocular segment from the cheliceral segment and corresponds in position to the vertebrate midbrain-hindbrain boundary (MHB) primordia (Posnien *et al.*, 2023) (see General Discussion Figure 41). The SHB, marked by *hh1*, is the first parasegments boundary to become visible and remains the most anterior one (Pechmann *et al.*, 2009). At the germ disc stage, several genes are positively regulated by Hh signaling, including *otd1*, *fgf8*, *opa*, and *lab-A*, and negatively regulated by Hh signaling, including *msx1* and *cubitus interruptus* (Akiyama-Oda and Oda, 2020). Here I show that *lim1a* and *Pax6.2* are initially expressed in stripes parallel to the SHB and later have their anterior or posterior expression boundary in that region, respectively (see General Discussion and Figure 41). Taken together, the SHB has an important embryonic signaling function like the IHB.

Spiders and *C. sculpturatus* have retained *lim1/5* duplicates from the arachnoplumonate WGD event (Aase-Remedios *et al.*, 2023). Here, I show *P. tepidariorum* *lim1a* homolog was expressed in a similar pattern to the *T. castaneum* *lim1/5* ortholog, initially in the deutocerebral segment later this domain shrinks, and additional expression is seen in the limb buds (Posnien *et al.*, 2011b). The insect olfactory organ emerges from the antennal segment. Interestingly, *lhx5* is also expressed in vertebrate olfactory primordia (Moreno *et al.*, 2003).

Unlike the beetle *lim1/5* ortholog, I did not detect expression in cells of the spider CNS. *P. tepidariorum* *lim1b* [g15693] was a marker for three clusters in the integrated dataset (not shown). One of these was cluster 8, the CNS cluster, where I do not see *lim1a* expression, and the other two clusters (2 and 11) correspond to posterior cell populations. *lim1/5* could be a good candidate for sub- and/or neo-functionalisation, but the role of *lim1b* remains unknown.

I propose *lim1a* has an early role in patterning the cheliceral segment. Future work in *P. tepidariorum* would involve parental and embryonic RNAi against *lim1a*. In *lim1a* knockdown embryos, I expect a loss of the cheliceral segment. Furthermore, I hypothesise *lim1a* eRNAi cell clones fail to abolish *hh1* expression between the splitting pre-cheliceral and cheliceral *hh1* stripes, however, to not induce *hh1* expression outside these *hh1* stripes. Thus, *lim1a* would act as a negative regulator of *hh1* expression to mediate the splitting dynamics for spider head segmentation, similar to *msx1* (Akiyama-Oda and Oda, 2020).

4. Chapter Four Results: Studying the development of the *pars intercerebralis* and *pars lateralis* in the spider brain

4.1 Background: Anterior neuroendocrine centres in insects and centipedes are established in the *six3* expressing neuroectoderm

The insect neuroendocrine system consists of several sets of neurones neurosecretory cells (NSCs) located in the brain and ventral nerve cord. The majority are found in the dorsal-medial protocerebrum, the *pars intercerebralis* (PI) and *pars lateralis* (PL) (Siegmund and Korge, 2001). NSCs of the PI and PL project their axons towards endocrine glands, the corpora cardiaca and corpora allata. In *Drosophila*, the corpora cardiaca, corpora allata, and the prothoracic gland, form a single complex, the ring gland, which surrounds the anterior tip of the aorta. The insect PI-PL/ring gland complex is required for several physiological functions critical in development, energy metabolism, growth, water regulation, and reproduction (Nässel, 2002; Hasebe and Shiga, 2021).

4.1.1 Embryonic development of the *D. melanogaster* PI and PL

In *D. melanogaster* the brain is mostly generated from a bilaterally symmetrical neuroectoderm, which gives rise to ~100 embryonic neuroblasts per brain hemisphere (Urbach and Technau, 2003b). Both ventral and head midline domains, stomatogastric nervous system (SNS) (Hartenstein *et al.*, 1994), and optic lobe (Green *et al.*, 1993), of insects contain neuronal progenitors that do not delaminate as individual cells like the neuroblasts, but that invaginate as an elongated furrow (ventral midline), several separate placodes (head midline, SNS), or individual placodes (optic lobe).

The *D. melanogaster* PI and PL are derived from adjacent bilateral placodes in the *six3*-expressing dorso-medial protocerebrum (de Velasco *et al.*, 2007; Steinmetz *et al.*, 2010) (Figure 22A). The head midline placodes invaginate during late embryogenesis (stage 13, long after brain neuroblast formation has ceased) and directly turn into neural cells (de Velasco *et al.*, 2007). They likely also give rise to type II neuroblasts involved in the development of the central complex (Walsh and Doe, 2017).

The PI placode and its derivatives express the PRD homeodomain protein Dvsx1/Dchx1 and can be followed until the late pupal stage (de Velasco *et al.*, 2007) (Figure 22A). The larval PI includes most if not all, NSCs that project towards the corpora cardiaca/corpora allata through the Nccl (de Velasco *et al.*, 2007; Siegmund and Korge, 2001). The PL placode and NSCs of the PL are marked by the adhesion protein Fasciclin 2 (Fas2) (de Velasco *et al.*, 2007) (Figure 22A). While in *D. melanogaster*, the functional relevance of this co-expression is unknown, in *T. castaneum* after *six3* knockdown, *vsx/chx* and *fas2*

expression corresponding to the presumptive PI and PL primordia were lost (Posnien *et al.*, 2011b). This suggests *six3* functions high in the hierarchy of neuroendocrine development.

The protocerebral expression domain of the terminal gap gene *tailless (tll)* includes both the PI and PL primordia (Figure 22A). Loss of *tll* function removes most of the protocerebrum including the PI, although the effect on the PL primordia was not investigated (de Velasco *et al.*, 2007). This suggests *tll* functions high in the hierarchy of neuroendocrine development.

Loss of Decapentaplegic (Dpp) results in the absence of the dorso-medial head epidermis that in wild-type separates the head midline placodes, resulting in hyperplastic placodes (de Velasco *et al.*, 2007).

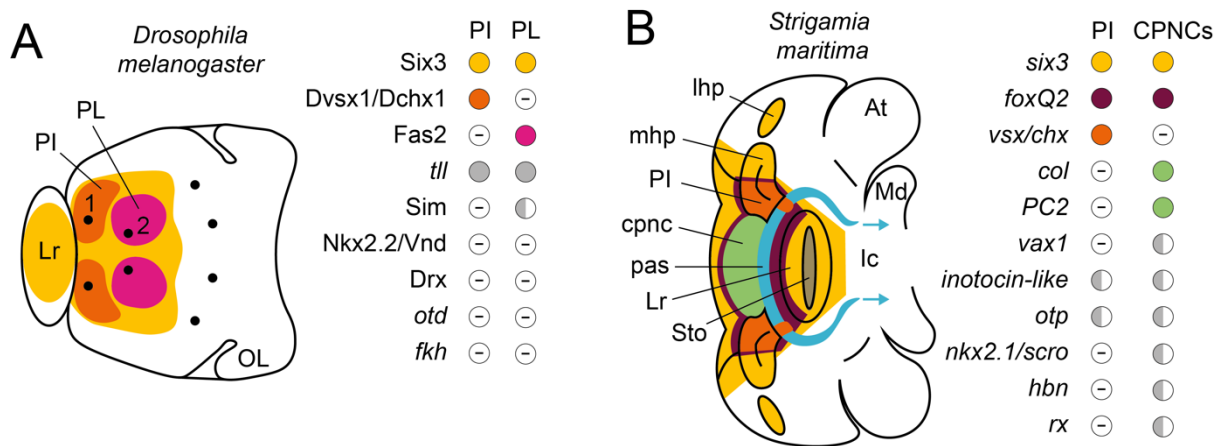


Figure 22: Anterior neurosecretory centres in insects and centipedes are established in the *six3* expressing neuroectoderm.

Heads orientated with the anterior to the left depicting NSC populations and their molecular signature. Protein expression or RNA expression in all cells (circle), a subset of cells (semi-circle), and no cells (-). **(A)** *Drosophila melanogaster*. Crumbs+ invaginating placodal cells (black dots) of the Dvsx1/Dchx1+ *pars intercerebralis* (PI, orange) and Fas2+ *pars lateralis* (PL, pink) in the Six3 (yellow) expressing territory. Schematic adapted from Steinmetz *et al.* (2010) based on protein expression. **(B)** *Strigamia maritima*. Nested domains of *foxQ2* (purple) and *six3* (yellow) expression in the anterior medial region. The *vsx/chx*+ PI primordia colocalises with the medial part of the main head placode (mhp). The *col*+ (and *PC2*+) central pioneer neuronal cells (CPNCs, green) lie between the bilateral parts of the PI. The primary axonal scaffold (pas, blue) extends from the CPNCs in the posterior direction (blue arrows). Schematic adapted from Hunnekuhl and Akam (2014) based on double fluorescent *in situs* hybridisation and anti-acetylated tubulin antibody stains. Other abbreviations: At: antennal segment; Ic: intercalary segment; Lr: labrum; lhp: lateral head placode; Md: mandibular segment; OL: optic lobe; Sto: stomodeum.

4.1.2 Embryonic development of the centipede *Strigamia maritima* PI and a central NSC population that pioneers the CNS

In the centipede *S. maritima*, *foxQ2* expression is entirely nested in the *six3* expressing domain (Hunnekuhl and Akam, 2014) (Figure 22B). The centipede PI is derived within this domain from *vsx/chx*-expressing placodal-like invagination sites (Hunnekuhl and Akam, 2014) (Figure 22B), similar to *D. melanogaster*. It is unknown whether the *S. maritima* has a PL, as there is no expression data for the *fas2* ortholog [SMAR008205] (Chipman *et al.*, 2014).

In addition to the PI, the centipede has central pioneer neuronal cells (CPNCs) characterised by the expression of *collier (col)* and *prohormone convertase 2 (PC2; also PHC2)* (Figure 22B). *col* marks post-mitotic cells committed to a neural fate. *PC2* encodes an enzyme specifically restricted to endocrine and neuroendocrine cells (Lovejoy, 2005; Tessmar-Raible, 2007) thus, the CPNCs form an active early neurosecretory centre (Hunnekuhl and Akam, 2014). The CPNCs form a dense cluster and once internalised project axonal tracts that pioneer the primary axonal scaffold of at least the anterior CNS (Hunnekuhl and Akam, 2014) (Figure 22B). At the stage of specification (first expression of *col* and *PC2*) and internalisation of these cells, along with *six3* and *foxQ2*, *nkx2.1/scarecrow (scro)*, *homeobrain (hbn)*, and *rax* are expressed in the territory from which the cells derive (Hunnekuhl and Akam, 2014). *nkx2.1/scro*, *hbn*, and *rax* are absent from the invagination sites and ectodermal compartments of the centipede PI (Hunnekuhl and Akam, 2014) (Figure 22B). Note that *hbn* is missing from the vertebrate genomes (Mazza *et al.*, 2010). Some CPNCs express the hypothalamic neuron markers *ventral anterior homeobox 1 (vax1)* and *orthopedia (otp)*, as well as the *oxytocin/vasopressin* ortholog *inotocin-like* (formerly *vtn*) (Hunnekuhl and Akam, 2014) (Figure 22B). *otp* and *inotocin-like* are also expressed in some cells of the centipede PI (Figure 22B). *vax* genes are absent from insect genomes (Tessmar-Raible, 2007) and the insect *oxytocin/vasopressin* ortholog *inotocin* has been lost from the *D. melanogaster* genome (Liutkeviciute *et al.*, 2016), precluding further analysis.

The *S. maritima* AMR shares transcriptional (*six3*, *foxQ2*, *nkx2.1/scro*, *rax*) similarities to the trochophore larva (24hpf) apical plate of the marine annelid *Platynereis dumerilii* (Tessmar-Raible *et al.*, 2007; Marlow *et al.*, 2014). From the center of this apical plate, the larval apical organ develops, a sensory structure that includes NSCs. Hunnekuhl and Akam (2014) suggest that the arthropod brain retained ancestrally an anterior medial population of NSCs homologous to those of the apical plate in other invertebrate phyla (Tessmar-Raible, 2004; Tessmar-Raible *et al.*, 2007; Tessmar-Raible, 2007; Marlow *et al.*, 2014), but that this cell population has been lost or greatly reduced in insects. No cell population corresponding to the *S. maritima* CPNCs has been characterised in any other arthropod.

Here I characterised the embryonic development of neuroendocrine cell populations in a chelicerate, the spider *Parasteatoda tepidariorum*. Neurogenesis in *P. tepidariorum* has been morphologically described by Mittman and Wolf (2012), beginning at late stage 9/early stage 10 with the appearance of point-like depressions of INPs, in the ventral neuroectoderm and pre-cheliceral lobes. In spiders, these INPs consist of 5 to 9 cells and directly differentiate into neurones (Stollewerk *et al.*, 2001). The pre-cheliceral region undergoes major morphogenetic movements, beginning at stage 10.2, which form the main brain compartments found in the adult spider (Doeffinger *et al.*, 2010). Recently, in *P. tepidariorum* I identified a cell population possibly corresponding to the insect PI (Leite *et al.*, 2024). In another entelegyne spider *C. salei*, Prospero antibody staining found many postmitotic cells committed to neural fate were located in the medial pre-cheliceral lobes that give rise to the central protocerebrum (Doeffinger *et al.*, 2010). However, it is not clear whether the spider's medial neural precursors are similar to the centipede's CPNCs (Hunnekuhl and Akam, 2014). Comparing the expression profile of the discussed genes in the spider to mandibulate arthropods may help elucidate the evolutionary relationships of embryonic cell populations across the arthropods and the structures they give rise to.

4.1.3 Chapter four aims

This chapter aimed to characterise the embryonic development of neuroendocrine cell populations in a chelicerate, the spider *P. tepidariorum*, by analysing the expression of known marker genes from insects and centipedes.

Objectives:

- Examine the spatial and relative expression of genes involved in the embryonic development of the insect PI using colourimetric ISH and dFISH
- Examine the spatial and relative expression of genes involved in the embryonic development of the insect PL using colourimetric ISH and dFISH
- Examine the spatial and relative expression of genes involved in the embryonic development of the centipede CPNCs using colourimetric ISH and dFISH

As in *D. melanogaster*, a *vax* ortholog (Aase-Remedios *et al.*, 2023) and the inotocin signalling system (Liutkeviciute *et al.*, 2016) has been lost in *P. tepidariorum*, precluding further analysis.

4.2 Results: Development of the spider *pars intercerebralis* from bilateral invaginating head placodes

P. tepidariorum has one *vsx/chx* ortholog [aug3.g27186], the expression of which was previously characterised at embryonic stages 8 and 9 (Leite *et al.*, 2024). To determine whether spiders have a PI, I looked at the later expression of *vsx/chx* and the relative expression of *vsx/chx* and the spider anterior median region (AMR) marker *six3.2* from the start of neurogenesis at stage 10.

I found that from stage 10.1, *vsx/chx* was no longer expressed along the ventral midline (Figure 23A). The C-shaped expression marked the outgrowing labrum (Figure 23A-C; Supplementary Figure 42). *vsx/chx* expression was nearly entirely nested in the *six3.2* expressing AMR (Figure 23D-F), similar to *T. castaneum* (Posnien *et al.*, 2011b), which suggests *six3.2* may act directly upstream of *vsx/chx*. At stage 10.1, *vsx/chx* and *six3.2* are co-expressed in and surrounding INPs (Figure 23D). At stage 10.2, *vsx/chx* expression also colocalised with the medial part of the anterior furrow (Figure 23E), similar to *vsx/chx* expression in the main head placode of *S. maritima* (Hunnekuhl and Akam, 2014). The combination of these genes at the corresponding position in insects and the centipede, suggests that spiders also have a PI.

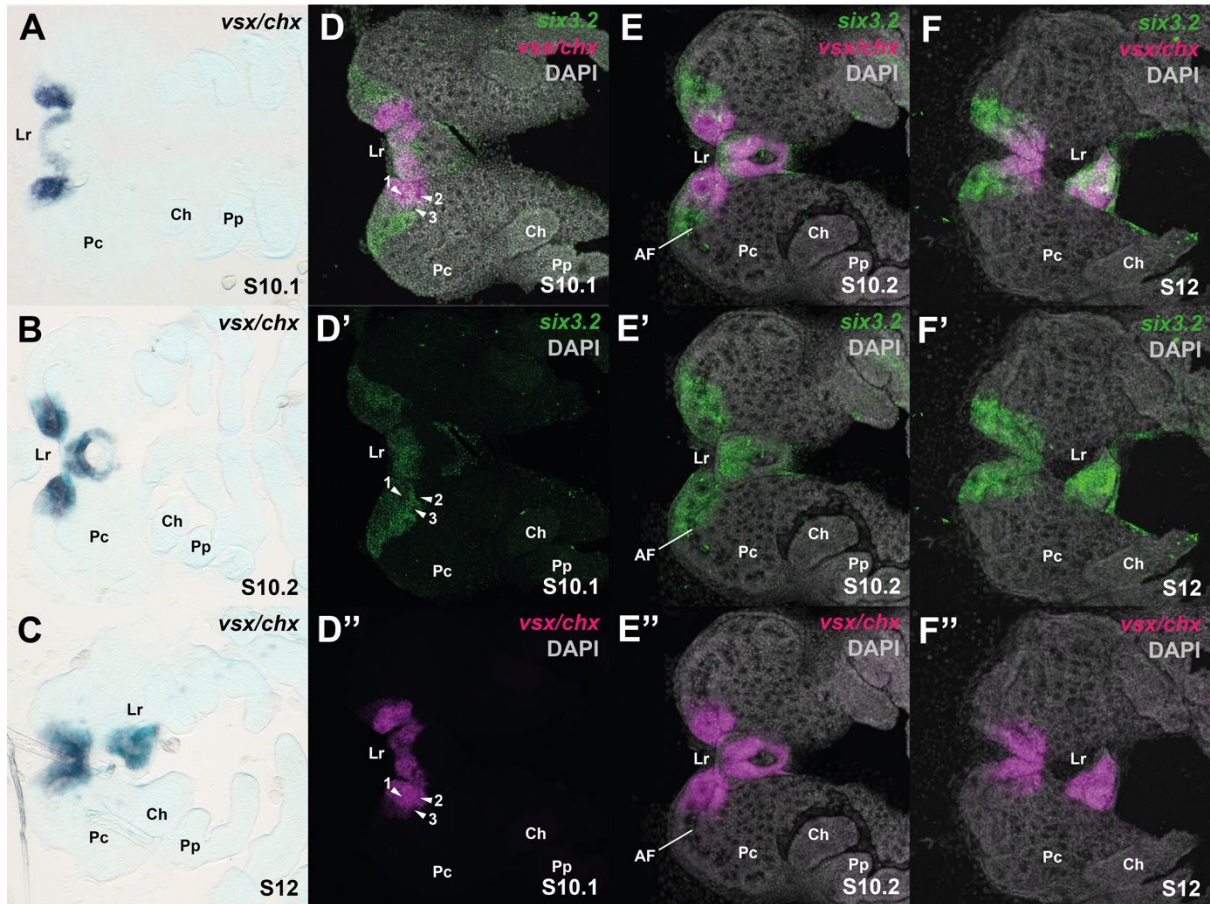


Figure 23: Expression of *vsx/chx* and *six3.2* in the developing head of *P. tepidariorum* embryos.

Embryonic heads were flat mounted with the anterior to the left. **(A-C)** Embryos stained for *vsx/chx* transcripts at stages 10.1 **(A)**, 10.2 **(B)**, and 12 **(C)**. **(D-F'')** Embryos stained for *six3.2* (green) and *vsx/chx* (magenta) transcripts with or without nuclei/DNA (DAPI: grey) at stages 10.1 **(D)**, 10.2 **(E)**, and 12 **(F)**. **(D-D'')** Three INPs in the AMR express *six3.2* and *vsx/chx* (1), *vsx/chx* (2), and *six3* (3). **(E-E'')** At stage 10.2, *vsx* expression colocalised with the medial part of the anterior furrow (AF). Other abbreviations: Ch: chelicerae; Lr: labrum; LF: lateral furrow; Pp: pedipalps; Pc: pre-chelicerai.

4.3 *foxQ2* and *vsx/chx* are expressed non-overlapping domains in the developing brain and labrum

P. tepidariorum has one *foxQ2* ortholog [g224], which is first expressed in a stripe at the anterior rim of the germ band at stage 7 (Schacht *et al.*, 2020). Therefore, to further characterise the molecular signature of the spider PI, I looked at the relative expression of *foxQ2* and *vsx/chx* from the onset of *vsx/chx* expression at stage 8 (Leite *et al.*, 2024) and onwards.

I found that *foxQ2* and *vsx/chx* were expressed in abutting mutually exclusive domains in the developing brain and labrum (Figure 24). Thus, *foxQ2* was absent from the invagination sites and expression domains of the spider PI, in contrast to *S. maritima* (Hunnekuhl and Akam, 2014) and *T. castaneum* (Kitzmann *et al.*, 2017).

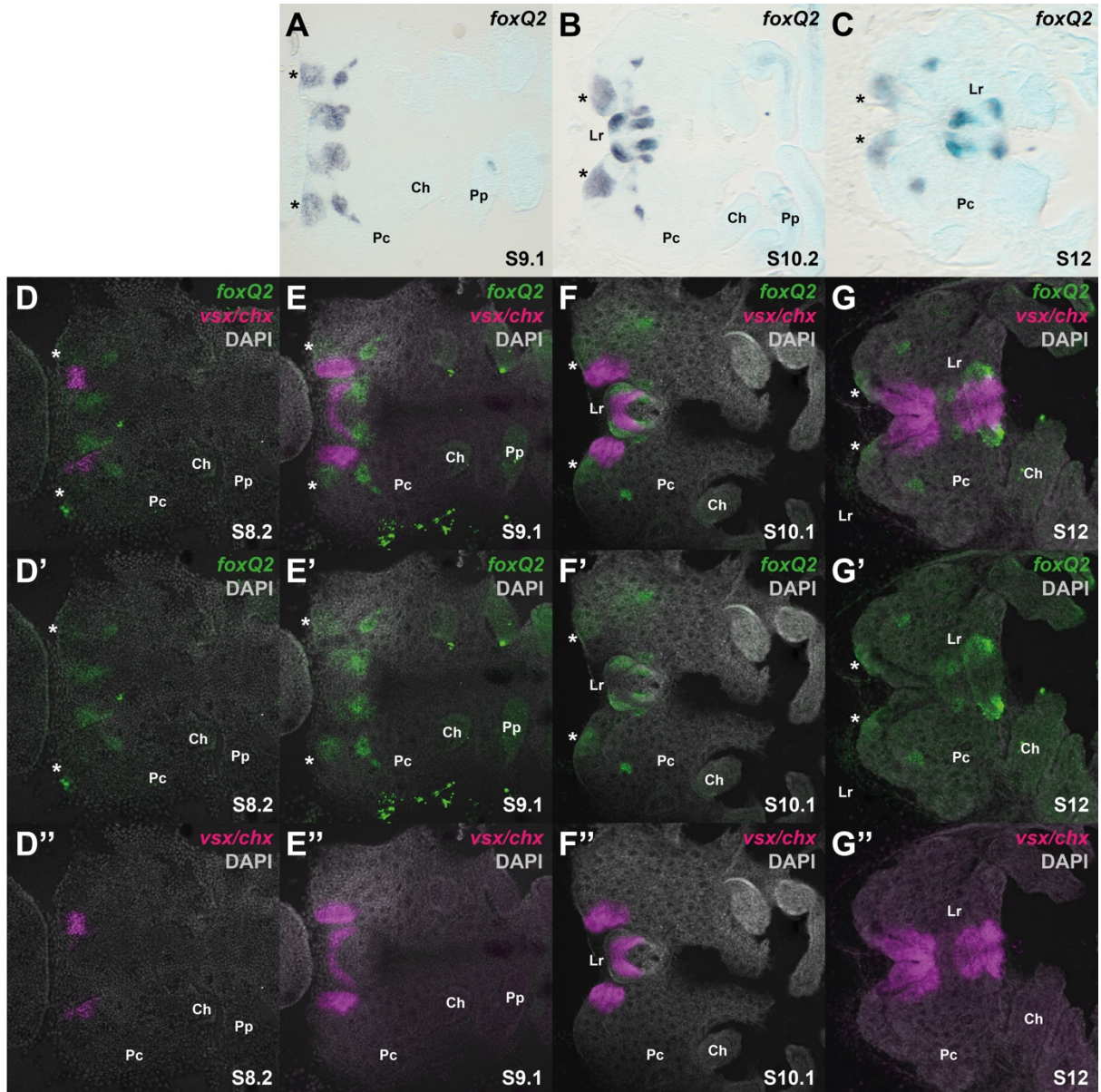


Figure 24: Expression of *vsx/chx* and *foxQ2* in the developing head of *P. tepidariorum* embryos. Embryos were flat mounted with the anterior to the left. Corresponding expression domains in all panels are indicated by an asterisk. **(A-C)** Embryos stained for *foxQ2* transcripts at stages 9.1 **(A)**, 10.1 **(B)**, and 12 **(C)**. **(D-G'')** Embryos stained for *foxQ2* (green) and *vsx/chx* (magenta) transcripts and nuclei/DNA (DAPI: grey) at stages 8.2 **(D)**, 9.1 **(E)**, 10.1 **(F)**, and 12 **(G)**. Other abbreviations: Ch: chelicerae; Lr: labrum; Pp: pedipalps; Pc: pre-chelicerai.

4.4 *fas2A* expression in an INP

P. tepidariorum has two *fas2* orthologs, *fas2A* [g31369] and *fas2B* [g18551], and their expression was previously characterised from embryonic stages 8 to 12 (Schomburg *et al.*, 2020). To determine whether spiders have a PL, I re-examined the expression of *P. tepidariorum* *fas2* orthologs. In addition to the relative expression of *fas2A* and the spider AMR marker *six3.2* and the spider PI marker *vsx/chx*.

I found *fas2A* was expressed from stage 9.2 in the neurogenic ectoderm of the pre-cheliceral lobes and segmentally repeated domains along the body (Figure 25A), earlier than previously reported (Schomburg *et al.*, 2020). From the onset of *fas2A* expression, the most-anterior medial spot of *fas2A* expression overlaps with *six3.2* (Figure 25E), similar to *T. castaneum* (Posnien *et al.*, 2011b). At stage 10.1, this spot of *fas2A* expression was observed in an INP (Figure 25F). The expression domains of *vsx/chx* and *fas2A* were mutually exclusive throughout development (Figure 25G&H), as in *D. melanogaster* (de Velasco *et al.*, 2007) and *T. castaneum* (Posnien *et al.*, 2011b). The combination of these genes at the corresponding position in insects, suggests that spiders also have a PL.

At stage 10.2, *fas2A* was expressed in three spots in the dorsal part of the stomodeum (Figure 25C). In spiders, INPs in a crescent-shaped array in the dorsal part of the stomodeum were suggested to give rise to the SNS (Doeffinger *et al.*, 2010) similar to *D. melanogaster*, in which the SNS primordium forms from three invaginating placodes in the roof of the foregut (Hartenstein *et al.*, 1994). Thus, *fas2A* likely contributes to the formation of the SNS in the spider. Furthermore, at stage 10.2 there was an additional *fas2A* domain at the rim of the head lobes next to the anterior furrow (Figure 25C). This *fas2A* expression domain remained at the leading edge of the non-neurogenic ectoderm and likely corresponded to the PEP (Figure 25D). I confirmed *fas2B* had uniform and ubiquitous expression throughout the embryo (not shown) (Schomburg *et al.*, 2020).

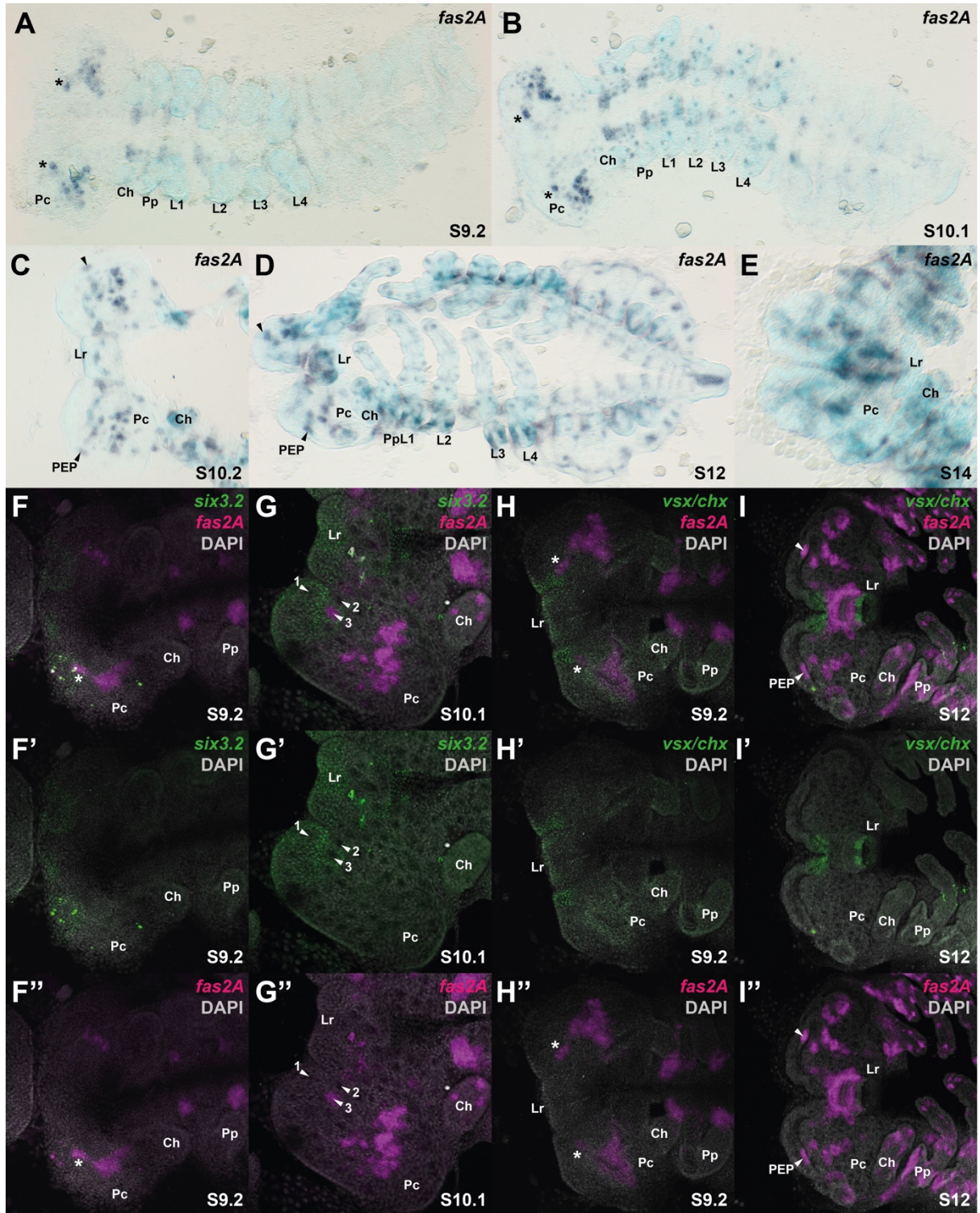


Figure 25: Expression of *fas2A*, *fas2A* and *six3.2*, and *fas2A* and *vsx/chx* in *P. tepidariorum* embryos.

Embryos were flat mounted with the anterior to the left. Corresponding expression domains in all panels are indicated by an asterisk. (A-E) Embryos stained for *fas2A* transcripts at stages 9.2 (A), 10.1 (B), 10.2 (C), 12 (D), and 14 (E). (F-G'') Embryos stained for *six3.2* (green) and *fas2A* (magenta) transcripts and nuclei/DNA (DAPI: grey) at stages 9.2 (F) and 10.1 (G). (G) Three point-like depressions in the AMR that express *six3.2* and *vsx/chx* (1, see Figure 23D), *vsx/chx* (2, Figure 23D), and *six3.2* and *fas2A*

(3). **(H-I'')** Embryos stained for *vsx/chx* (green) and *fas2A* (magenta) transcripts and nuclei/DNA (DAPI: grey) at stages 9.2 **(H)** and 12 **(I)**. Other abbreviations: Ch: chelicerae; Lr: labrum; Pp: pedipalps; Pc: pre-chelicerai; PEP: principal eye primordia; L1-4: walking legs 1-4.

4.5 *tll* marks the developing protocerebrum and encompasses the spider PI and PL

P. tepidariorum has a single *tll* ortholog [g18090] (Leite *et al.*, 2018), and although expression of this gene was previously characterised from embryonic stages 8 to 12, the pattern was difficult to resolve (Schomburg *et al.*, 2020). To further characterise the molecular signature of the *P. tepidariorum* PI and PL, I re-examined the expression of *P. tepidariorum tll* and the relative expression of *tll* and the spider head segmentation gene *hedgehog* (*hh*).

I found that *tll* was expressed from stage 6, earlier than previously reported (Schomburg *et al.*, 2020), in a stripe at the anterior rim of the germband (Figure 16A&B). This expression was anterior to the splitting pre-chelicerai/chelicerai and pedipalp *hh1* stripe and thus, marks the pre-chelicerai region (Figure 16F). Similarly, *tll* expression in the *D. melanogaster* embryo marks the future anterior (Pignoni *et al.*, 1990). Unlike in *D. melanogaster* (Pignoni *et al.*, 1990) and *T. castaneum* (Schroder *et al.*, 2000) however, *tll* was not expressed in the future/early posterior of the spider germ band (Figure 16A). Subsequently, *P. tepidariorum tll* expression marks the neurogenic regions of the pre-chelicerai lobes laterally to the labrum and stomodeum (Figure 16D,E&G), in a similar pattern to *T. castaneum tll*. This suggests *tll* is required for protocerebrum development and encompasses the spider PI and PL, as in *D. melanogaster* (de Velasco *et al.*, 2007).

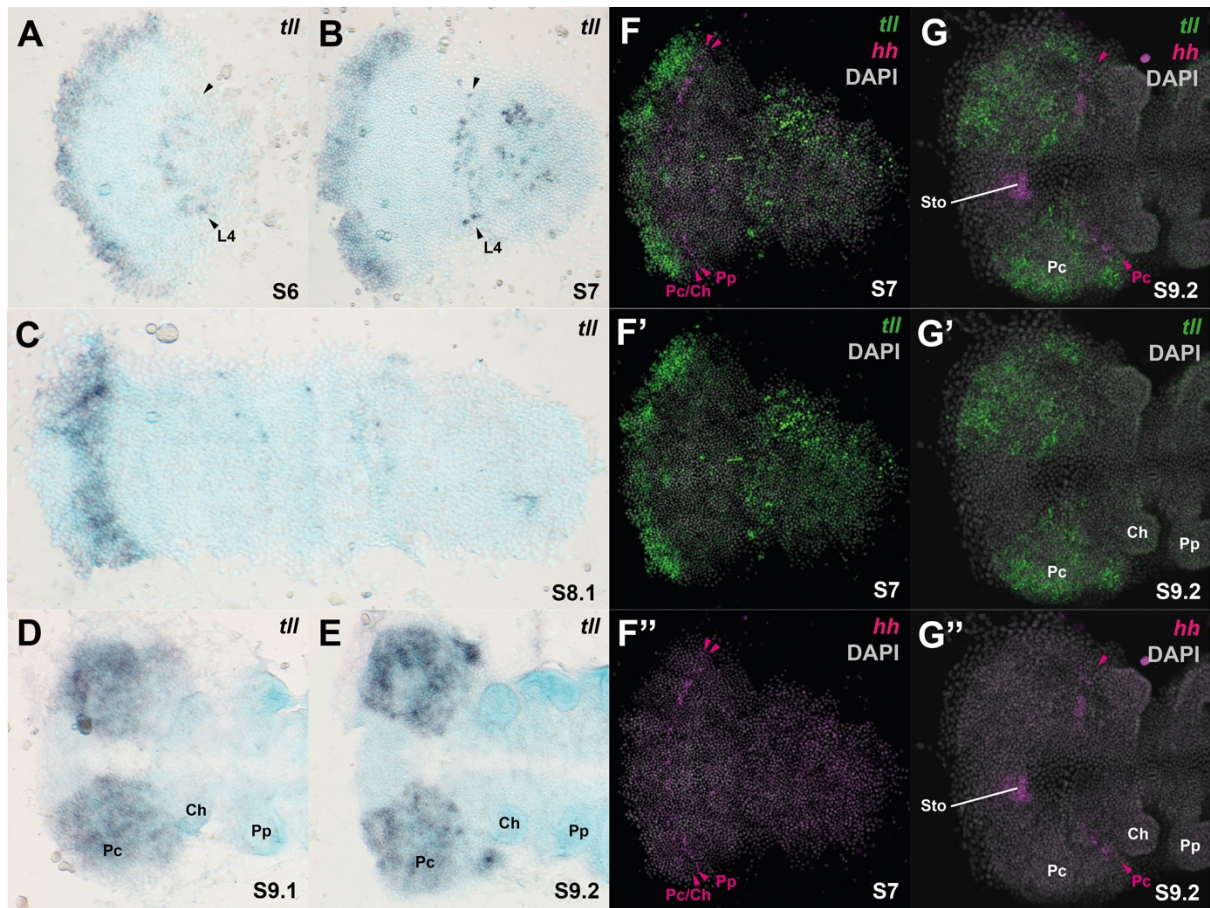


Figure 26: Expression of *tll* and *hh1* in *P. tepidariorum* embryos.

Embryos were flat mounted with the anterior to the left. **(A-E)** Embryos stained for *tll* transcripts at stages 6 **(A)**, 7 **(B)**, 8.1 **(C)**, 9.1 **(D)**, and 9.2 **(E)**. **(F-G'')** Embryos stained for *tll* (green) and *hh* (magenta) transcripts and nuclei/DNA (DAPI: grey) at stages 7 **(F)** and 9.2 **(G)**. **(A)** *tll* expression was first seen at stage 6, in a stripe spanning the anterior rim of the germ band in a broad stripe and in a transient mesodermal stripe (black arrowhead) potentially marking the fourth walking leg segment (L4). **(B)** *tll* expression at the anterior rim expands. **(C)** *tll* expression begins to clear from the AMR and the mesodermal expression disappears. **(D)** *tll* expression splits into two domains and marks the lateral neurogenic regions of the pre-chelicer segment. **(F)** At stage 7, *tll* was expressed anterior to the splitting *hh1* stripes (magenta arrowhead). **(G)** At stage 9.2, *tll* was expressed directly adjacent to the pre-chelicer *hh1* stripe and was not co-expressed with stomodeal *hh1* expression. Other abbreviations: Ch: chelicerae; Pp: pedipalps; Pc: pre-chelicer; Sto: stomodeum.

4.6 Spiders have a *col*-expressing anterior neuronal cell population

P. tepidariorum has two *col* orthologs, which are likely ohnologs retained from the ancestral WGD (Schwager *et al.*, 2017), and their expression was characterised previously at stage 12. *col1* [g10955] was reported to be expressed in a small number of cells at the anterior rim of the head lobes at stage 12 (Schaeper *et al.*, 2010). While *col2* [g31501] was described as expressed at stage 12 in bilateral spots in the L4, O1, and O2 segments (Schomburg *et al.*,

2020). To determine whether spiders have a similar cell population to the *S. maritima* CPNCs, I first needed to determine whether spiders have a neurogenic *six3* and *foxQ2* expressing domain thus, I first looked at the relative expression of *foxQ2* and the spider AMR marker *six3.2*. Following this I re-examined the embryonic expression of the *P. tepidariorum* *col* ohnologs.

I found that only the anterior-most domain of *foxQ2* expression was nested in the *six3.2* domain (Figure 27A-D). At stage 10.1, several point-like depressions of internalised neural precursor cells were observed where their expression overlaps (Figure 27C), indicating the anterior-most domain of *foxQ2* is in the neurogenic ectoderm, which was previously misinterpreted as non-neurogenic by Schacht et al. (2020).

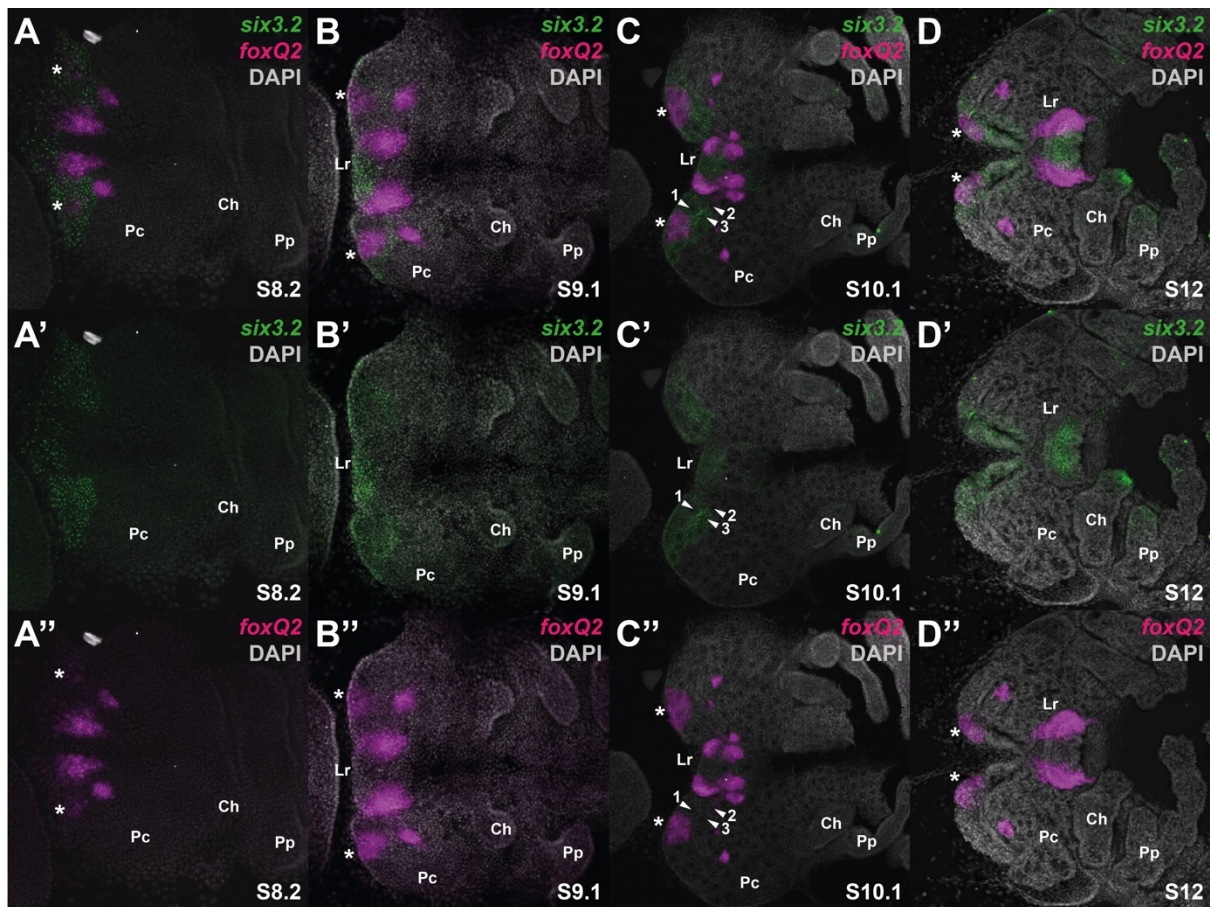


Figure 27: Expression of *six3.2* and *foxQ2* in the developing head of *P. tepidariorum* embryos.

Embryos were flat mounted with the anterior to the left. Corresponding expression domains in all panels are indicated by an asterisk. **(A-D'')** Embryos stained for *six3.2* (green) and *foxQ2* (magenta) transcripts and nuclei/DNA (DAPI: grey) at stages 8.2 (**A**), 9.2 (**B**), 10.1 (**C**), and 12 (**D**). **(C)** Three point-like depressions in the AMR that express *six3.2* and *vsx/chx* (1, see Figure 23D), *vsx/chx* (2, see Figure 23D), and *six3.2* and *fas2A* (3, see Figure 25F) are all free of *foxQ2* expression. Other abbreviations: Ch: chelicerae; Lr: labrum; Pp: pedipalps; Pc: pre-chelicerai.

col1 was not expressed in the pre-cheliceral lobes until stage 10.2 (Figure 28A&B). At stage 11, *col1* was expressed in the medial pre-cheliceral lobes (Figure 28C&F). At stage 12, *col1* is expressed in four bilateral domains at the anterior rim of the pre-cheliceral lobes (Figure 28D&G). The two medial *col1* expression domains at the anterior rim of the pre-cheliceral lobes overlap with *six3.2* (Figure 28H), and likely the anterior-most expression domain of *foxQ2* (Figure 27D).

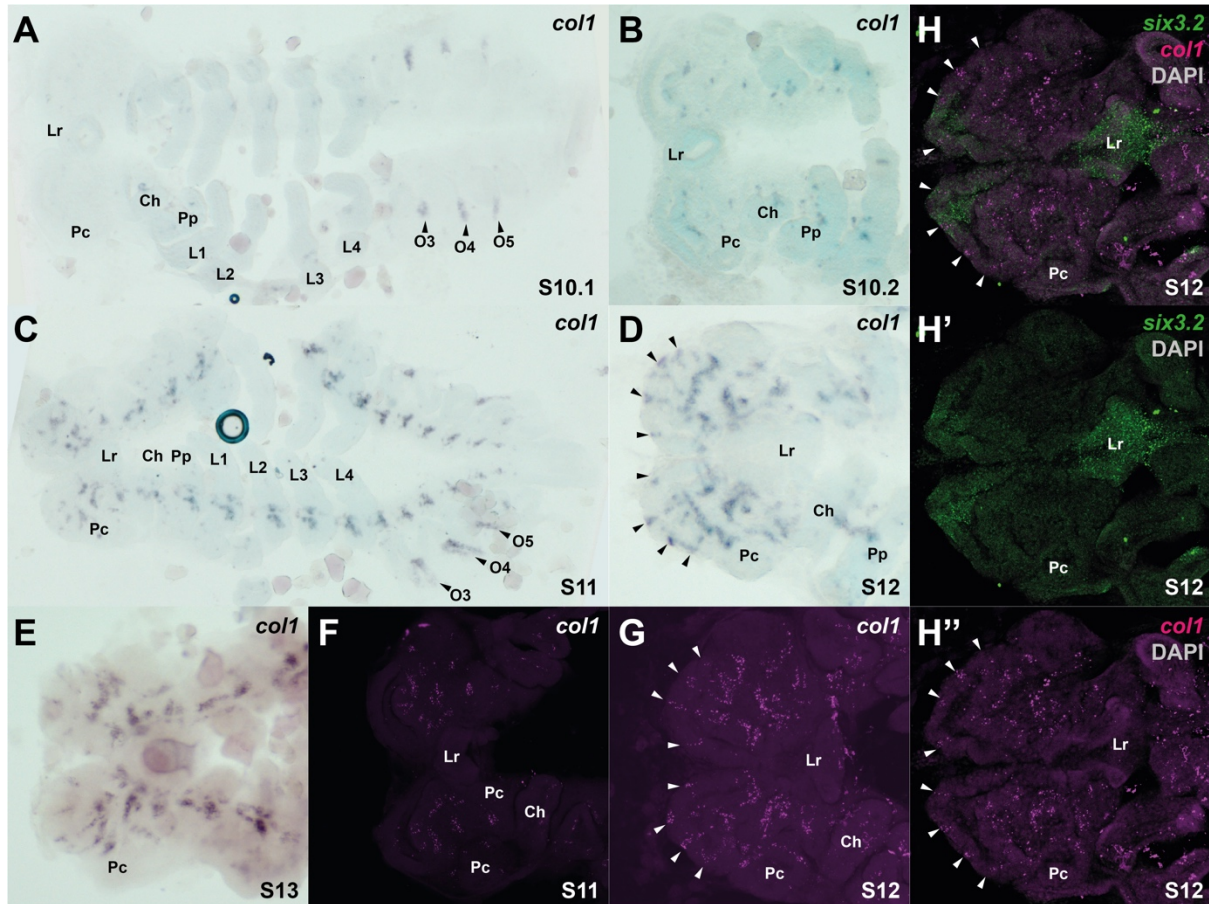


Figure 28: Expression of *col1* in the developing head of *P. tepidariorum* embryos.

Embryos were flat mounted with the anterior to the left. Stages prior to the onset of expression were omitted. Corresponding expression domains in all panels are indicated by arrowheads. **(A-G)** Embryos stained for *col1* transcripts at stages 10.1 **(A)**, 10.2 **(B)**, 11 **(C&F)**, 12 **(D&G)**, and 13 **(E)**. **(A)** *col1* expression was first seen at stage 10.1 in the anterior boundary of the third, fourth, and fifth opisthosomal segments (O3, O4 and O5; black arrowheads). **(H-H'')** Embryos stained for *six3.2* (green) and *col1* (magenta) transcripts and nuclei/DNA (DAPI: grey) at stages 12. Other abbreviations: Ch: chelicerae; Lr: labrum; Pp: pedipalps; Pc: pre-cheliceral.

4.6.1 *P. tepidariorum* PC2 has no detectable embryonic expression

P. tepidariorum has a single PC2 ortholog [g24083] (Fritzsche and Hunnekuhl, 2021), but its expression has not been characterised in this spider or any other chelicerate. ISH with a PC2 probe produced no distinguishable signal from stages 8 to 12.

4.7 *nkx2.1/scro* is expressed in some cells of the spider PI placode

P. tepidariorum has one *nkx2.1/scro* ortholog [g16180] (Leite *et al.*, 2018), but its expression has not been characterised in this spider or any other chelicerate.

I found *nkx2.1/scro* was exclusively expressed in the pre-chelicerar region, in the stomodeum primordia and developing brain (Figure 29). At stage 10.2, *nkx2.1/scro* was expressed in distinct elevations in the developing brain: the medial and lateral subdivisions abutting the anterior and lateral furrow, respectively (Figure 29E). From stage 10.2, *nkx2.1/scro* overlaps with the posterior part of *vsx/chx* expression corresponding to the spider PI placode (Figure 23D-F and Figure 29E-F). At stages 10.2 and 12, *nkx2.1/scro* expression partly overlaps with *col1* expression in the medial pre-chelicerar lobes (Figure 28C&F and Figure 29E&F).

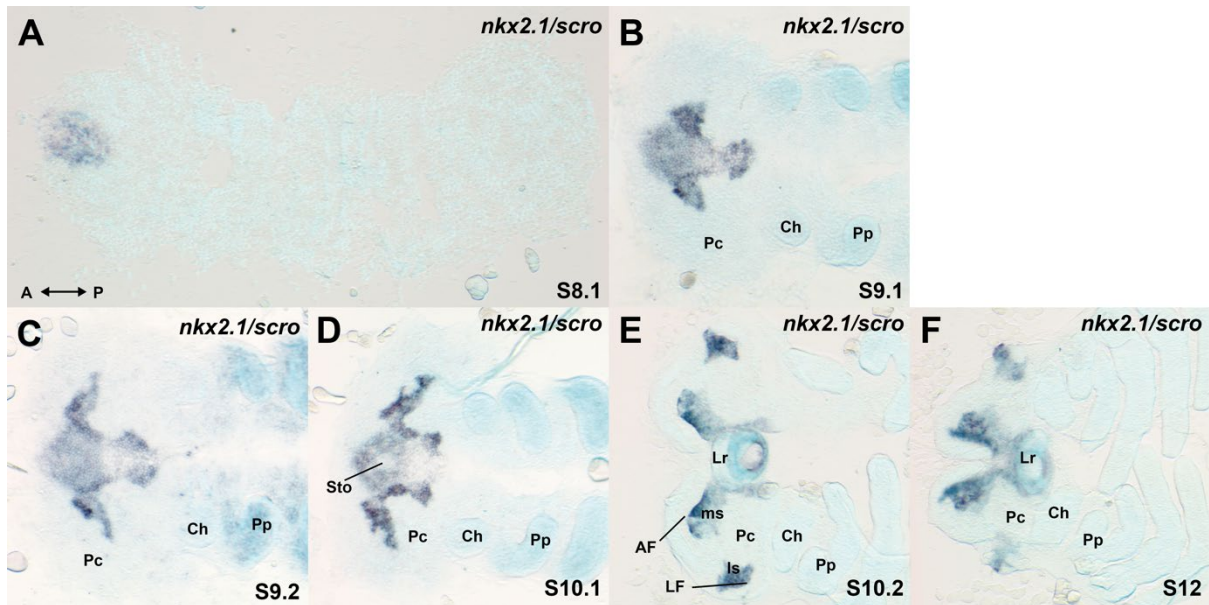


Figure 29: Expression of *nkx2.1/scro* in the developing head of *P. tepidariorum* embryos.

Embryos were flat mounted with the anterior to the left. Stages prior to the onset of expression were omitted. **(A-F)** Embryos stained for *nkx2.1/scro* transcripts at stages 8.1 **(A)**, 9.1 **(B)**, 9.2 **(C)**, 10.1 **(D)**, 10.2 **(E)**, and 12 **(F)**. **(A)** *nkx2.1/scro* expression was first seen at stage 8.1, in a circular domain at the anterior rim of the germ band. **(B-D)** From stage 9.1, *nkx2.1/scro* expression retracts from the anterior rim and extends from the circular domain into the neurogenic ectoderm. **(D)** At stage 10.1, *nkx2.1/scro* expression clears at the centre of the circular domain, resulting in expression surrounding the stomodeum (Sto) primordia. **(E)** At stage 10.2, *nkx2.1/scro* was expressed in the medial subdivisions (ms) abutting the anterior furrow (AF) and an additional expression domain arises in the lateral subdivisions (ls) abutting the lateral furrow (LF). **(F)** At stage 12, the AF and LF were closed by the expansion of the respective subdivisions and the two halves of the brain fuse at the midline. Other abbreviations: Ch: chelicerae; Lr: labrum; Pp: pedipalps; Pc: pre-chelicerar.

To elucidate the role of *nkx2.1/scro* in stomodeum development, I then looked at the relative expression of *nkx2.1/scro* with spider stomodeal markers *FoxA1/forkhead1* (*FoxA1/fkh1*) and *hh1* (Feitosa *et al.*, 2017) (Figure 30). From the onset of *nkx2.1/scro* expression at the anterior rim of the germband, I see co-expression *FoxA1/fkh1* (Figure 30A-B). This suggests part of the early anterior *nkx2.1/scro* domain is endodermal. From stage 8, *hh1* is a specific marker for the developing stomodeum (Feitosa *et al.*, 2017). *nkx2.1/scro* expression appears to be underneath stomodeal *hh1* expression before *nkx2.1/scro* expression clears at the centre of the domain (Figure 30C-D). *nkx2.1/scro* is expressed in the crescent-shaped array of INPs in the dorsal part of the stomodeum, which gives rise to the SNS in the spider (Doeffinger *et al.*, 2010).

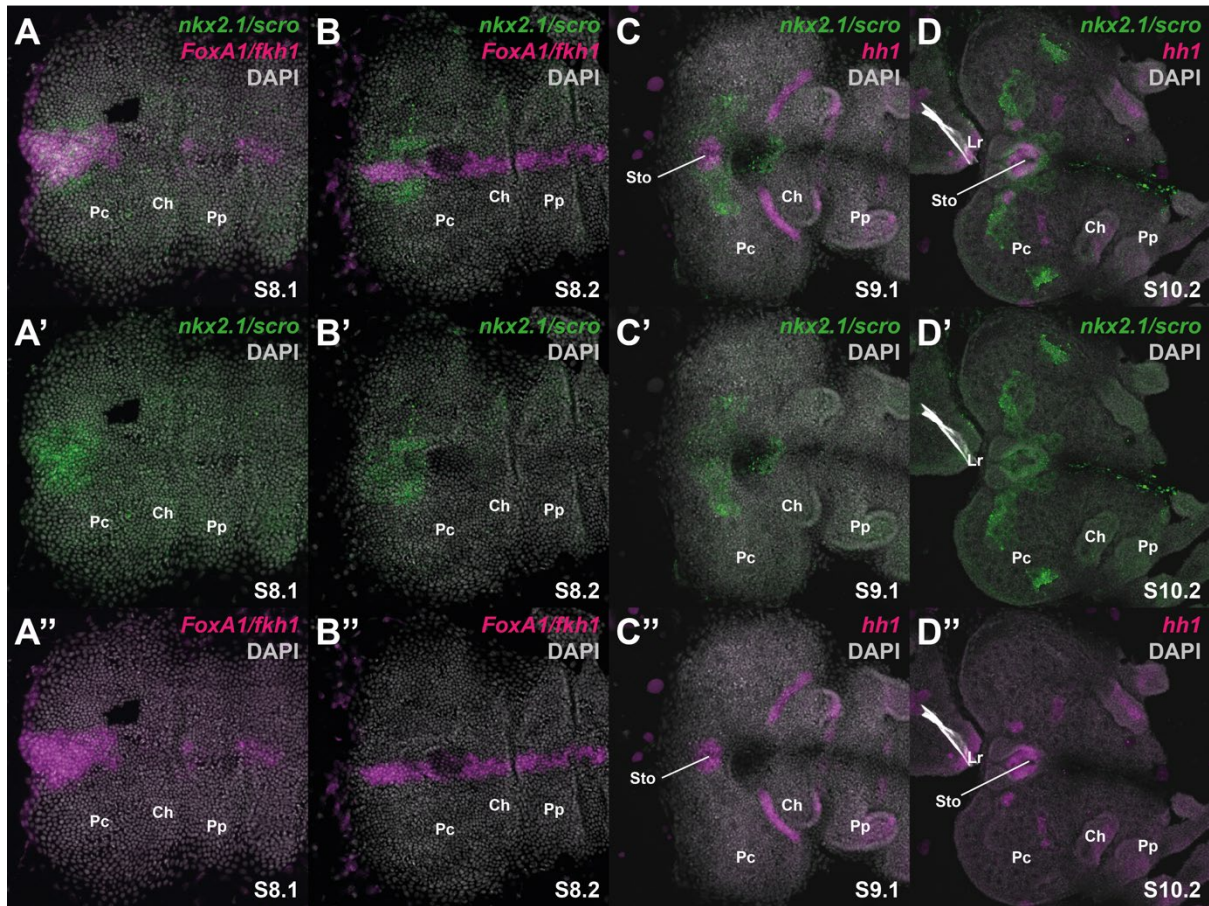


Figure 30: Expression of *nkx2.1/scro* and spider stomodeal markers in the developing head of *P. tepidariorum* embryos.

Embryos were flat mounted with the anterior to the left. (A-D'') Embryos stained for *nkx2.1/scro* (green) and *FoxA1/fkh1* (magenta) transcripts and DNA (grey) at stages 8.1 (A) and 8.2 (B). (C-D'') Embryos stained for *nkx2.1/scro* (green) and *hh1* (magenta) transcripts and DNA (grey) at stages 9.1 (C) and 10.2 (D). Other abbreviations: Ch: chelicerae; Lr: labrum; Pp: pedipalps; Pc: pre-chelicerar; Sto: Stomodeum.

4.8 *hbn* and *raxA* are absent from the spider PI/PL

P. tepidariorum has one *hbn* ortholog [g8762], one *otp* ortholog [g8767], and two *rax* orthologs, *raxA* [g8760] and *raxB* [g23262] (Leite *et al.*, 2018), but their expression has not been characterised in this spider or any other chelicerate.

I found *hbn* and *raxA* were exclusively expressed in the pre-chelicerel region and in a similar pattern in the developing brain (Figure 31). At stage 10.1 expression of *hbn* (Figure 31D), and *raxA* (Figure 31J) are mutually exclusive of *vsx/chx* (Figure 23D) and *fas2A* (Figure 25G) expression domains corresponding to the spider PI and PL primordia, respectively. Furthermore, similar to *D. melanogaster* (de Velasco *et al.*, 2007) and *Tribolium* (Posnien *et al.*, 2011b), I found *P. tepidariorum* *raxA* (and *hbn*) is expressed adjacent to the *six3.2* domain. In the fly, Rax at least partly overlaps with the two head midline placodes posterior to the PL (de Velasco *et al.*, 2007) (Figure 22A). Similar to *D. melanogaster*, I found *P. tepidariorum* *raxA* (and *hbn*) colocalises with the lateral part of the anterior furrow, the neural primordium of the arcuate body (Doeffinger *et al.*, 2010). *hbn* (Figure 31F) and *raxA* (Figure 31L) likely overlap with the two lateral *col1* expression domains at the anterior rim of the pre-chelicerel lobes at stage 12 (Figure 28D&G).

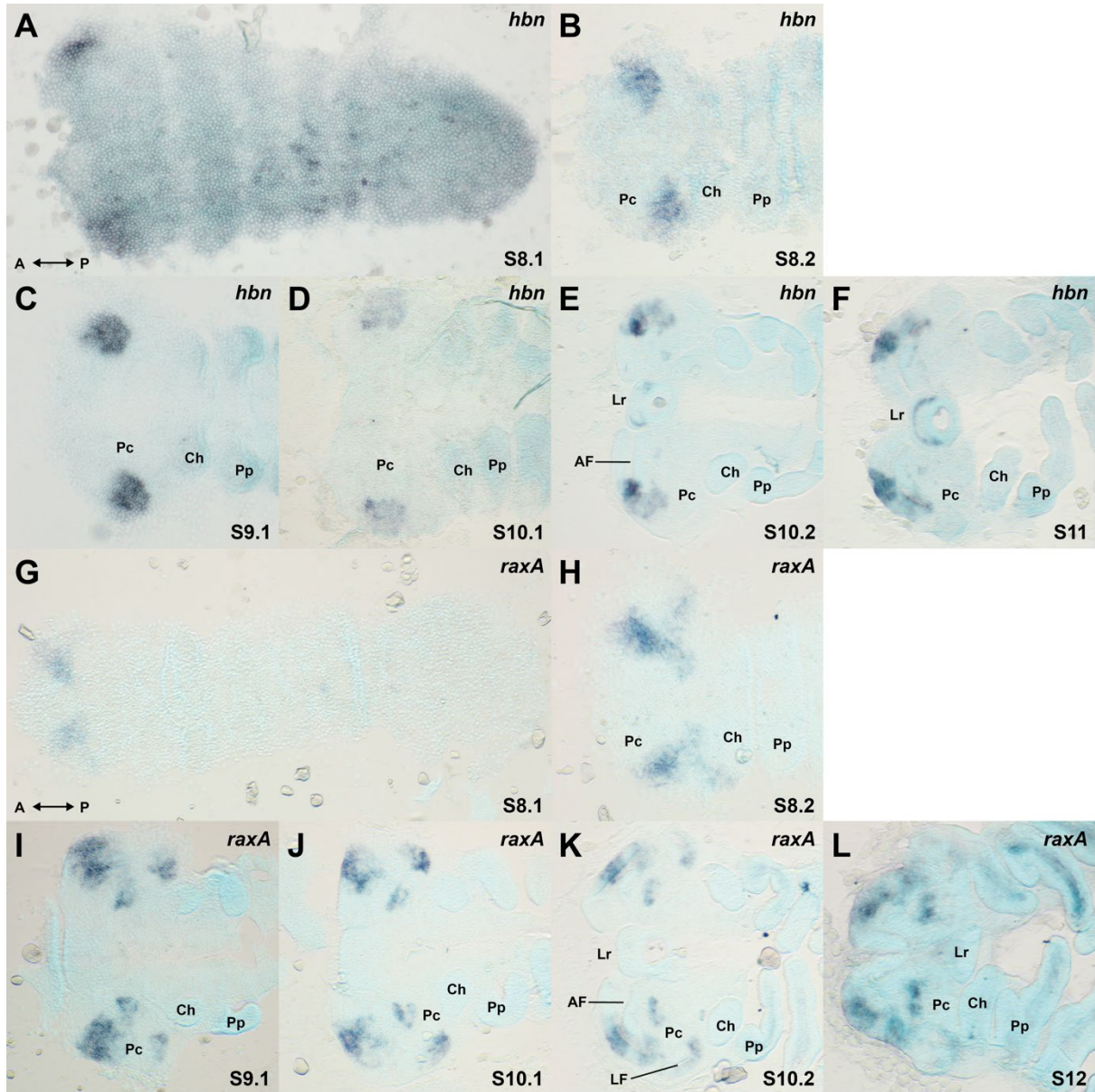


Figure 31: Expression of *hbn* and *raxA* in the developing head of *P. tepidariorum* embryos.

Embryos were flat mounted with the anterior to the left. Stages prior to the onset of expression were omitted. **(A-F)** Embryos stained for *hbn* transcripts at stages 8.1 **(A)**, 8.2 **(B)**, 9.1 **(C)**, 10.1 **(D)**, 10.2 **(E)**, and 11 **(F)**. **(A and B)** *hbn* was first expressed at stage 8.1, in two bilateral patches in the anterior lateral neurogenic ectoderm of the pre-chelicer lobes. **(C and D)** From stage 10.2, *hbn* was expressed in and surrounding the lateral part of the anterior furrow (AF) and the labrum (Lr). **(G-L)** Embryos stained for *raxA* transcripts at stages 8.1 **(G)**, 8.2 **(H)**, 9.1 **(I)**, 10.1 **(J)**, 10.2 **(K)**, and 12 **(L)**. **(G)** *raxA* was first expressed at stage 8.1, in two bilateral patches in the anterior lateral neurogenic ectoderm of the pre-chelicer lobes. **(H-L)** *raxA* was subdivided into three domains, in and surrounding the lateral part of the AF, the medial pre-chelicer lobes, and lateral of the lateral furrow (LF). Other abbreviations: Ch: chelicerae; Lr: labrum; LF: lateral furrow; Pp: pedipalps; Pc: pre-chelicer.

ISH with a *raxB* and *otp* probe produced no distinguishable signal from stages 8 to 12.

By analysing the expression patterns of genes including *tll*, *six3.2*, *vsx/chx*, and *fas2A* I have identified that spiders have PI and PL, consistent with the *D. melanogaster* definition (de Velasco *et al.*, 2007) (Figure 32). From previously published expression data on *P. tepidariorum* of *otd* (Pechmann *et al.*, 2009), *FoxA/fkh* (Feitosa *et al.*, 2017; Janssen *et al.*, 2022), and *nkx2.2/vnd* (Leite *et al.*, 2018) paralogs are also absent from the invagination sites and expression domains of the presumptive spider PI and PL (Figure 32), consistent with what has been observed in *D. melanogaster* (de Velasco *et al.*, 2007).



In the grasshopper *Schistocerca gregaria*, at least four type II neuroblasts contributing to the WXYZ tracts of the central complex are located in the developing PI (Boyan and Reichert, 2011; Boyan and Williams, 1997). I found INPs in the spider PI, which may contribute to the development of the arcuate body (Figure 32). In insects, cells of the PI merge to form the most medial neurosecretory structure (de Velasco *et al.*, 2007), whereas in *S. maritima*, the bilateral parts of the PI remain separated by the *col*-expressing population (Hunnekuhl and Akam, 2014). At later developmental stages in the spider, cells of the also PI merge. Future work in *P. tepidariorum* would involve confirming that NSCs of the classically defined PI fall within the

vsx/chx domain by labelling with neuropeptide markers. For example, FMRFamides, which are found in NSCs of the *D. melanogaster* larval PI and vertebrate hypothalamus.

The identification of PI and PL in the spider, a representative of chelicerates, suggests that these neuroendocrine centres date back to at least the arthropod common ancestor. Expression data in the onychophoran *Euperipatoides kanangrensis* suggests the PI may have a panarthropod origin (Janssen, 2017). It would be interesting to see where *fas2* expression localises during head development in the *S. maritima* and *E. kanangrensis* to determine whether they have retained a PL. Furthermore, the expression of *fas2A* in an INP provides the first evidence that they encode spatial information for neural subtype identity.

4.10.1 *foxQ2* in the development of the arcuate body and PI

In *T. castaneum*, *foxQ2* is initially co-expressed with *six3*, *foxQ2*-positive neurons contribute to the central complex, and *foxQ2* knockdown leads to aberrant central complex development (Kitzmann *et al.*, 2017; He *et al.*, 2019). In addition to aberrant MB, but not their loss, although Posnien *et al.* (2023) suggested MB defects could result from indirect effects.

It has previously been shown that *six3.1* (Leite *et al.*, 2024) and *foxQ2* (Schacht *et al.*, 2020) are co-expressed at the anterior rim of the germ band at stage 7 (see General Discussion Figure 41). Here I show that anterior-most domain of *foxQ2* expression was nested in the *six3.2* domain and thus, *foxQ2* likely contributes to the development of the spider arcuate body. It is unclear if *foxQ2* is involved in the development of the spider MB, this may be resolved by co-expression of *foxQ2* and *dac*, which marks the developing MB (Doeffinger *et al.*, 2010), in a spider with robust MB such as *C. salei*.

In the marine annelid *P. dumerilii* *six3* (and *foxQ2*) mark the apical plate of trochophore larvae (Steinmetz *et al.*, 2010; Marlow *et al.*, 2014; Leite *et al.*, 2024). In *S. maritima*, *foxQ2*, along with *six3*, is expressed in invagination sites and ectodermal compartments of the centipede PI (Hunnekuhl and Akam, 2014). In *D. melanogaster*, *foxQ2* (*fd102C*) is also anteriorly expressed (Lee and Frasch, 2004) and may correlate with the location of PI/PL. In *T. castaneum*, *chx* expression initially largely overlaps with *foxQ2*, but they later resolve to mainly non-overlapping patterns (Kitzmann *et al.*, 2017). In *T. castaneum* *foxQ2* knock-down embryos, *chx* expression in the early elongating germ bands was completely lost and at later stages *chx* expression corresponding to the PI primordia was highly reduced (Kitzmann *et al.*, 2017). The role of *foxQ2* on beetle PL primordia was not investigated. This suggests *foxQ2*, like *six3*, functions high in the hierarchy of neuroendocrine development.

Previously, in *P. tepidariorum*, parental RNAi against the *six3* ohnologs and *foxQ2* was performed to investigate their role in labrum development (Schacht *et al.*, 2020). Future work in *P. tepidariorum*, would involve staining *six3.1*, *six3.2*, and *foxQ2* knockdown embryos for *vsx/chx* and *fas2A* to determine their role in PI and PL development, respectively. I show that

vsx/chx expression was nearly entirely nested in the *six3.2* expressing AMR. Thus, in *six3.2* knockdown embryos I expect *vsx/chx* expression corresponding to both the PI and labrum to be affected, similar to *six3* knockdown in the beetle (Posnien *et al.*, 2011b). I show *foxQ2* and *vsx/chx* look to be non-overlapping in the developing brain and labrum. Nevertheless, due to the early embryonic expression of *foxQ2*, in *foxQ2* knockdown embryos, I expect *vsx/chx* expression corresponding to both the PI and labrum to be affected, similar to *foxQ2* knockdown in the beetle (Kitzmann *et al.*, 2017).

4.10.2 *fas2A* in the development of the principal eyes and pioneer neurones

Here I show *P. tepidariorum fas2A* is expressed in the PEP at stage 10.2 and 12. The expression of *fas2A* in the PEP at later embryonic stage and SEP would be easily resolved with *so1* co-expression. In *D. melanogaster* Fas2 acts specifically to inhibit epidermal growth factor receptor (EGFR) signalling during the normal development of the eye (Mao and Freeman, 2009). The EGFR has multiple functions in eye development but one of its key roles is the recruitment of cells into the developing ommatidia. Loss of *fas2* function in the fly results in the recruitment of ectopic clusters and excess photoreceptors, specific characteristics of EGFR hyperactivity.

In *D. melanogaster*, Fas2 appears on the surface of clusters of neural somata before axon outgrowth (Nassif *et al.*, 1998). These clusters, called fibre tract founder clusters, pioneer the main axonal tracts in the brain (Nassif *et al.*, 1998). Thus, in the *P. tepidariorum* segmentally reiterated expression of *fas2A* could correspond to the position of the commissures and connectives.

4.10.3 *col*-expressing anterior neuronal cell population

In *S. maritima*, a *col/PC2*-expressing central cell population located in a *six3* and *foxQ2*-expressing domain pioneers the anterior CNS (Hunnekuhl and Akam, 2014). While in *P. tepidariorum*, I identified bilateral domains of *foxQ2* nested within the *six3.2* expressing neuroectoderm and within this domain *col1*-expressing cells. I cannot confirm whether this neuronal cell population forms an active early neurosecretory centre as the spider *PC2* ortholog has no detectable embryonic expression.

The *T. castaneum PC2* ortholog is first expressed during late embryogenesis shortly before hatching in the developing brain and ventral neuroectoderm (Fritzsche and Hunnekuhl, 2021). In larvae, *PC2* is expressed in the nervous system with stronger expression in the anterior medial brain, where the PI is located (Fritzsche and Hunnekuhl, 2021). Thus, insect and spider *PC2* orthologs are not expressed at early embryonic stages, in contrast to *S.*

maritima PC2. Based on the expression of the PC2 in *T. castaneum*, the spider ortholog may be expressed at later embryonic stages when cuticle formation prevents detection via ISH.

There is another subtilisin-like protease, PC1/3, that like PC2 is specifically restricted to endocrine and neuroendocrine cells (Lovejoy, 2005; Tessmar-Raible, 2007). PC1/3 has a pre-bilaterian origin and is found in representatives of each arthropod group but has been lost in *D. melanogaster* (Fritzsche and Hunnekuhl, 2021). The only arthropod PC1/3 ortholog that has been in *T. castaneum*. In the beetle, PC1/3 was not expressed embryonically and in larvae, no expression was observed in the PI or other cells of the protocerebrum (Fritzsche and Hunnekuhl, 2021). *P. tepidariorum* has a single PC1/3 ortholog [XP_015924123.1] (Fritzsche and Hunnekuhl, 2021) but its expression has not been characterised. In the spider I expect PC1/3 to be expressed similarly to the beetle ortholog. In conclusion, I expect that the spider *col1*-expressing cells I identified here, in contrast to the centipede *cpnc*, do not form an active early neurosecretory centre.

The long-range pioneer neurons originating from the *S. maritima* CPNCs have not been characterised in any other arthropod (Hunnekuhl and Akam, 2014). In most insects, neuroblasts that pioneer the axon tracts of the anterior CNS are located in the bilateral head neuroectoderm (Posnien *et al.*, 2011b; Urbach and Technau, 2003a). In the grasshopper *Schistocerca gregaria*, two neurons that originate directly from the epithelium within the anterior medial domain of the head pioneer the primary brain commissure (Boyan and Williams, 2008). Further work in *P. tepidariorum* would involve labelling the CNS with an anti-acetylated tubulin antibody to determine which cells are the first to differentiate into neurones. The spider *col1*-expressing cell identified here may not form an active early neurosecretory centre, however, they may correspond to the long-range pioneer neurons of the centipede and/or short-range pioneer neurons of the grasshopper. Consistent with this hypothesis, the later expression of spider *col1* looks as though it is involved in establishing the anterior commissure.

4.10.4 The spider anterior furrow and its molecular similarities to the vertebrate cranial placodes and insect head placode

The anterior-most vertebrate placodes are the adenohipophyseal and olfactory placodes, from which the anterior pituitary gland and olfactory system develop, respectively. Nearly all placodal cells and their primordia are marked by the expression of pan-placodal markers: *six1/2/so*, *eya*, and *six4/5*. Specific adenohipophyseal placode markers are *pitx* and *lhx3/lim3*.

The *D. melanogaster* placodes are morphologically visible, while the insect head placode (IHP) of *T. castaneum* is defined by its molecular similarity to vertebrate cranial placodes (Posnien *et al.*, 2011a). The beetle IHP is a morphological unit marked by *six4* expression located at the anterior-median position of the pre-antennal region (Posnien *et al.*,

2011a). The pre-placodal marker *Dlx3/5/Distal-less (Dll)* also appears expressed in the IHP (Posnien *et al.*, 2023). The IHP also expresses pan-placodal and small clusters of cells express specific adenohypophyseal placode markers (Posnien *et al.*, 2011a). The anterior and posterior IHP are marked by *six3* and *otd1* expression, respectively. Furthermore, *ey* and *bf1/sloppy paired (slp)* are small cell clusters or individual cells of the IHP. Expression of *vsx/chx* and *fas2* corresponding to the presumptive beetle PI and PL, respectively, fall within the IHP and *six4* contributes to the beetle PI. In *D. melanogaster* *six4* is a marker for type II neuroblast-derived lineages (Chen *et al.*, 2021) which produce the columnar cells of the PI that innervate the central complex (Boyan and Reichert, 2011).

The spider anterior furrow also has molecular similarities to the beetle IHP and vertebrate adenohypophyseal placode. Similar to the beetle, in *P. tepidariorum*, *so1* and *eya* are expressed at the periphery of the pre-chelicerar neuroectoderm (Schomburg *et al.*, 2015; Baudouin Gonzalez *et al.*, 2024). Later *so1* expression splits into two domains corresponding to the PEP and SEP (Schomburg *et al.*, 2015). *eya* remains expressed in the non-neurogenic ectoderm, PEP and SEP, and future anterior furrow (Baudouin Gonzalez *et al.*, 2024). *P. tepidariorum* has two *pitx* orthologs, which are ohnologs retained from the arachnoplumonate WGD event (Aase-Remedios *et al.*, 2023), *pitx2* is expressed in the future anterior furrow at stage 10.1 overlapping with *vsx/chx* and *fas2* expression domains corresponding to the spider PI and PL primordia (Leite *et al.*, 2018). The pre-placodal marker *Dlx3/5/Dll* is also expressed in the future anterior furrow from stage 7 (Pechmann *et al.*, 2011). *P. tepidariorum* has a single *six4* ortholog, the spider ohnolog has been lost, and two *lhx3/lim3* orthologs, which are the result of a recent tandem duplication event (Aase-Remedios *et al.*, 2023), but their expression has not been characterised in this spider or any other chelicerate.

Future work in *P. tepidariorum* would involve characterising the expression of *six4* and *lhx3/lim3* paralogs to provide further evidence for a similar region to the insect IHP and vertebrate adenohypophyseal placode in the spider. *T. castaneum*, does not have ocelli, but the IHP has molecular similarity to the region from which the ocelli emerge in the grasshopper *S. gregaria* (Posnien *et al.*, 2023). Thus, *six4* may also be a candidate for principal eye development in the spider.

4.10.5 *nkx2.1/scro* is expressed in the primordia of arthropod, polychaete, and vertebrate neurosecretory centres

In vertebrates, *nkx2.1* is a marker for the medial forebrain and is expressed in the anterior neuroectoderm in most bilaterians (Tessmar-Raible, 2007). In the annelid *P. dumerilii*, *nkx2.1/scro* is expressed in the apical plate including the apical organ (Tessmar-Raible, 2007; Tessmar-Raible *et al.*, 2007).

In *D. melanogaster*, from stage 13 *nkx2.1/scro* is expressed in bilateral domains in the developing protocerebrum (Zaffran *et al.*, 2000) however, this expression has not been described with respect to the PI/PL primordia. At larval stages, *nkx2.1/scro*-expressing neurones are present in the CNS including the PI region (Yoo *et al.*, 2020). By comparing single *in situ* staining's in *T. castaneum*, some cells of the developing PI appear to express *nkx2.1/scro* (Posnien *et al.*, 2011b). Similarly, I found in *P. tepidariorum*, that some cells of the PI placode express *nkx2.1/scro*. This contrasts observations in *S. maritima*, where *nkx2.1* is absent from invagination sites and ectodermal compartments of the PI (Hunnekuhl and Akam, 2014). In the centipede, the anterior part of the ring-like *nkx2.1* domain overlaps with *foxQ2* expression and the stomodeum develops in the centre of this domain (Hunnekuhl and Akam, 2014). Thus, *nkx2.1/scro*, like *six3* (Tessmar-Raible, 2007), is expressed in primordia of arthropod, polychaete, and vertebrate neurosecretory centres. Their exact evolutionary relationships are under discussion (Tessmar-Raible, 2007). Consistent with other invertebrates, I found *P. tepidariorum nkx2.1/scro* is expressed in the stomodeum primordia.

4.10.6 *raxA* and *hbn* candidates for spider arcuate body development

In vertebrates, *rx* (and *otp*) are expressed in the medial forebrain and its later derivatives, including the hypothalamus, preoptic area, and cells of the eye and optic stalk (Tessmar-Raible, 2007). In the annelid *P. dumerilii*, *rax* is expressed peripherally in the apical plate of the trochophore larva and in the centre of the apical plate in metatrochophore larva (48 hpf) (Tessmar-Raible *et al.*, 2007; Marlow *et al.*, 2014). In *Strigamia rax* (and *hbn*) are initially co-expressed with *six3* (and *foxQ2*) in the medial forebrain during the specification and internalisation of the CPNCs (Hunnekuhl and Akam, 2014).

In insects and *P. tepidariorum rax* is not expressed in the medial forebrain. In both *T. castaneum* and *D. melanogaster*, *rax* seems to be largely or even exclusively co-expressed with *otd* and there is a strong contribution of *rax*-positive neurons to both, columnar and tangential neurons of the central complex (Farnworth *et al.*, 2020; Davis *et al.*, 2003). Furthermore, in both beetles and flies *rax* lack of function leads to aberrant central complex development (Davis *et al.*, 2003; Koniszewski, 2011). Here I show *P. tepidariorum raxA* (and *hbn*) are mutually exclusive of *six3.2* expression and localise to the lateral part of the anterior furrow and thus, likely contribute to the development of the spider arcuate body. This provides evidence that the spider arcuate body, like the insect central complex (Posnien *et al.*, 2023), is derived from both *six3*- and *otd*-expressing protocerebrum (see General Discussion Figure 41).

4.10.7 The role *otp* in spiders is unclear

In the vertebrate hypothalamus, *otp* is required for the specification of neurosecretory hypothalamic neurons (Wang and Lufkin, 2000). Similarly, in the marine annelid *P. dumerilii* *otp* demarcates various sensory-neurosecretory cells in the central *six3*-free territory of the apical plate, which project to the neurosecretory fibre plexus (Tessmar-Raible *et al.*, 2007; Marlow *et al.*, 2014). In *S. maritima*, *otp* is expressed in some PI cells and CPNCs (Hunnekuhl and Akam, 2014).

In *D. melanogaster*, Otp has a role in hindgut development (Hildebrandt *et al.*, 2020). Otp is also expressed in lineages of the protocerebrum and deutocerebrum in the fruit fly (Hildebrandt *et al.*, 2020; Hildebrandt *et al.*, 2022). A small portion of the *otp*-expressing neural cell populations is located in the region which gives rise to the neuroendocrine cell populations (Hildebrandt *et al.*, 2022). However, Hildebrandt *et al.* (2022) suggest it is highly unlikely that any of the NSCs are Otp-positive although, no double-labeling of Otp in conjunction with neuropeptide markers was performed. The spider *otp* ortholog may be expressed at later embryonic stages when cuticle formation prevents detection via ISH. Thus, the role of *otp* in neuroendocrine development in insects and spiders is unclear.

5. Chapter Five Results: Analysis of the retention of ohnologs in key developmental genes following whole-genome duplication in arachnopolmonates

5.1 Background: Evidence of WGD in arachnopolmonates

Whole-genome duplications (WGDs) have occurred multiple times during animal evolution, including in lineages leading to vertebrates, teleosts (Glasauer and Neuhauss, 2014), horseshoe crabs (Kenny *et al.*, 2016; Nong *et al.*, 2021), and arachnopolmonates (Schwager *et al.*, 2017). Following WGD, gene copy numbers are doubled and retained paralogs (ohnologs) can then share ancestral functions (sub-functionalisation) and/or evolve new roles (neo-functionalization) or return to a single copy (non-functionalisation). The two rounds of WGD that occurred in the vertebrate (R1) and gnathostome stem-lineage (R2) have been suggested to facilitate the evolution of vertebrate novelties (Shimeld and Holland, 2000; Cañestro *et al.*, 2013; Yu *et al.*, 2024). However, in other animal lineages how the WGD has contributed to diversification is not well understood.

While WGD is typically followed by extensive gene loss, many ohnologs retained belong to the developmental toolkit of TFs, signaling pathway components, and microRNAs (Erwin, 2009). One TF family that has stood out are the Hox homeobox genes, which regulate the identity of the body plan along the anterior-posterior axis of all bilaterian animals (Pearson *et al.*, 2005; Hueber and Lohmann, 2008; Turetzek *et al.*, 2024). In vertebrates, four Hox clusters have partially retained following 1R and 2R (Holland *et al.*, 1994; Meyer and Scharl, 1999; Pascual-Anaya *et al.*, 2013). Evidence of a WGD in the lineage leading to arachnopolmonates, a clade of five arachnids that have book lungs, has predominantly come from two lineages, spiders and scorpions. Many spider homeobox genes and clusters, including two Hox clusters, were retained in duplicate following WGD (Schwager *et al.*, 2017; Leite *et al.*, 2018; Aase-Remedios *et al.*, 2023). However, the derived retrolateral tibial apophysis (RTA) clade, which includes half of all extant spider species has not been surveyed. The RTA clade includes families of hunting spiders, jumping spiders (Salticidae) and wolf spiders (Lycosidae). Duplicates of homeobox genes have also been retained in scorpions while arachnids lacking WGD, such as ticks, mites, and harvestmen, exhibit single copies (Leite *et al.*, 2018; Aase-Remedios *et al.*, 2023). In spiders, Hox ohnologs exhibit differences in spatial and temporal expression but for each Hox cluster the overall spatial and temporal collinearity is largely intact (Turetzek *et al.*, 2024). In *P. tepidariorum*, several other homeobox genes exhibit divergent expression consistent with both sub- and neo-functionalisation (Leite *et al.*, 2018; Aase-Remedios *et al.*, 2023). Thus, homeobox duplications have played a role in the evolution of spider development.

P. tepidariorum has also been shown to have retained duplicates of signaling pathway components and microRNAs (Janssen *et al.*, 2010; Janssen *et al.*, 2015; Leite *et al.*, 2016; Schwager *et al.*, 2017). Wnt genes encode secreted glycoprotein ligands that regulate several developmental processes, including segmentation (Murat *et al.*, 2010). There are thirteen subfamilies of metazoan *Wnt* genes (*Wnt1-11*, *Wnt16* and *WntA*), with *Wnt3* lost in the lineage leading to protostomes (Janssen *et al.*, 2010; Cho *et al.*, 2010). Further *Wnt* gene losses have been suggested within arthropods: *Wnt2* and *Wnt4* in Insecta, *Wnt16* in holometabolous orders, *Wnt11* in *Diptera*, and *WntA* in *D. melanogaster* (Janssen *et al.*, 2010). In arthropods, *Wnt* gene duplications were previously only observed in *P. tepidariorum*, the only arachnoplumonate surveyed (Janssen *et al.*, 2010). *P. tepidariorum* possesses eleven *Wnt* genes, including duplicates of *Wnt7* and *Wnt11*, while orthologs of *Wnt9* and *Wnt10* were not recovered (Janssen *et al.*, 2010). However, whether this pattern of retention and loss is similar in other spiders and arachnoplumonates lineages is unknown. In vertebrates, several *Wnt* families have been retained in duplicate after the first two rounds of WGD for example, humans have two copies of *Wnt2*, *Wnt3*, *Wnt5*, *Wnt7*, *Wnt8*, *Wnt9* and *Wnt10* (Janssen *et al.*, 2010). Thus, *Wnt* genes may be more conservative markers of WGD, to be used in combination with the *Hox* genes as well as other homeobox genes.

There are four subfamilies of metazoan *frizzled (fz)* genes (*FzI-IV*) (Janssen *et al.*, 2015). In arthropods, there have been lineage-specific losses such as *fz3* in the beetle *T. castaneum* (Beermann *et al.*, 2011). However, again duplications were restricted to the arachnoplumonates *P. tepidariorum* and the scorpion *Me. martensii*, which both possess a duplication of *fz4* and appear to have lost orthologs of *fz2* and *fz3*, respectively (Janssen *et al.*, 2015). Again, whether this pattern of retention and loss is similar in other spiders, scorpions, or other chelicerates with and without an ancestral WGD is unknown.

Beyond spiders and scorpions, genomic and transcriptomic information for the three other arachnoplumonate lineage is lacking for comparison (see Gainett and Sharma (2020), Gainett *et al.* (2020), and Ontano *et al.* (2021) for recent advances). One such lineage is the amblypygids or whip spiders with 190 extant species (Garwood *et al.*, 2017).

5.1.1 Chapter five aims

This chapter aimed to better understand the genomic consequences of the WGD in a greater diversity of arachnoplumonate lineages. I surveyed the newly sequenced transcriptomes of a jumping spider *Marpissa muscosa* and wolf spider *Pardosa amentata*, which both belong to the derived RTA clade, in addition to two amblypygids *Charinus acosta* (Charinidae) and *Euphrynichus bacillifer* (Neoamblypygi), for key developmental gene families and sensory/neurosecretory TFs. In addition to twelve existing transcriptomic and genomic resources for chelicerates, with and without an ancestral WGD.

Objectives:

- Survey the Hox gene repertoires and determine the origin of duplicates by phylogenetic analysis
- Survey the Wnt gene repertoires and determine the origin of duplicates by phylogenetic analysis
- Survey the frizzled gene repertoires and determine the origin of duplicates by phylogenetic analysis
- Survey homeobox genes *nkx2.1/scro*, *otp*, *rax*, *vax*, and *vsx/chx* and determine the origin of duplicates by phylogenetic analysis

The results presented here are published in Harper et al. (2021) except for sensory/neurosecretory TF repertoires and their origins.

5.2 Results: *Hox* repertoires and their origins

It was previously shown that duplicates of all ten Hox genes ancestral to Panarthropoda have been retained in the scorpion *Mesobuthus martensii* (Di et al., 2015); all except *Hox3* in the scorpion *Centruroides sculpturatus* (Leite et al., 2018); and all except *fushi tarazu* (*ftz*) in the spider *P. tepidariorum* (Schwager et al., 2017) (Figure 33). While single copy orthologs were found in the non-arachnopulmonate arachnids: the tick *Ixodes scapularis* (Leite et al., 2018) and the harvestman *Phalangium opilio* (Sharma et al., 2012) (Figure 33).

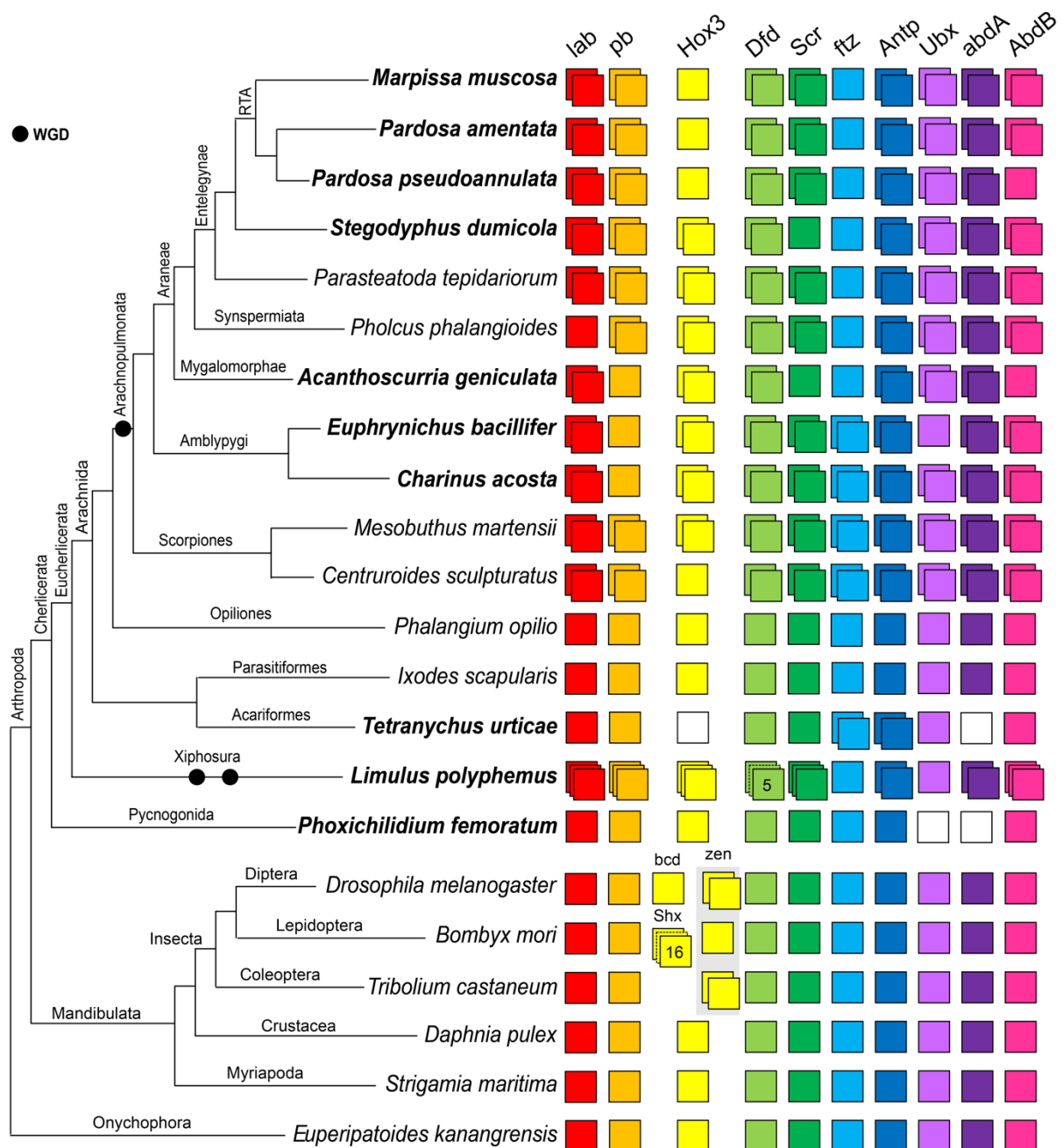


Figure 33. Repertoires of Hox genes in arachnids and other selected arthropods.

Species surveyed (bold) for Hox genes (coloured boxes), depicting duplicates (overlapping boxes) and putative losses (white boxes). The insect *Hox3* homolog *zen* has undergone independent tandem duplications in *T. castaneum* to yield *zen* and *zen2*; in cyclorrhaphan flies to yield *zen* and *bicoid*; and in the genus *Drosophila* to yield *zen2*. *Bombyx mori* is not representative of all species of ditrysian Lepidoptera, which typically possess four distinct *Hox3* genes termed *Special homeobox* genes (*ShxA*, *ShxB*, *ShxC*, and *ShxD*) and the canonical *zen* gene. Adapted from Harper et al. (2021).

Using reciprocal blast and phylogenetic analysis I found the spider *Hox* gene repertoires were largely consistent with *P. tepidariorum* (Figure 33). There were several exceptions: single copies of *Hox3* in *M. muscosa*, *P. amentata*, and the genome of *Pardosa*

pseudoannulata, of *Sex combs reduced* (*Scr*) in transcriptome for *Acanthoscurria geniculata* and genome of *Stegodyphus dumicola*, of *proboscipedia* (*pb*) in the transcriptome of *A. geniculata*, and of *labial* (*lab*) in the transcriptome of *Pholcus phalangioides* (Figure 33).

Both amblypygids had duplicates of all *Hox* genes except *pb* and *E. bacillifer* only had one copy of *Ultrabithorax* (*Ubx*) (Figure 33).

In the non-arachnopulmonate arachnid surveyed, the mite *Tetranychus urticae*, I did not identify *Hox3* and *abdA* candidates but recovered *ftz* duplicates, consistent with Grbic et al. (2011).

I surveyed two non-arachnid chelicerates, the transcriptome of a pycnogonid *Phoxichilidium femoratum* and the genome of the horseshoe crab *Limulus polyphemus*. In *P. femoratum*, I recovered single copies of all *Hox* genes except *Ubx* and *abdA*, consistent with (Ballesteros et al., 2021) (Figure 33). *L. polyphemus* has undergone multiple rounds of WGD independently of the arachnopulmonates (Nong et al., 2021; Kenny et al., 2016). Consistent with this, while I identified single copies of *ftz* and *Ubx*, I found three copies of *lab*, *pb*, *Hox3*, *Scr*, and *AbdB*, five potential copies of *Dfd*, and two copies of *Antp* and *abdA* (Figure 33).

I used both homeobox and homeodomain trees to confirm *Hox* gene identity relationships (Harper et al., 2021). Full protein sequences produced the best-supported phylogeny (Figure 34). The homedomain tree contained insufficient phylogenetic information to resolve within-clade relationships (Harper et al., 2021). The homeobox tree provided more within-clade resolution but with lower support than full protein sequences (Harper et al., 2021). *T. urticae* Ftz-1, Ftz-2, AbdB, and Ubx sequences (Grbic et al., 2011) did not resolve correctly in the full *Hox* tree (Figure 34) but did in the homeodomain and homeobox trees (Harper et al., 2021). I recovered *Antp* as a paraphyletic group using full-length protein sequences, albeit with low support (61%; Figure 34). Both homeobox and homeodomain phylogenies confirm the identity of these sequences, indicating substantial sequence divergence outside of this conserved region in several species.

The resolution of duplicate *Hox* sequences indicates that the majority are likely to be ohnologs. In many cases, duplicates form well-supported clades of orthologs, containing sequences from multiple orders, suggesting a shared origin for duplication. For instance, spider and scorpion *Pb* duplicates are candidate ohnologs, resolving in two well-supported ($\geq 77\%$) clades of orthologs whose topologies reflect phylogenetic relationships. I found few duplicates that formed species paralog pairs or lineage ortholog pairs as would be expected from lineage-specific duplications. However, there are exceptions such as scorpion *Lab* sequences (Figure 34).

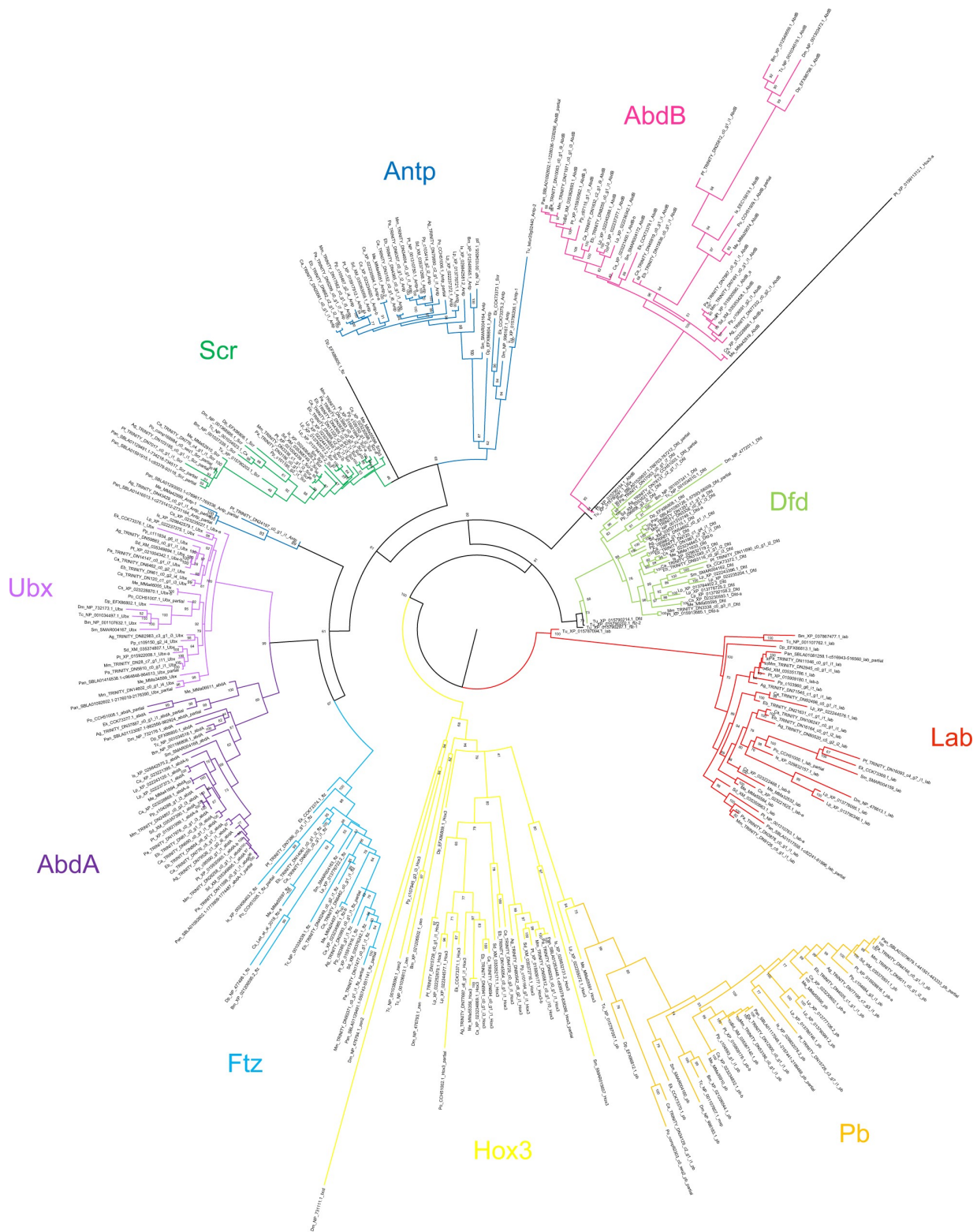


Figure 34. Maximum-likelihood phylogeny of Hox amino acid sequences.

The Hox subfamilies are shown in different colours (after Figure 33). Panarthropods included: *Acanthoscurria geniculata* (Ag), *Bombyx mori* (Bm), *Centruroides sculpturatus* (Cs), *Charinus acosta* (Ca), *Drosophila melanogaster* (Dm), *Euperipatoides kanangrensis* (Ek), *Euphrynichus bacillifer* (Eb), *Ixodes scapularis* (Is), *Limulus polyphemus* (Lp), *Marpissa muscosa* (Mm), *Mesobuthus martensii* (Me), *Parasteatoda tepidariorum* (Pt), *Pardosa amentata* (Pa), *Pardosa pseudoannulata* (Pan), *Phalangium opilio* (Po), *Pholcus phalangioides* (Pp), *Phoxichildium femoratum* (Pf), *Stegodyphus dumicola* (Sd), *Strigamia maritima* (Sm), *Tetranychus urticae* (Tu), and *Tribolium castaneum* (Tc). Node labels indicate ultrafast bootstrap support values. Published in Harper et al. (2021).

5.3 Wnt repertoires and their origins

Again, using reciprocal blast and phylogenetic analysis, I found the spider *Wnt* gene repertoires were largely consistent with *P. tepidariorum* (Janssen *et al.*, 2010) (Figure 35). There were a few exceptions: a single copy of *Wnt11* in *A. geniculata* and a second copy of *Wnt4* in *M. muscosa*, *P. amentata*, *P. pseudoannulata*, *S. dumicola*, and *A. geniculata* (Figure 35). Furthermore, I also found duplicates of *Wnt1/wg* in both *P. pseudoannulata* and *A. geniculata*, and a duplicate of *WntA* in *P. pseudoannulata* (Figure 35).

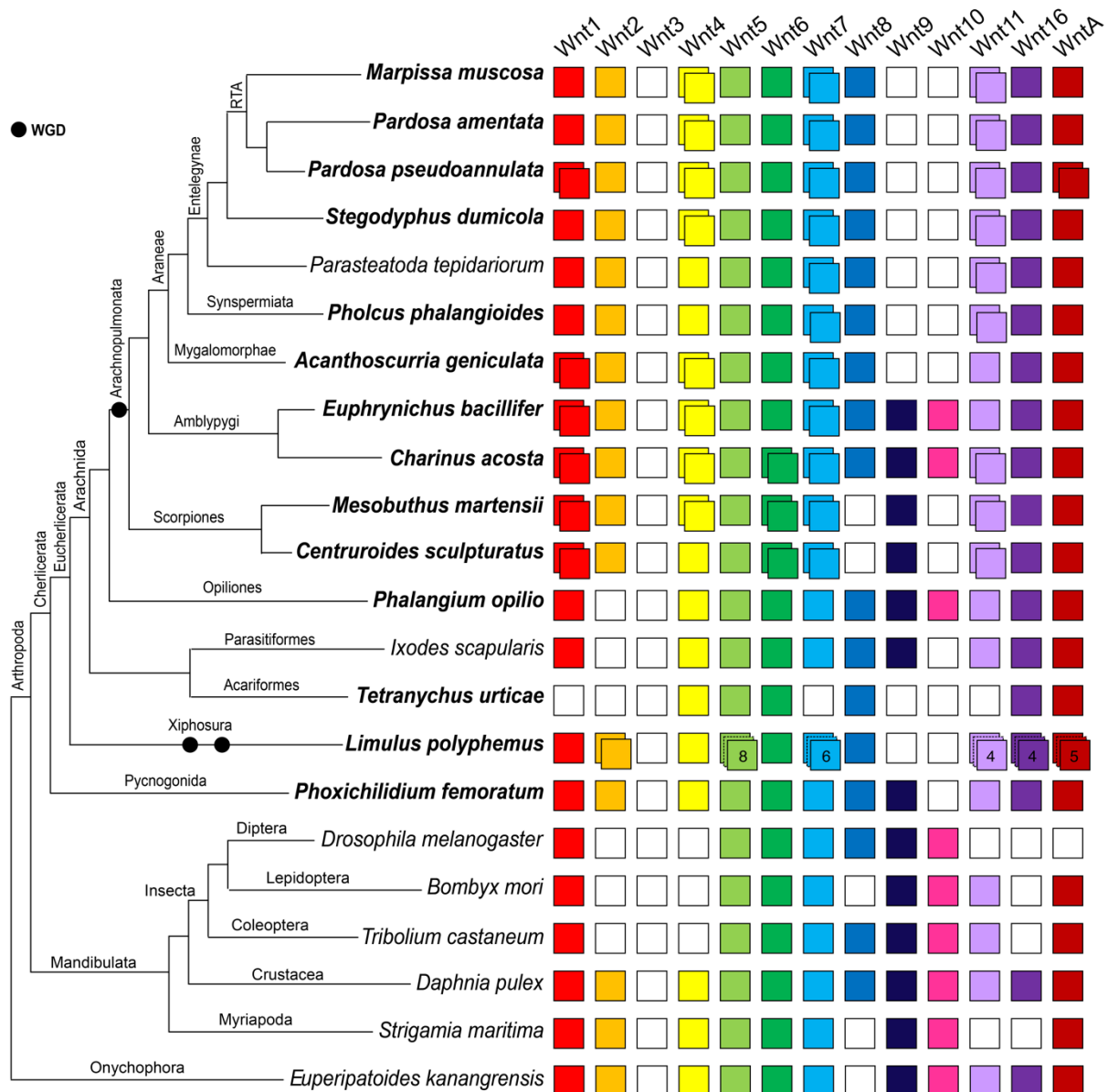


Figure 35. Repertoires of *Wnt* subfamilies in arachnids and other selected arthropods.

Species I surveyed (bold) for *Wnt* genes (coloured boxes), depicting duplicates (overlapping boxes) and putative losses (white boxes). Adapted from Harper et al. (2021).

In the amblypygids, I recovered a complete complement of arthropod *Wnt* subfamilies and duplicates of *Wnt1/wg*, *Wnt4*, and *Wnt7* in both species (Figure 35). Additional duplicates of *Wnt6* and *Wnt11* were recovered from *C. acosta* (Figure 35).

In the two scorpion genomes, representatives of ten out of twelve arthropod *Wnt* subfamilies were found, with both missing *Wnt8* and *Wnt9* (Figure 35). Two copies of *Wnt1/wg*, *Wnt6*, *Wnt7*, and *Wnt11* were retrieved in both species, with an additional copy of *Wnt4* in *Me. martensii* (Figure 35).

In the non-arachnopulmonate arachnids, I recovered single copies of all *Wnt* genes except *Wnt2* (Figure 35). I found the tick *I. scapularis* was also missing *Wnt10* and in the mite *T. urticae* I did not recover these or *Wnt1/wg*, *Wnt7*, *Wnt9*, and *Wnt11* (Figure 35).

In the non-arachnid chelicerates, *P. femoratum* and *L. polyphemus*, I found similar representation of the subfamilies, with both missing *Wnt10*, and *Wnt9* missing in *L. polyphemus*. Large numbers of duplicates were found in *L. polyphemus* including six potential copies of *Wnt5* and *Wnt7*, four of *Wnt11* and *Wnt16*, and five of *WntA* (Figure 35).

Based on phylogenetic analysis of full *Wnt* peptide sequences, *Wnt* paralogs appear to stem from the arachnopulmonate WGD event, although confirmation requires synteny analysis (Figure 36).

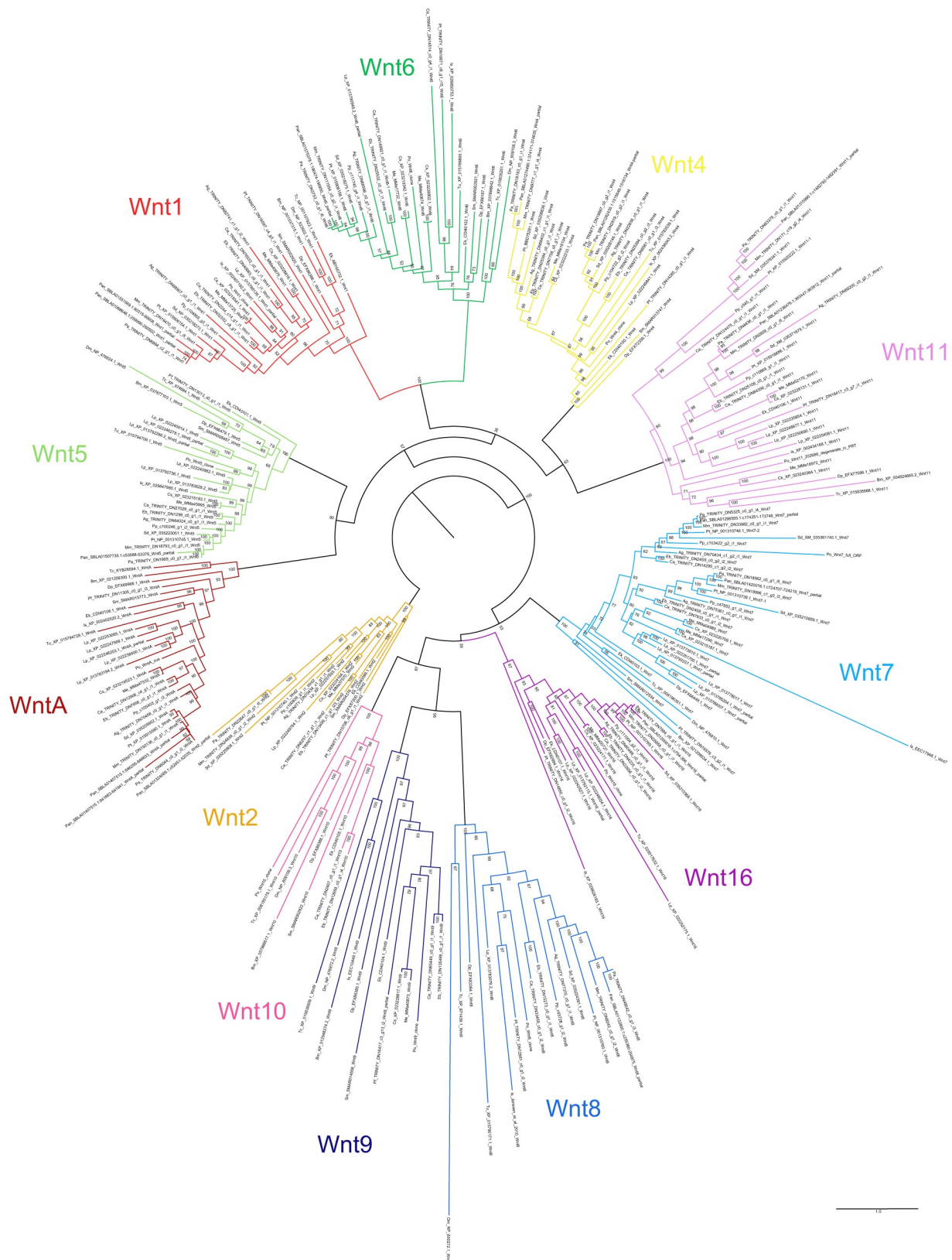


Figure 36. Maximum-likelihood phylogeny of Wnt amino acid sequences.

The 12 Wnt subfamilies are shown in different colours (after Figure 35). Panarthropods included: *Acanthoscurria geniculata* (Ag), *Bombyx mori* (Bm), *Centruroides sculpturatus* (Cs), *Charinus acosta* (Ca), *Drosophila melanogaster* (Dm), *Euperipatoides kanangrensis* (Ek), *Euphrynichus bacillifer* (Eb), *Ixodes scapularis* (Is), *Limulus polyphemus* (Lp), *Marpissa muscosa* (Mm), *Mesobuthus martensii* (Me),

Parasteatoda tepidariorum (Pt), *Pardosa amentata* (Pa), *Pardosa pseudoannulata* (Pan), *Phalangium opilio* (Po), *Pholcus phalangioides* (Pp), *Phoxichildium femoratum* (Pf), *Stegodyphus dumicola* (Sd), *Strigamia maritima* (Sm), *Tetranychus urticae* (Tu), and *Tribolium castaneum* (Tc). Node labels indicate ultrafast bootstrap support values. Node labels indicate ultrafast bootstrap support values. Published in Harper et al. (2021).

5.3.1 Arachnopulmonates retained *Wnt1/wg* ohnologs

I identified two copies of *Wnt1/wg* in the transcriptomes of both amblypygids and *A. geniculata*, and the genomes of *P. pseudoannulata*, *C. sculpturatus*, and *M. martensii*. The putative duplicates in *P. pseudoannulata* resolved together with 100% support (Figure 36). These two sequences are partial and have a short (19 amino acids) nonidentical overlapping region therefore, it is unclear whether they represent a genuine, lineage-specific duplication. The amblypygid *Wnt1/wg* ortholog pairs resolved together with 100% support. The scorpion *Wnt1/wg* ortholog pairs resolved together with 100% support. The *A. geniculata* *Wnt1/wg* duplicates form separate clades with sequences from spiders (92%), and one ortholog pair from scorpions and amblypygids ($\geq 70\%$; Figure 36). The other scorpion and amblypygid ortholog pair resolved with single copies from the non-arachnopulmonate arachnids and the horseshoe crab ($\geq 84\%$; Figure 36). This suggests that the *Wnt1/wg* duplicates are ohnologs retained from the arachnopulmonate WGD event.

5.3.2 Arachnopulmonates retained *Wnt4* ohnologs

Wnt4 paralogs from *M. muscosa*, *P. amentata*, *A. geniculata*, *P. pseudoannulata*, and *S. dumicola* form two separate and well-supported clades (100% and $\geq 91\%$; Figure 36). The amblypygid *Wnt4* ortholog pairs resolved together with 100% support and sisters to the spider clades (99% and 100%; Figure 36). *Wnt4* duplicates likely represent retained ohnologs following the arachnopulmonate WGD, despite being lost in *P. tepidariorum* (Janssen et al., 2010) and *P. phalangioides*.

5.3.3 Scorpions and amblypygids retained *Wnt6* ohnologs

The scorpion *Wnt6* ortholog pairs resolved together with 100% support. As expected, one Ca-*Wnt6* duplicate and the *E. bacillifer* *Wnt6* sequence resolved together with 100% support. These resolved in a clade with the single copy spider, horseshoe crab, and harvestman *Wnt6* sequences with low support ($\geq 50\%$; Figure 36). The other Ca-*Wnt6* duplicate resolved in a clade with the remaining non-arachnopulmonate arachnid and pycnogonid sequences ($\geq 84\%$) and as sister to the former clade (95%; Figure 36). Although it is not completely clear from the phylogenetic analysis the *Wnt6* duplicates are likely ohnologs retained from the arachnopulmonate WGD event.

5.3.4 Spiders and amblypygids retained *Wnt7* ohnologs

Duplicates of *Wnt7* were previously identified in *P. tepidariorum* (Janssen *et al.*, 2010) and were recovered from all surveyed arachnopulmonates. The spider *Wnt7* duplicates formed two clades with high support ($\geq 61\%$ and $\geq 96\%$; Figure 36) suggesting the retention of ohnologs. The amblypygid *Wnt7* ortholog pairs resolved together with 100% support and sisters to the spider clades but with lower support (62% and 50%; Figure 36). The scorpion *Wnt7* ortholog pairs resolved together with 100% support and formed their own separate clade (92%), possibly indicating a lineage-specific duplication perhaps subsequent to the WGD (Figure 36).

5.3.5 Arachnopulmonates retained *Wnt11* ohnologs

Duplicates of *Wnt11* previously identified in *P. tepidariorum* (Janssen *et al.*, 2010) were recovered from all surveyed arachnopulmonates except from the transcriptomes of *A. geniculata* and *E. bacillifer* (Figure 36). The spider *Wnt11* duplicates formed two separate well-supported clades ($\geq 85\%$ and 100%) with each *C. acosta* *Wnt11* duplicate (98% and 99%; Figure 36). One *Ca-Wnt11* duplicate and the *E. bacillifer* *Wnt11* sequence resolved together with 100% support (Figure 36). I propose that the spider and amblypygid *Wnt11* duplicates are likely ohnologs retained from the ancestral WGD. The resolution of the scorpion *Wnt11* duplicates is less clear; ortholog pairs resolved together with 100% support, but only one resolved as sister to the two clear clades occupied by the spider and amblypygid duplicates (98%; Figure 36). The other scorpion *Wnt11* ortholog pair grouped with insect sequences (72%; Figure 36).

5.3.6 Unclear if *P. pseudoannulata* *WntA* duplication is true

The putative *WntA* duplicates recovered from the *P. pseudoannulata* genome were located on the same scaffold. These two sequences are partial and have a short (28 amino acids) nonidentical overlapping region therefore, it is unclear whether they represent a genuine, lineage-specific duplication.

5.4 *frizzled* repertoires and their origins

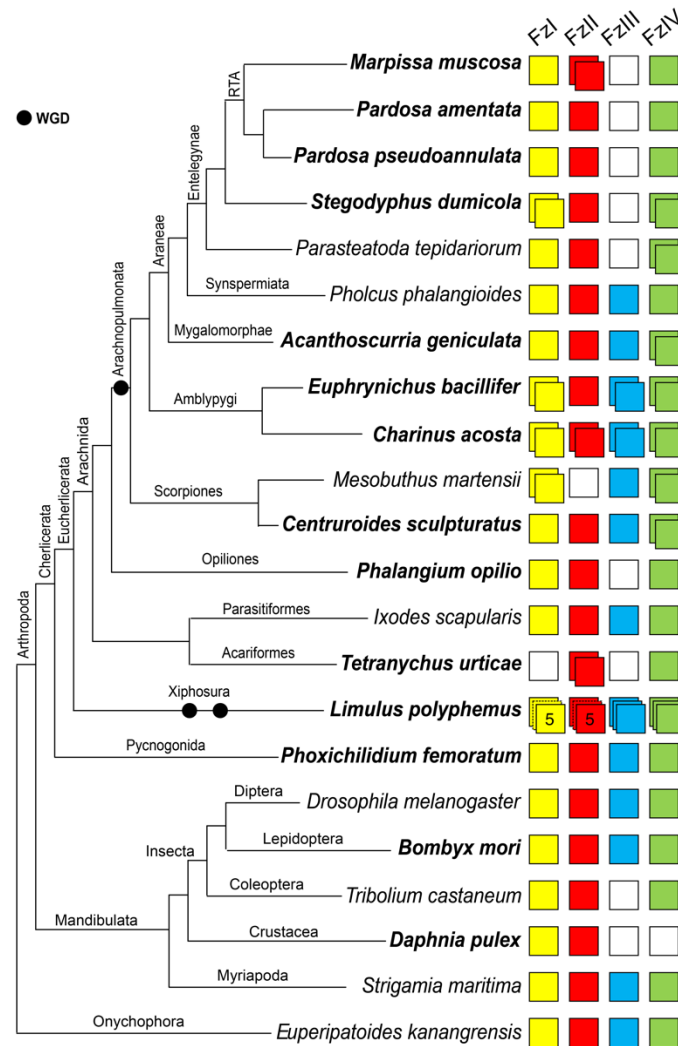


Figure 37. Repertoires of *frizzled* subfamilies in arachnids and other selected arthropods.

Species I surveyed (bold) for *frizzled* genes (coloured boxes), depicting duplicates (overlapping boxes) and putative losses (white boxes). Adapted from Harper et al. (2021).

Using reciprocal blast and phylogenetic analysis, I found in all surveyed spiders possess at least one ortholog of *fz1*, *fz2*, and *fz4* (Figure 37). A duplicate of *fz1* and *fz2* was recovered from *S. dumicola* and *M. muscosa*, respectively (Figure 37). *fz4* duplicates were recovered from *S. dumicola* and *A. geniculata* (Figure 37), consistent with *P. tepidariorum* (Janssen et al., 2015). *fz3* orthologs were not recovered from *M. muscosa*, *P. amentata*, *P. pseudoannulata*, and *S. dumicola* (Figure 37), consistent with *P. tepidariorum* (Janssen et al., 2015). A single *fz3* ortholog was recovered from *A. geniculata* (Figure 37), consistent with *P. phalangioides* (Janssen et al., 2015).

The two amblypygids surveyed have a complete complement of *frizzled* genes, with duplicates of *fz1*, *fz3*, and *fz4* orthologs in both species and a duplicate of *fz2* in *C. acosta* (Figure 37).

Consistent with the scorpion *Me. Martensii* (Janssen *et al.*, 2015), *C. sculpturatus* possessed a single *fz3* ortholog and a duplication of *fz4* (Figure 37). I recovered a *fz2* ortholog in *C. sculpturatus*, which was absent in *Me. martensii* (Figure 37). I also identified a *fz1* duplicate in *Me. martensii* not reported previously by Janssen *et al.* (2015), while a single copy was recovered from *C. sculpturatus* (Figure 37).

Among the non-arachnopulmonate arachnids, I recovered *fz3* in *I. scapularis* not reported previously by Janssen *et al.* (2015) and thus, *I. scapularis* possessed a single copy of all four *frizzled* genes, but *fz3* was absent in *P. opilio* (Figure 37). In *T. urticae*, *fz3* and *fz1* were absent, but I recovered a duplicate of *fz2* (Figure 37).

Both non-arachnid chelicerates surveyed have a complete complement of *frizzled* genes (Figure 37). I recovered five copies of *fz1* and *fz2* and three copies of *fz3* and *fz4* from *L. polyphemus* (Figure 37).

Based on phylogenetic analysis of full Frizzled peptide sequences, *frizzled* paralogs appear to stem from both lineage-specific duplications and the arachnopulmonate WGD event, although confirmation requires synteny analysis (Figure 38).

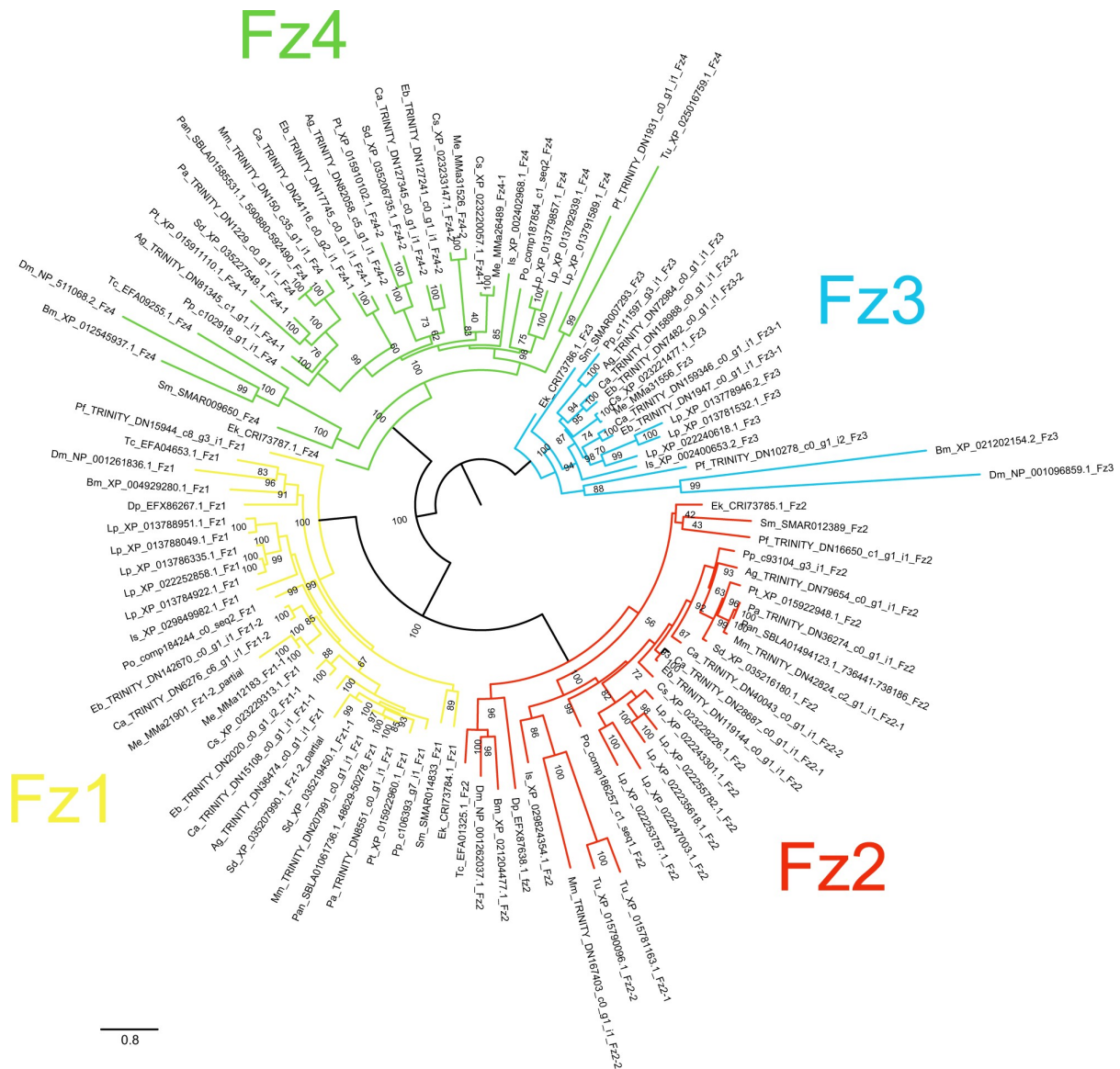


Figure 38. Maximum-likelihood phylogeny of Frizzled amino acid sequences.

The Frizzled subfamilies are shown in different colours (after Figure 37). Panarthropods included: *Acanthoscurria geniculata* (Ag), *Bombyx mori* (Bm), *Centruroides sculpturatus* (Cs), *Charinus acosta* (Ca), *Drosophila melanogaster* (Dm), *Euperipatoides kanangrensis* (Ek), *Euphrynichus bacillifer* (Eb), *Ixodes scapularis* (Is), *Limulus polyphemus* (Lp), *Marpissa muscosa* (Mm), *Mesobuthus martensii* (Me), *Parasteatoda tepidariorum* (Pt), *Pardosa amentata* (Pa), *Pardosa pseudoannulata* (Pan), *Phalangium opilio* (Po), *Pholcus phalangioides* (Pp), *Phoxichildium femoratum* (Pf), *Stegodyphus dumicola* (Sd), *Strigamia maritima* (Sm), *Tetranychus urticae* (Tu), and *Tribolium castaneum* (Tc). Node labels indicate ultrafast bootstrap support values. Node labels indicate ultrafast bootstrap support values. Published in Harper et al. (2021).

5.4.1 Amblypygids retained *fz1* ohnologs

I recovered *fz1* duplicates from *S. dumicola*, *M. martensii*, and both amblypygids. The Sd-Fz1 and Me-Fz1 paralog pairs resolved together with high support and are likely the result of

lineage-specific duplications (Figure 38). Conversely, the amblypygid Fz1 ortholog pairs resolved together with 100% support and formed separate clades with sequences from spiders ($\geq 67\%$) and scorpions ($\geq 85\%$; Figure 38). This suggests the amblypygid *fz1* duplicates are ohnologs.

5.4.2 Origin of spider and amblypygid *fz2* duplicates is unclear

The *T. urticae* Fz2 duplicates resolved together with 100% support, indicative of a lineage-specific duplication (Figure 38). One *M. muscosa* Fz2 copy and both *C. acosta* Fz2 duplicates resolved in a clade with the other spider Fz2 sequences ($\geq 63\%$; Figure 38). As expected, one Ca-Fz2 duplicate and the *E. bacillifer* Fz2 sequence resolved together with 100% support. The other Mm-Fz2 duplicate resolved in a clade with *I. scapularis* Fz2 and *T. urticae* Fz2 duplicates ($\geq 86\%$; Figure 38). This suggests that the amblypygid *fz2* duplicates were likely the result of a lineage-specific duplication while the spider duplicates may be ohnologs.

5.4.3 Amblypygids retained *fz3* ohnologs

The amblypygid Fz3 ortholog pairs resolved together with 100% support and formed separate well-supported clades with sequences from *L. polyphemus* ($\geq 70\%$) and arachnoplumonates ($\geq 94\%$) suggesting the retention of ohnologs (Figure 38).

5.4.4 Arachnoplumonates retained *fz4* ohnologs

Duplicates of *fz4* were previously identified in *P. tepidariorum* and *Me. martensii* (Janssen *et al.*, 2015) and were recovered here in the spiders *S. dumicola* and *A. geniculata*, the scorpion *C. sculpturatus*, and both amblypygids. The spider Fz4-1 duplicates formed a clade with the single copy spider Fz4 sequences ($\geq 76\%$; Figure 38). The spider Fz4-2 duplicates formed a well well-supported clade (100%; Figure 38). The amblypygid Fz4 ortholog pairs resolved together with 100% support and as sister to the two spider clades (99% and 73%; Figure 38). The scorpion Fz4 ortholog pairs resolved together with 100% support and formed their own separate clade with low support (40%), which was sister to all the spider and amblypygid Fz4 sequences (62%), possibly indicating a lineage-specific duplication (Figure 38). This suggests that the spider and amblypygid *fz4* duplicates were retained from the ancestral WGD.

5.5 Sensory/neurosecretory TF repertoires and their origins

I also surveyed the focal chelicerates for five homeobox TFs present in sensory/neurosecretory cells in Bilateria: *nkx2.1/scro*, *otp*, *rax*, *vax*, and *vsx/chx* (Tessmar-Raible, 2007). Spiders and the scorpion *C. sculpturatus* have retained a single copy of *nkx2.1* (Aase-Remedios *et al.*, 2023) (Figure 39). Spiders have retained single copies of *otp* while *C.*

sculpturatus possesses two copies (Aase-Remedios *et al.*, 2023) (Figure 39). Spiders and *C. sculpturatus* have retained duplicates of *rax* (Aase-Remedios *et al.*, 2023). In chelicerates, a single *vax* ortholog was previously only recovered in *P. opilio* (Aase-Remedios *et al.*, 2023) (Figure 39). Single orthologs of *vsx/chx* have been recovered in spiders while *C. sculpturatus* possesses two copies (Aase-Remedios *et al.*, 2023) (Figure 39).

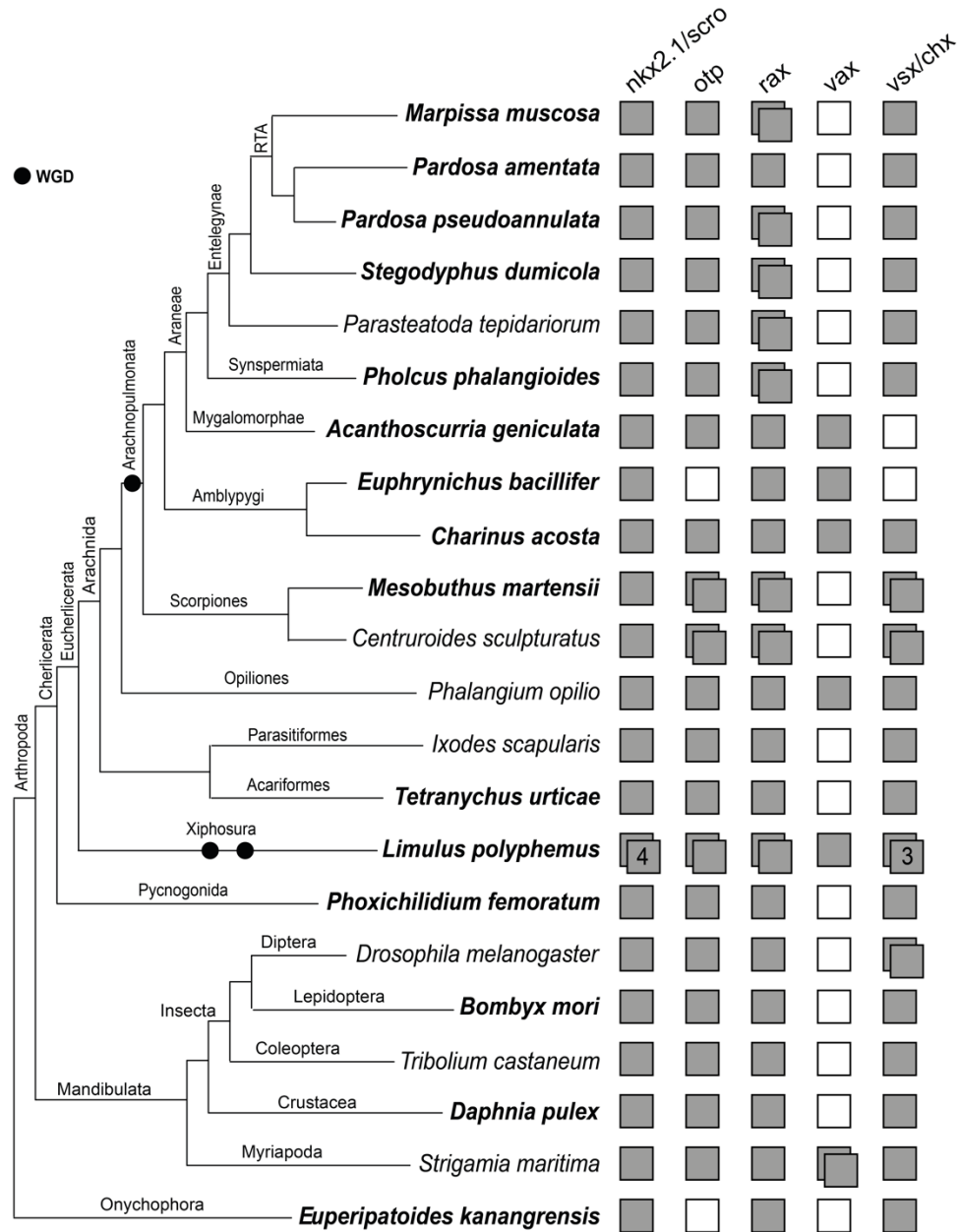


Figure 39. Repertoires of sensory/neurosecretory genes in arachnids and other selected arthropods.

Species I surveyed (bold) for sensory/neurosecretory genes (grey boxes), depicting duplicates (overlapping boxes) and putative losses (white boxes).

Using reciprocal blast and phylogenetic analysis, I found the spider sensory/neurosecretory gene repertoires were largely consistent with *P. tepidariorum* except

for a single copy of *rax* in *P. amentata* and *A. geniculata* (Figure 39). I did not recover a *vsx/chx* ortholog in *A. geniculata* however, I did cover a single *vax* ortholog.

In the amblypygid *C. acosta*, I recovered a complete complement of these five TFs while *otp* and *vsx* were not recovered from *E. bacillifer* (Figure 39).

Consistent with the scorpion *C. sculpturatus*, *Me. martensii* possessed duplicates of *rax*, *otp*, and *vsx/chx* (Figure 39).

In the non-arachnopulmonate arachnid, *T. urticae* I recovered single copies of all except *vax* (Figure 39), consistent with *I. scapularis*.

In the non-arachnid chelicerates, *P. femoratum* I recovered single copies of all except *vax* (Figure 39). In *L. polyphemus* I recovered a single copy of *vax*, two copies of *otp* and *rax*, three of *vsx/chx*, and four of *nkx2.1/scro* (Figure 39).

5.4.4 Spiders and scorpions retained *rax* ohnologs and scorpions retained *vsx/chx* and *otp* ohnologs

Based on phylogenetic analysis of full *rax*, *vsx/chx*, and *otp* peptide sequences. The scorpion *vsx/chx* ortholog pairs resolved together with 100% support and formed separate clades with sequences from *L. polyphemus*, *T. urticae*, spiders, and amblypygids (60%) and *I. scapularis* and *P. opilio* ($\geq 92\%$) suggesting the retention of ohnologs (Figure 40). The scorpion *otp* ortholog pairs resolved together with 100% support and although it is not completely clear from the phylogenetic analysis the *otp* duplicates are likely ohnologs retained from the arachnopulmonate WGD event (Figure 40). The spider *raxA* sequences formed a well-supported clade as did the arachnopulmonate *raxB* sequences (86% and $\geq 70\%$; Figure 40), suggesting the retention of ohnologs.

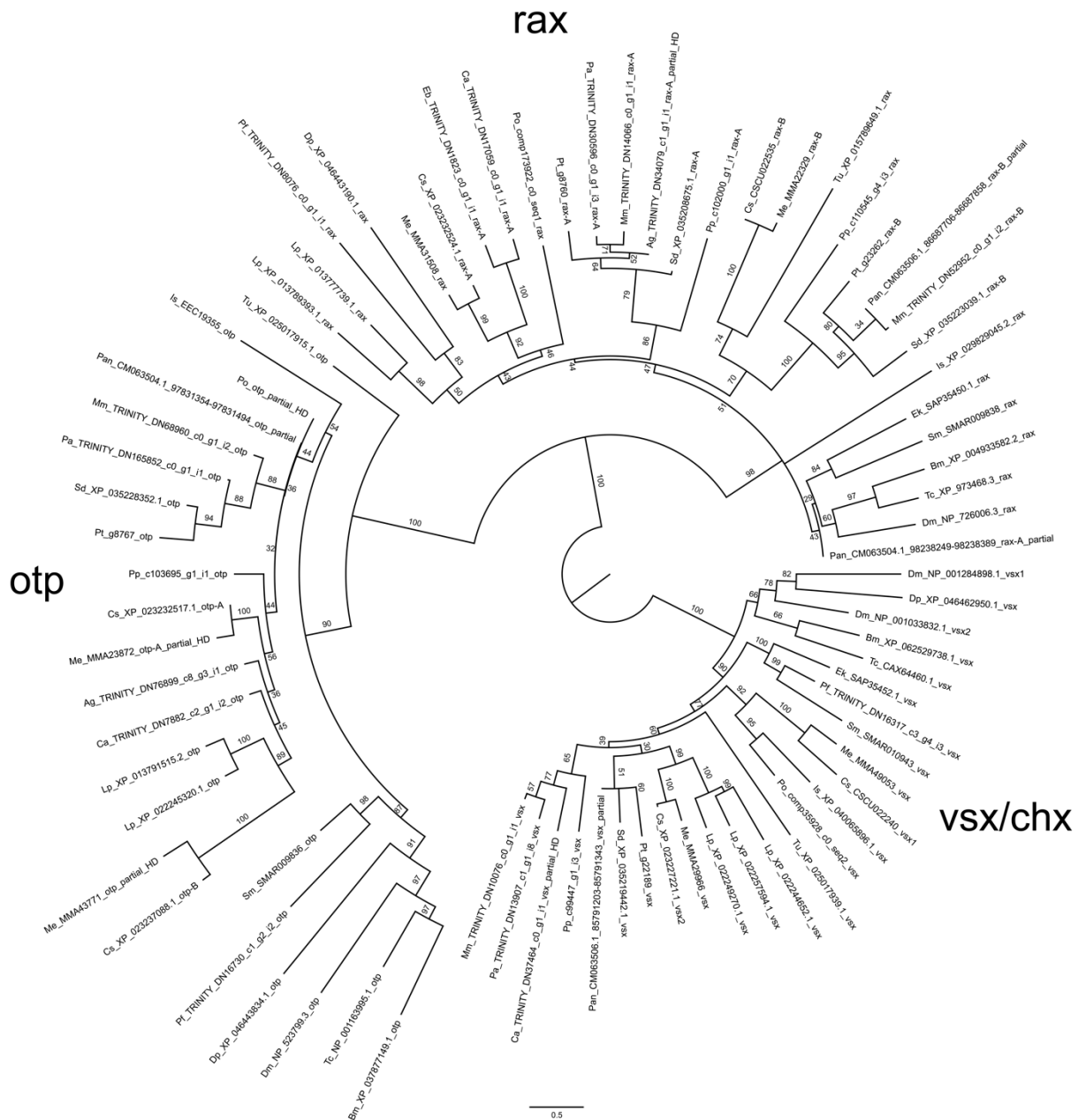


Figure 40. Maximum-likelihood phylogeny of *rax*, *otp*, and *vsx/chx* amino acid sequences.

Panarthropods included: *Acanthoscurria geniculata* (Ag), *Bombyx mori* (Bm), *Centruroides sculpturatus* (Cs), *Charinus acosta* (Ca), *Drosophila melanogaster* (Dm), *Euperipatoides kanangrensis* (Ek), *Euphrynichus bacillifer* (Eb), *Ixodes scapularis* (Is), *Limulus polyphemus* (Lp), *Marpissa muscosa* (Mm), *Mesobuthus martensii* (Me), *Parasteatoda tepidariorum* (Pt), *Pardosa amentata* (Pa), *Pardosa pseudoannulata* (Pan), *Phalangium opilio* (Po), *Pholcus phalangioides* (Pp), *Phoxichildium femoratum* (Pf), *Stegodyphus dumicola* (Sd), *Strigamia maritima* (Sm), *Tetranychus urticae* (Tu), and *Tribolium castaneum* (Tc). Node labels indicate ultrafast bootstrap support values. Node labels indicate ultrafast bootstrap support values.

5.6 Discussion

The improved taxonomic sampling provided further evidence for the ancestral arachnopulmonate WGD and allowed me to identify likely patterns of ohnolog retention and loss. Furthermore, I identified the first reported duplicates of *Wnt1/wg* in any animal and the first *Wnt10* in any arachnid.

5.6.1 Hox duplicates are broadly retained

The absence of a second copy of *pb* in two distantly related amblypygids is suggestive of a loss in the common ancestor of amblypygids. Gainett and Sharma (2020) also recovered a single copy of *pb* from the *Phrynos marginemaculatus* transcriptome, but two copies in *Charinus israelensis* transcriptome. However, one of these sequences (32 aa) had an incomplete homeodomain that was identical to the other copy; therefore, I am hesitant about its status as a duplicate. Embryos of *C. acosta* were collected at multiple stages of development, supporting the hypothesis that this is a genuine loss, rather than the absence of expression at a particular developmental stage.

A *ftz* duplicate has been retained in amblypygids and scorpions. A recent survey of homeobox genes across eight chromosome-level spider genomes also reports a single copy of *ftz* (Aase-Remedios *et al.*, 2023). This suggests, a loss *ftz* in the common ancestor of spiders. In both *P. opilio* (Sharma *et al.*, 2012) and *P. tepidariorum* (Schwager *et al.*, 2017), the single copy of *ftz* is expressed in walking leg pairs 2 to 4. Therefore, any sub- or neo-functionalisation that could be evident in scorpion and amblypygid *ftz* will have taken place after their divergence from spiders. Expression patterns of the two paralogs in these groups would be of interest for comparison.

I report an absence of a *Hox3* duplicate from the transcriptomes of *M. muscosa* and *P. amentata* and the genome of *P. pseudoannulata*, consistent with another RTA spider *C. salei* (Schwager *et al.*, 2017). Recently, a single copy of *Hox3* was recovered from the chromosome-level genomes of a Linyphiidae and three Araneidae spiders (Aase-Remedios *et al.*, 2023). Based on current phylogenies (Krehenwinkel *et al.*, 2019), this suggests the *Hox3* duplicate has been lost twice, once in the lineage leading to the RTA clade and another in the common ancestor of Linyphiidae and Araneidae. Spider genomes contain two Hox clusters, Hox Cluster B is typically intact, organised, and contains orthologs of all ten arthropod Hox genes. While Hox Cluster A lacks *ftz*, as well as *Hox3* in species missing a *Hox3* duplicate, and is sometimes rearranged. In *P. tepidariorum*, *Hox3-B* is predominantly expressed in the pedipalpal through to walking leg 4 from stage 8 (Schwager *et al.*, 2017) while *Hox3-A* was recently shown to be expressed earlier in the second and third walking leg segments (Leite *et*

al., 2024). The *Hox3* duplicate is highly divergent in spiders, which may have led to eventual pseudogenisation or neo-functionalisation.

Other apparent losses of *Hox* gene duplicates are restricted to transcriptomes of individual species, except for *Scr* in *S. dumicola* which could reflect a failure to capture additional sequences or indicate lineage-specific loss. If *lab* duplicates from *P. phalangioides* and *pb*, *Scr*, and *AbdB* duplicates from *A. geniculata* are absent from their genomes this could represent an interesting divergence between the three major groups of spiders. While these genes are unlikely to be directly responsible, the divergence might provide a starting point for understanding the important morphological differences between mygalomorphs, synspermiatans, and entelegynes. However, genomic information for additional taxa in both plesiomorphic clades is required to verify these potential losses.

5.6.2 Limited retention of Wnt duplicates

Across the arachnopulmonates surveyed, I recovered single copies of *Wnt2*, *Wnt5*, *Wnt8-10*, *Wnt16*, and *WntA*. This likely reflects early losses in the arachnopulmonate ancestor following WGD. This suggests that the interaction of *Wnt* genes is dose-dependent and may be disturbed by the presence and transcription of a new duplicate (Janssen *et al.*, 2021).

I discovered a duplicate *Wnt1/wg* in scorpions, amblypygids, and a mygalomorph spider: duplicates of this *Wnt* gene have not yet been detected in any other metazoans, even following multiple rounds of WGD in vertebrates, teleosts (see <https://web.stanford.edu/group/nusselab/cgi-bin/wnt/vertebrate>), or horseshoe crabs. This suggests duplicates of *Wnt1/wg* have been lost in the lineage leading to araneomorphs (true spiders). In spiders, *Wnt1/wg* has lost the segment-polarity-like expression pattern which is conserved in the harvestman and other arthropods (Janssen *et al.*, 2021). In the mygalomorph *A. geniculata*, *Wnt1.2* is exclusively expressed in the SAZ (Janssen *et al.*, 2021). The other paralog is expressed in a similar pattern to the single *Wnt1* genes in true spiders but is lacking expression in the SAZ (Janssen *et al.*, 2021), an example of sub-functionalization after WGD.

The presence of two *Wnt4* ohnologs in amblypygids, a mygalomorph spider, and some lineages of entelegyne spider, contrasts with the retention of only one *Wnt4* in *P. tepidariorum* (Janssen *et al.*, 2010) and *P. phalangioides*. This suggests that the loss of a second *Wnt4* gene occurred independently in at least two lineages of spiders. Loss of a second *Wnt4* gene in *P. phalangioides* may be representative of Haplogynae as Janssen *et al.* (2021) did not identify a second copy in the genome of another basally branching haplogyne spider *Loxosceles reclusa*. The expression patterns of *Wnt4* genes in *P. tepidariorum*, *P. phalangioides*, and *P. opilio* are too diverse to identify possible patterns of sub- or neo-functionalisation in *A. geniculata* (Janssen *et al.*, 2021). Expression patterns of other arachnopulmonates could help to clarify the ancestral expression of *Wnt4*.

Wnt6 ohnologs were likely retained in scorpions and amblypygids, suggesting a *Wnt6* duplicate was lost in the lineage leading to spiders. Janssen et al. (2021) suggested *Wnt6* may have had an ancestral role in posterior elongation, like its proposed ancient paralog *Wnt1*. *Wnt6* could be a good candidate for sub- or neo-functionalisation.

I found *Wnt7* and *Wnt11* ohnologs were retained across arachnopulmonates however, I recovered a single copy of *Wnt11* from a mygalomorph spider and *P. phalangioides* transcriptome. This is consistent with synteny analysis in *P. tepidariorum*, which found *Wnt7* and *Wnt11* paralogs on different scaffolds, suggesting they are ohnologs (Schwager et al., 2017). Attempts to characterise the expression patterns of *Wnt11* paralogs in *P. tepidariorum* and *P. phalangioides*, only detected the expression of *Wnt11.2* (Janssen et al., 2021). *Wnt11* could be a good candidate for sub- or neo-functionalisation, but the role of *Wnt11.1* remains unknown.

The presence of *Wnt10* in both amblypygids and *P. opilio* is also intriguing because it is absent from all other chelicerates surveyed so far. In *P. opilio* *Wnt9* and *Wnt10* are expressed in a segment-polarity gene-like pattern and along the ventral side of the appendages, respectively (Janssen et al., 2021). Function-shuffling occurs regularly in Wnt genes and is often, but not necessarily, accompanied by gene loss and could explain the loss of *Wnt9* and *Wnt10* in spiders (Janssen et al., 2021) as well as the loss of *Wnt8* and *Wnt10* in scorpions.

5.6.3 *frizzled* repertoires vary substantially

I identified the presence of two *fz1* and *fz3* ohnologs in both amblypygids. This suggests that the loss of a second *fz1* and *fz3* gene occurred independently in spiders and scorpions. In *P. tepidariorum*, *fz1* is expressed in the ventral neuroectoderm (Janssen et al., 2015). *fz1* could be a good candidate for sub- or neo-functionalisation however, expression data is missing in other arachnopulmonates. The presence of *fz3* in *A. geniculata* and *P. phalangioides* suggests *fz3* was lost again in the more derived entelegyne lineages. Janssen et al. (2015) suggested *fz3* has an ancestral role in the nervous system, eye, and appendage development. Studying the expression of spider and other chelicerates *fz3* orthologs is required to address this more fully.

I found *fz4* ohnologs were retained across arachnopulmonates however, I recovered a single copy of *fz4* from *S. dumicola* and the RTA clade spiders. This is consistent with synteny analysis in *P. tepidariorum*, which found *fz4* paralogs on different scaffolds suggesting they are ohnologs (Schwager et al., 2017). In *P. tepidariorum* the *fz4* orthologs have been subject to sub-functionalisation (Janssen et al., 2015).

The expanded *frizzled* repertoire in amblypygids is intriguing since they also have the largest Wnt repertoires, via both duplication and representation of the subfamilies. While

frizzled genes encode key receptors for Wnt ligands, they also have other Wnt-independent functions, so the expanded *frizzled* repertoire could be equally related to the evolution of alternative signaling roles.

5.6.4 Sensory/neurosecretory TFs expanded in scorpions

Across the arachnoplumonates surveyed, *nkx2.1/scro* has been retained as a single copy. This likely reflects an early loss of the *nkx2.1/scro* duplicate in the arachnoplumonte ancestor following WGD.

rax and *otp*, along with *hbn*, belong to the HRO cluster named after the gene order in *D. melanogaster* (Walldorf *et al.*, 2000). This order is conserved in other insects, *C. sculpturatus*, and the tick genome, indicating that it is likely the ancestral arthropod arrangement (Aase-Remedios *et al.*, 2023). Across spiders, the HRO cluster is found intact but reordered to *rax*, *hbn*, then *otp*, suggesting this rearrangement occurred once ancestrally (Aase-Remedios *et al.*, 2023).

Loss of *rax* duplicates are restricted to transcriptomes of individual species. I studied the expression of *P. tepidariorum raxA* however, could not detect *raxB* expression at embryonic stages 7-12 (see Chapter Four). *rax* could be a good candidate for sub- or neo-functionalisation, but the role of *raxB* remains unknown (see Chapter Four). Studying the expression of *rax* ohnologs in other arachnids with and without a WGD will help elucidate its role in role in chelicerate brain development.

Both scorpions retained *otp* and *vsx/chx* ohnologs, suggestive of a loss of an ohnolog in a spider and amblypygid ancestor. I studied the expression of the single *otp* ortholog in *P. tepidariorum* however, could not detect *otp* expression at embryonic stages 7-12 (see Chapter Four). *otp* could be a good candidate for sub- or neo-functionalisation in the scorpion however, the role of *otp* in chelicerates is unknown. I found *vsx/chx* expression marks the developing neuroendocrine PI in *P. tepidariorum* (see Chapter Four), it would be interesting to see if scorpions have a PI and the impact of retaining a *vsx/chx* ohnolog on its development.

vax genes are also absent from insect genomes (Tessmar-Raible, 2007) and I found they are also missing in the crustacean. I recovered a single *vax* orthologs in the mygalomorph spider, both amblypygids, and the horseshoe crab. This suggests an early loss of the *vax* duplicate in the arachnoplumonate ancestor following WGD, followed by additional losses in the lineage leading to scorpions and araneomorph spiders. Vertebrate *vax* is exclusively expressed in the medial forebrain and its developmental derivatives the hypothalamus and preoptic area (Take-uchi *et al.*, 2003). The centipede *S. maritima* has two *vax* orthologs, *vax1* is expressed in some CPNCs and laterally in the main head placode, but not in the PI (Hunnekuhl and Akam, 2014). Furthermore, in *P. dumerilii* *vax* is expressed in and around the apical organ from the trochophore larval stage onwards (Tessmar-Raible, 2004; Tessmar-

Raible, 2007). The expression of chelicerate *vax* orthologs has not been characterised and would help elucidate its role in neuroendocrine development in arthropods.

foxQ2 was not included in the survey, despite being one of the highly conserved anterior genes, because it was lost in the lineage leading to amphibians and mammals. *foxQ2* is retained as a single copy in *P. tepidariorum*, future work could involve retrieving data on its duplication status in mygalomorph spiders and other arachnoplumonates.

Building on the candidate approach use here, in the future an unbiased approaches should be used to systematically identify ohnologs and compare commonalities and differences between the arachnoplumonate and vertebrate WGDs (Singh *et al.*, 2015).

6.1 Anterior neuroectoderm patterning in *P. tepidariorum* and mapping of the vertebrate mid-hindbrain boundary homolog

non-segmental pre-chelicer			SHB	Ch	Pp	L1	
S7							<i>six3.1</i> (C14)
S8							<i>six3.2</i> (C14)
	S5						<i>otd1</i> (C16)
	S8						<i>otd2</i> (C13)
							<i>gbx*</i>
S6							<i>tll</i> (C13 & C14)
S7							<i>foxQ2</i>
	S7						<i>Pax6.1</i> (C13)
	S6						<i>Pax6.2</i> (C13 & C16)
				S7			<i>lim1a</i> (C16)
					S4		<i>labA</i> (C15 & C18)
					S6		<i>labB</i> (C15 & C18)
							Other Hox
	S5		1	2	3		<i>hh1</i> (C16)
				3	2		<i>en1</i>
						S8	<i>Wnt8/Wnt16</i> (C13)
							<i>vnd*</i>
							<i>msx1</i> (C16)
pc				dc	tc		
PI	AB	MB					
PL	PE?	SE					

106

Expression in the spider neuroectoderm is based on *Parasteatoda tepidariorum* data. Where the data diverge, the harvestman *Phalangium opilio* pattern (*) is indicated by thin lines (Leite *et al.*, 2018). For each gene the initial and strongest expression domain (darkest colour) and further late expansions (lightest colour) are shown. Colours correspond to Figure 2 in Posnien *et al.* (2023). For each gene, I indicated at what position (vertical red broken lines) the SHB would be located when using only that gene as marker. Previously published expression data for Hox genes (Schwager *et al.*, 2017), *otd1* and *hh1* (Pechmann *et al.*, 2009), *en1* (Schwager *et al.*, 2015), Wnt genes (Janssen *et al.*, 2022) and *msx1* (Leite *et al.*, 2018). Data summarised from clusters 13, 14, and 16 from Leite *et al.* (2024) and unpublished data presented in this thesis. *P. tepidariorum* does not have an *Sp5/btd* ortholog (Setton and Sharma, 2018). Abbreviations: AB: arcuate body; Ch: chelicerel segment; dc: deutocerebrum; MB: mushroom bodies; pc: protocerebrum; PE: principal eyes; Pp: pedipalpal segment; SE: secondary eyes; SHB: spider head boundary; tc: tritocerebrum; L1: first walking leg segment.

Expression data in the spider corroborates the vertebrate MHB homolog being placed at the arthropod pre-antennal/chelicerel and antennal/chelicerel boundary (Posnien *et al.*, 2023). In vertebrates, the MHB emerges in the anterior intervening zone (IZ), devoid of Hox gene and *otx* expression, at the developmental boundary defined by adjacent non-overlapping expression of *otx* and *gastrulation brain homeobox (gbx)* (Rhinn *et al.*, 2005). *Pax2/5/8* genes mark the MHB and the first rhombomere of the hindbrain (Kiecker and Lumsden, 2005).

The antennal/chelicerel segment corresponds to the vertebrate IZ as it lacks Hox gene expression and initially *otd* expression (Posnien *et al.*, 2023) (Figure 41). *P. tepidariorum* has two *gbx* paralogs, which are ohnologs retained from the arachnoplumonate WGD event (Aase-Remedios *et al.*, 2023), *gbx2* is expressed in the prosomal appendages while *gbx1* could not be detected during embryogenesis by ISH (Leite *et al.*, 2018). The single *gbx* ortholog in the harvestman *P. opilio* is expressed up to the pre-chelicerel/chelicerel boundary, and likely represents the ancestral expression pattern (Leite *et al.*, 2018) (Figure 41). I show *P. tepidariorum Pax2.1* is expressed before *Pax2.2*, in segmental domains up to the pre-chelicerel/chelicerel boundary (Supplementary Figure 43), there may be an overlap, but this would be easily resolved by double staining with *hh1*. *Pax2.2* is expressed in the region of the SHB (Supplementary Figure 43). It would be interesting to characterise the expression of *Pax2* in an arachnid without a WGD, such as the harvestman, to better understand the ancestral expression pattern.

The expression of insect and arachnid columnar genes, those expressed along the body axis and involved in animal neural patterning, such as *nkx2.2/vnd* and *msx/msh*, reflect vertebrate findings. Vertebrate *Nkx2.2* is expressed along the body axis but in the MHB region this stripe is interrupted. In *D. melanogaster*, *vnd* shows ventral expression domains in all segments except for the antennal segment. *P. tepidariorum* has two *nkx2.2/vnd* paralogs, which are ohnologs retained from the arachnoplumonate WGD event (Aase-Remedios *et al.*,

2023), *vnd1* was expressed along the ventral neuroectoderm and adjacent to the PI/PL primordia, while *vnd2* could not be detected during embryogenesis by ISH (Leite *et al.*, 2018). In contrast, the single harvestman *vnd* ortholog is expressed in ventral all segments except in the pre-chelicerai/chelicerai boundary region (Figure 41) and likely represents the ancestral expression pattern (Leite *et al.*, 2018). Vertebrate *msx* has its anterior expression boundary at the MHB, similarly, the *D. melanogaster muscle segment homeobox (msh)* ortholog is expressed up to the pre-antennal/antennal boundary (Posnien *et al.*, 2023). *P. tepidariorum* has three *msx* paralogs, *msx1* is expressed up to the pre-chelicerai/chelicerai boundary and likely represents the ancestral expression pattern (Leite *et al.*, 2018) (Figure 41).

In vertebrates, *FGF8* expression marks the region posterior to the MHB. No expression data is available for *P. tepidariorum fgf8* after the germ disc stage nor is *fgf8* expression data available for any other chelicerate.

6.2 Further analysis of SCS cell types in spiders

Recently Medina-Jiménez *et al.* (2024a; 2024b) performed live-cell dissociation and droplet-based scRNA-seq on pooled stage 10-12 embryos of *P. tepidariorum* with the processed dataset containing a low cell number. Furthermore, live-cell dissociation can induce artefacts and stress-related transcriptional signatures. Droplet-based scRNA-seq methods (e.g., Drop-seq, inDrop and Chromium 10X) are sensitive to the introduction of noise from ambient RNA and cellular debris, which can falsely increase unique molecular identifier and gene counts per cell. In *P. tepidariorum*, we know there is molecular distinction between the principal eyes and each secondary eye pair (Schomburg *et al.*, 2015). Medina-Jiménez *et al.* (2024a; 2024b) did not identify cell clusters corresponding to the developing principal eyes nor individual clusters corresponding to each secondary eye pair, this is likely due to the pooling of different staged embryos.

To gain further insights into embryonic head, brain, and eye development in spiders I suggest carrying out scRNA-seq on several more embryonic stages (9.2-14) of *P. tepidariorum*, using ACME dissociation and SPLiT-seq (García-Castro *et al.*, 2021). ACME dissociation overcomes the limitations of live cell dissociations by simultaneously fixing and dissociating cells, preserving their RNA (García-Castro *et al.*, 2021). SPLiT-seq allows for sample multiplexing (Rosenberg *et al.*, 2018). Ambient RNA is eliminated with the supernatant in each of the successive centrifugation steps for each of the split-pool rounds. A FACS step after the third round of barcoding removes cellular debris and sorts cells into sub-libraries. These stages encompass the onset of neurogenesis (stage 9.2) when INPs become visible over the pre-chelicerai lobes and the formation of the anterior furrow (stage 10.2), the neural primordium of the arcuate body (Mittmann and Wolff, 2012). In addition to the formation of the

PEP and SEP primordia (stage 9.2/10.1), spitting of the SEP (stage 12), and when the principal and secondary eye pairs reach their final position (stage 14) (Schomburg *et al.*, 2015). Once identified, the evolutionary origin of new candidate genes involved in the development of morphological structures can be tested by means of comparative studies, such as *hh2* in spiders (Medina-Jiménez *et al.*, 2024b).

6.3 Further analysis of the PI and PL in spiders and other arachnids

In *P. tepidariorum*, later embryonic expression of *otp*, *raxB*, and *PC2* is missing. In addition to *PC1/3*, which like *PC2*, is specifically restricted to endocrine and neuroendocrine cells has not been studied in this spider or any other chelicerate. Furthermore, the expression of the pan-placodal *six4* and adenohypophyseal placode marker *lhx3/lim3* have not been studied in this spider or any other chelicerate.

In vertebrates, another TF expressed in the medial forebrain, and its developmental derivatives, the hypothalamus and preoptic area, as well as several other forebrain regions is *single-minded (sim)* (Tessmar-Raible, 2007). In the annelid *P. dumerilii*, *sim* is expressed in the apical organ of metatrochophore larva (Tessmar-Raible, 2004). The *D. melanogaster* midline determinant *Sim* is expressed faintly from stage 13 onward in a subset of cells in the PI placode (de Velasco *et al.*, 2007). *P. tepidariorum* and the centipede *S. maritima* *sim* orthologs have been studied concerning the evolution of the ventral midline in arthropods (Linne *et al.*, 2012). In *P. tepidariorum*, *sim* was expressed in the developing protocerebrum at later embryonic stages however, it is unclear whether this expression overlaps with the PI primordia (Linne *et al.*, 2012). This would be easily resolved by co-expression of *sim* and *vsx/chx* in *P. tepidariorum*.

In *P. dumerilii* *vtn* is expressed in a pair of cells in the apical organ of metatrochophore larva, each sends an axon that projects to the neurosecretory fibre plexus (Tessmar-Raible *et al.*, 2007). Similarly in *S. maritima*, *inotocin-like* is expressed in some cells of CPNCs and PI (Hunnekuhl and Akam, 2014). The inotocin signalling system has been lost in Diptera and araneomorph spiders. It has also been lost several other times during insect evolution, such as in Hymenoptera and Lepidoptera (Liutkeviciute *et al.*, 2016). While in chelicerates araneomorph spiders seem to be the exception with the inotocin signalling system recovered in horseshoe crabs, mites, ticks, and scorpions (Liutkeviciute *et al.*, 2016). It would be interesting to see if other arachnopulmonates have retained the inotocin signalling system, including mygalomorph spiders and amblypygids, which I have shown have both retained key developmental genes lost in araneomorph spiders. Furthermore, to characterise the embryonic expression of insect *inotocin* and chelicerate *inotocin-like* orthologs. Studies on inotocin signalling system insects suggest a possible role in diuresis in *T. castaneum* (Aikins

et al., 2008) and locomotor activity in ants (Liutkeviciute *et al.*, 2018). I expect the beetle *inotocin* ortholog to be expressed in the developing PI, similar to the centipede *inotocin-like* ortholog (Hunnekuhl and Akam, 2014).

Future work would involve determining if the PI and PL have been retained in other chelicerates and the possible role WGD has played in the development of neuroendocrine centres. I suggest studying the relative embryonic expression of *six3*, *vsx/chx*, and *fas2*, as well as other genes involved in the development of insect and centipede neuroendocrine centres in mygalomorph spiders, arachnopulmonates, and the harvestman.

7 Bibliography

- Aase-Remedios, M. E., Janssen, R., Leite, D. J., *et al.* (2023). Evolution of the Spider Homeobox Gene Repertoire by Tandem and Whole Genome Duplication. *Mol Biol Evol*, 40(12).
- Aikins, M. J., Schooley, D. A., Begum, K., *et al.* (2008). Vasopressin-like peptide and its receptor function in an indirect diuretic signaling pathway in the red flour beetle. *Insect Biochem Mol Biol*, 38(7):740-8.
- Akiyama-Oda, Y., Akaiwa, T. & Oda, H. (2022). Reconstruction of the Global Polarity of an Early Spider Embryo by Single-Cell and Single-Nucleus Transcriptome Analysis. *Front Cell Dev Biol*, 10:933220.
- 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):1735-47.
- Akiyama-Oda, Y. & Oda, H. (2006). Axis specification in the spider embryo: dpp is required for radial-to-axial symmetry transformation and sog for ventral patterning. *Development*, 133(12):2347-57.
- Akiyama-Oda, Y. & Oda, H. (2016). Multi-color FISH facilitates analysis of cell-type diversification and developmental gene regulation in the Parasteatoda spider embryo. *Dev Growth Differ*, 58(2):215-24.
- Akiyama-Oda, Y. & Oda, H. (2020). Hedgehog signaling controls segmentation dynamics and diversity via *msx1* in a spider embryo. *Sci Adv*, 6(37).
- Atkins, M., Jiang, Y., Sansores-Garcia, L., *et al.* (2013). Dynamic rewiring of the Drosophila retinal determination network switches its function from selector to differentiation. *PLoS Genet*, 9(8):e1003731.
- Ballesteros, J. A., Setton, E. V. W., Santibanez-Lopez, C. E., *et al.* (2021). Phylogenomic Resolution of Sea Spider Diversification through Integration of Multiple Data Classes. *Mol Biol Evol*, 38(2):686-701.
- Baudouin Gonzalez, L., Schoenauer, A., Harper, A., *et al.* (2024). Development and patterning of a highly versatile visual system in spiders. *bioRxiv*:2023.12.22.572789.
- Baudouin-Gonzalez, L., Harper, A., McGregor, A. P., *et al.* (2022). Regulation of Eye Determination and Regionalization in the Spider Parasteatoda tepidarium. *Cells*, 11(4).
- Beermann, A., Prühs, R., Lutz, R., *et al.* (2011). A context-dependent combination of Wnt receptors controls axis elongation and leg development in a short germ insect. *Development*, 138(13):2793-805.
- Birkan, M., Schaeper, N. D. & Chipman, A. D. (2011). Early patterning and blastodermal fate map of the head in the milkweed bug *Oncopeltus fasciatus*. *Evol Dev*, 13(5):436-47.
- Boyan, G. S. & Reichert, H. (2011). Mechanisms for complexity in the brain: generating the insect central complex. *Trends Neurosci*, 34(5):247-57.
- Boyan, G. S. & Williams, J. L. (1997). Embryonic development of the pars intercerebralis/central complex of the grasshopper. *Dev Genes Evol*, 207(5):317-329.
- Boyan, G. S. & Williams, J. L. (2008). Evidence that the primary brain commissure is pioneered by neurons with a peripheral-like ontogeny in the grasshopper *Schistocerca gregaria*. *Arthropod Struct Dev*, 37(3):186-98.
- Browne, W. E., Patel, N. H. & Martindale, M. Q. Developmental Expression of *lab* and *ems* in the head and CNS of *Parhyale hawaiiensis*. Society for Developmental Biology 63rd Annual Meeting, University of Calgary, Alberta, Canada.
- Browne, W. E., Schmid, B. G., Wimmer, E. A., *et al.* (2006). Expression of *otd* orthologs in the amphipod crustacean, *Parhyale hawaiiensis*. *Dev Genes Evol*, 216(10):581-95.
- Bruce, A. E. E. & Shankland, M. (1998). Expression of the Head Gene *Lox22-Otxin* in the Leech *Helobdella* and the Origin of the Bilateral Body Plan. *Developmental Biology*, 201(1):101-112.
- Cañestro, C., Albalat, R., Irimia, M., *et al.* (2013). Impact of gene gains, losses and duplication modes on the origin and diversification of vertebrates. *Semin Cell Dev Biol*, 24(2):83-94.
- Cao, Z., Yu, Y., Wu, Y., *et al.* (2013). The genome of *Mesobuthus martensii* reveals a unique adaptation model of arthropods. *Nat Commun*, 4:2602.
- Chen, R., Hou, Y., Connell, M., *et al.* (2021). Homeodomain protein Six4 prevents the generation of supernumerary Drosophila type II neuroblasts and premature differentiation of intermediate neural progenitors. *PLoS Genet*, 17(2):e1009371.
- Chipman, A. D. & Edgecombe, G. D. (2019). Developing an integrated understanding of the evolution of arthropod segmentation using fossils and evo-devo. *Proc Biol Sci*, 286(1912):20191881.

- Chipman, A. D., Ferrier, D. E., Brena, C., *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.
- Cho, S. J., Vallès, Y., Giani, V. C., Jr., *et al.* (2010). Evolutionary dynamics of the wnt gene family: a lophotrochozoan perspective. *Mol Biol Evol*, 27(7):1645-58.
- Clark, E. & Akam, M. (2016). Odd-paired controls frequency doubling in *Drosophila* segmentation by altering the pair-rule gene regulatory network. *Elife*, 5.
- Clark, E., Peel, A. D. & Akam, M. (2019). Arthropod segmentation. *Development*, 146(18).
- Cohen, S. M. & Jürgens, G. (1990). Mediation of *Drosophila* head development by gap-like segmentation genes. *Nature*, 346(6283):482-5.
- Coulcher, J. F. & Telford, M. J. (2012). Cap'n'collar differentiates the mandible from the maxilla in the beetle *Tribolium castaneum*. *EvoDevo*, 3(1):25.
- Crozatier, M., Valle, D., Dubois, L., *et al.* (1999). Head versus trunk patterning in the *Drosophila* embryo; collier requirement for formation of the intercalary segment. *Development*, 126(19):4385-94.
- Dalton, D., Chadwick, R. & McGinnis, W. (1989). Expression and embryonic function of empty spiracles: a *Drosophila* homeo box gene with two patterning functions on the anterior-posterior axis of the embryo. *Genes Dev*, 3(12a):1940-56.
- Davis, R. J., Tavsanli, B. C., Dittrich, C., *et al.* (2003). *Drosophila* retinal homeobox (*drx*) is not required for establishment of the visual system, but is required for brain and clypeus development. *Dev Biol*, 259(2):272-87.
- Davis, R. L. (2005). Olfactory memory formation in *Drosophila*: from molecular to systems neuroscience. *Annu Rev Neurosci*, 28:275-302.
- De Velasco, B., Erclik, T., Shy, D., *et al.* (2007). Specification and development of the pars intercerebralis and pars lateralis, neuroendocrine command centers in the *Drosophila* brain. *Developmental Biology*, 302(1):309-323.
- Di, Z., Yu, Y., Wu, Y., *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.
- Doeffinger, C., Hartenstein, V. & Stollewerk, A. (2010). Compartmentalization of the precheliceral neuroectoderm in the spider *Cupiennius salei*: development of the arcuate body, optic ganglia, and mushroom body. *J Comp Neurol*, 518(13):2612-32.
- Economou, A. D. & Telford, M. J. (2009). Comparative gene expression in the heads of *Drosophila melanogaster* and *Tribolium castaneum* and the segmental affinity of the *Drosophila* hypopharyngeal lobes. *Evol Dev*, 11(1):88-96.
- Erwin, D. H. (2009). Early origin of the bilaterian developmental toolkit. *Philos Trans R Soc Lond B Biol Sci*, 364(1527):2253-61.
- Farnworth, M. S., Eckermann, K. N. & Bucher, G. (2020). Sequence heterochrony led to a gain of functionality in an immature stage of the central complex: A fly-beetle insight. *PLoS Biol*, 18(10):e3000881.
- Farzana, L. & Brown, S. J. (2008). Hedgehog signaling pathway function conserved in *Tribolium* segmentation. *Dev Genes Evol*, 218(3-4):181-92.
- Feitosa, N. M., Pechmann, M., Schwager, E. E., *et al.* (2017). Molecular control of gut formation in the spider *Parasteatoda tepidariorum*. *Genesis*, 55(5).
- Finkelstein, R. & Perrimon, N. (1990). The orthodenticle gene is regulated by bicoid and torso and specifies *Drosophila* head development. *Nature*, 346(6283):485-8.
- Friedrich, M. (2017). Ancient genetic redundancy of eyeless and twin of eyeless in the arthropod ocular segment. *Dev Biol*, 432(1):192-200.
- Friedrich, M. (2022). Coming into clear sight at last: Ancestral and derived events during chelicerate visual system development. *Bioessays*, 44(12):e2200163.
- Fritzsche, S. & Hunnekuhl, V. S. (2021). Cell-specific expression and individual function of prohormone convertase PC1/3 in *Tribolium* larval growth highlights major evolutionary changes between beetle and fly neuroendocrine systems. *EvoDevo*, 12(1):9.
- Gainett, G., Ballesteros, J. A., Kanzler, C. R., *et al.* (2020). Systemic paralogy and function of retinal determination network homologs in arachnids. *BMC Genomics*, 21(1):811.
- Gainett, G. & Sharma, P. P. (2020). Genomic resources and toolkits for developmental study of whip spiders (Amblypygi) provide insights into arachnid genome evolution and antenniform leg patterning. *EvoDevo*, 11:18.
- Gallitano-Mendel, A. & Finkelstein, R. (1998). Ectopic orthodenticle expression alters segment polarity gene expression but not head segment identity in the *Drosophila* embryo. *Dev Biol*, 199(1):125-37.

- García-Castro, H., Kenny, N. J., Iglesias, M., *et al.* (2021). ACME dissociation: a versatile cell fixation-dissociation method for single-cell transcriptomics. *Genome Biol*, 22(1):89.
- Garwood, R. J., Dunlop, J. A., Knecht, B. J., *et al.* (2017). The phylogeny of fossil whip spiders. *BMC Evol Biol*, 17(1):105.
- Glasauer, S. M. & Neuhauss, S. C. (2014). Whole-genome duplication in teleost fishes and its evolutionary consequences. *Mol Genet Genomics*, 289(6):1045-60.
- Grbic, M., Van Leeuwen, T., Clark, R. M., *et al.* (2011). The genome of *Tetranychus urticae* reveals herbivorous pest adaptations. *Nature*, 479(7374):487-92.
- Green, P., Hartenstein, A. Y. & Hartenstein, V. (1993). The embryonic development of the *Drosophila* visual system. *Cell Tissue Res*, 273(3):583-98.
- Harper, A., Baudouin Gonzalez, L., Schonauer, A., *et al.* (2021). Widespread retention of ohnologs in key developmental gene families following whole-genome duplication in arachnopolmonates. *G3 (Bethesda)*, 11(12).
- Hartenstein, V., Tepass, U. & Gruszynski-Defeo, E. (1994). Embryonic development of the stomatogastric nervous system in *Drosophila*. *J Comp Neurol*, 350(3):367-81.
- Hasebe, M. & Shiga, S. (2021). Oviposition-promoting pars intercerebralis neurons show period-dependent photoperiodic changes in their firing activity in the bean bug. *Proc Natl Acad Sci U S A*, 118(9).
- He, B., Buescher, M., Farnworth, M. S., *et al.* (2019). An ancestral apical brain region contributes to the central complex under the control of foxQ2 in the beetle *Tribolium*. *eLife*, 8:e49065.
- Heisenberg, M. (2003). Mushroom body memoir: from maps to models. *Nat Rev Neurosci*, 4(4):266-75.
- Hilbrant, M., Damen, W. G. & Mcgregor, A. P. (2012). Evolutionary crossroads in developmental biology: the spider *Parasteatoda tepidariorum*. *Development*, 139(15):2655-62.
- Hildebrandt, K., Bach, N., Kolb, D., *et al.* (2020). The homeodomain transcription factor *Orthopedia* is involved in development of the *Drosophila* hindgut. *Hereditas*, 157(1):46.
- Hildebrandt, K., Klöppel, C., Gogel, J., *et al.* (2022). *Orthopedia* expression during *Drosophila melanogaster* nervous system development and its regulation by microRNA-252. *Dev Biol*, 492:87-100.
- Holland, P. W., Garcia-Fernández, J., Williams, N. A., *et al.* (1994). Gene duplications and the origins of vertebrate development. *Dev Suppl*:125-33.
- Homann, H. (1971). Die augen der Araneae. *Zeitschrift für Morphologie der Tiere*, 69(3):201-272.
- Homberg, U. (2008). Evolution of the central complex in the arthropod brain with respect to the visual system. *Arthropod Struct Dev*, 37(5):347-62.
- Hueber, S. D. & Lohmann, I. (2008). Shaping segments: Hox gene function in the genomic age. *Bioessays*, 30(10):965-79.
- Hughes, C. L. & Kaufman, T. C. (2002). Hox genes and the evolution of the arthropod body plan. *Evol Dev*, 4(6):459-99.
- Hunnekuhl, V. S. & Akam, M. (2014). An anterior medial cell population with an apical-organ-like transcriptional profile that pioneers the central nervous system in the centipede *Strigamia maritima*. *Developmental Biology*, 396(1):136-149.
- Hunnekuhl, V. S. & Akam, M. (2017). Formation and subdivision of the head field in the centipede *Strigamia maritima*, as revealed by the expression of head gap gene orthologues and hedgehog dynamics. *EvoDevo*, 8(1):18.
- Jager, M., Murienne, J., Clabaut, C., *et al.* (2006). Homology of arthropod anterior appendages revealed by Hox gene expression in a sea spider. *Nature*, 441(7092):506-508.
- Janeschik, M., Schacht, M. I., Platten, F., *et al.* (2022). It takes Two: Discovery of Spider Pax2 Duplicates Indicates Prominent Role in Chelicerate Central Nervous System, Eye, as Well as External Sense Organ Precursor Formation and Diversification After Neo- and Subfunctionalization. *Frontiers in Ecology and Evolution*, 10.
- Janssen, R. (2012). Segment polarity gene expression in a myriapod reveals conserved and diverged aspects of early head patterning in arthropods. *Dev Genes Evol*, 222(5):299-309.
- Janssen, R. (2017). Comparative analysis of gene expression patterns in the arthropod labrum and the onychophoran frontal appendages, and its implications for the arthropod head problem. *EvoDevo*, 8(1):1.
- Janssen, R., Budd, G. E. & Damen, W. G. M. (2011). Gene expression suggests conserved mechanisms patterning the heads of insects and myriapods. *Developmental Biology*, 357(1):64-72.
- Janssen, R., Le Gouar, M., Pechmann, M., *et al.* (2010). Conservation, loss, and redeployment of Wnt ligands in protostomes: implications for understanding the evolution of segment formation. *BMC Evol Biol*, 10:374.

- Janssen, R., Pechmann, M. & Turetzek, N. (2021). A chelicerate Wnt gene expression atlas: novel insights into the complexity of arthropod Wnt-patterning. *Evodevo*, 12(1):12.
- Janssen, R., Schomburg, C., Prpic, N. M., *et al.* (2022). A comprehensive study of arthropod and onychophoran Fox gene expression patterns. *PLoS One*, 17(7):e0270790.
- Janssen, R., Schönauer, A., Weber, M., *et al.* (2015). The evolution and expression of panarthropod *frizzled* genes. *Frontiers in Ecology and Evolution*, 3.
- Jeon, H., O, J., Jin, S., *et al.* (2019). A Role for buttonhead in the Early Head and Trunk Development in the Beetle *Tribolium castaneum*. *Dev Reprod*, 23(1):63-72.
- Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., *et al.* (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods*, 14(6):587-589.
- Kanayama, M., Akiyama-Oda, Y., Nishimura, O., *et al.* (2011). Travelling and splitting of a wave of *hedgehog* expression involved in spider-head segmentation. *Nature Communications*, 2:500.
- Kenny, N. J., Chan, K. W., Nong, W., *et al.* (2016). Ancestral whole-genome duplication in the marine chelicerate horseshoe crabs. *Heredity (Edinb)*, 116(2):190-9.
- Kiecker, C. & Lumsden, A. (2005). Compartments and their boundaries in vertebrate brain development. *Nat Rev Neurosci*, 6(7):553-64.
- Kittelmann, S. 2012. *Formation of the Clypeolabral Region During Embryonic Head Development of the Red Flour Beetle Tribolium castaneum*. Doctor rerum naturalium, Georg-August-Universität.
- Kittelmann, S., Ulrich, J., Posnien, N., *et al.* (2013). Changes in anterior head patterning underlie the evolution of long germ embryogenesis. *Developmental Biology*, 374(1):174-184.
- Kitzmann, P., Weißkopf, M., Schacht, M. I., *et al.* (2017). A key role for foxQ2 in anterior head and central brain patterning in insects. *Development*, 144(16):2969-2981.
- Koniszewski, N. 2011. *Functional analysis of embryonic brain development in Tribolium castaneum*. Georg-August Universität Göttingen.
- Krehenwinkel, H., Pomerantz, A., Henderson, J. B., *et al.* (2019). Nanopore sequencing of long ribosomal DNA amplicons enables portable and simple biodiversity assessments with high phylogenetic resolution across broad taxonomic scale. *Gigascience*, 8(5).
- Kunz, T., Kraft, K. F., Technau, G. M., *et al.* (2012). Origin of *Drosophila* mushroom body neuroblasts and generation of divergent embryonic lineages. *Development*, 139(14):2510-22.
- Land, M. F. 1985. The Morphology and Optics of Spider Eyes. In: BARTH, F. G. (ed.) *Neurobiology of Arachnids*. Berlin, Heidelberg: Springer Berlin Heidelberg.
- Lee, H. H. & Frasch, M. (2004). Survey of forkhead domain encoding genes in the *Drosophila* genome: Classification and embryonic expression patterns. *Dev Dyn*, 229(2):357-66.
- Lee, J. J., Von Kessler, D. P., Parks, S., *et al.* (1992). Secretion and localized transcription suggest a role in positional signaling for products of the segmentation gene *hedgehog*. *Cell*, 71(1):33-50.
- Leite, D. J., Baudouin-Gonzalez, L., Iwasaki-Yokozawa, S., *et al.* (2018). Homeobox Gene Duplication and Divergence in Arachnids. *Mol Biol Evol*, 35(9):2240-2253.
- Leite, D. J., Ninova, M., Hilbrant, M., *et al.* (2016). Pervasive microRNA Duplication in Chelicerates: Insights from the Embryonic microRNA Repertoire of the Spider Parasteatoda tepidariorum. *Genome Biol Evol*, 8(7):2133-44.
- Leite, D. J., Schönauer, A., Blakeley, G., *et al.* (2024). An atlas of spider development at single-cell resolution provides new insights into arthropod embryogenesis. *Evodevo*, 15(1):5.
- Lev, O. & Chipman, A. D. (2021). Development of the Pre-gnathal Segments in the Milkweed Bug *Oncopeltus fasciatus* Suggests They Are Not Serial Homologs of Trunk Segments. *Front Cell Dev Biol*, 9:695135.
- Linne, V., Eriksson, B. J. & Stollewerk, A. (2012). Single-minded and the evolution of the ventral midline in arthropods. *Dev Biol*, 364(1):66-76.
- Liutkeviciute, Z., Gil-Mansilla, E., Eder, T., *et al.* (2018). Oxytocin-like signaling in ants influences metabolic gene expression and locomotor activity. *Faseb j*, 32(12):fj201800443.
- Liutkeviciute, Z., Koehbach, J., Eder, T., *et al.* (2016). Global map of oxytocin/vasopressin-like neuropeptide signalling in insects. *Sci Rep*, 6:39177.
- Long, S. M. (2021). Variations on a theme: Morphological variation in the secondary eye visual pathway across the order of Araneae. *Journal of Comparative Neurology*, 529(2):259-280.
- Lovejoy, D. A. 2005. *Neuroendocrinology: An Integrated Approach*, John Wiley and Sons.
- Luan, Q., Chen, Q. & Friedrich, M. (2014). The Pax6 genes *eyeless* and *twin of eyeless* are required for global patterning of the ocular segment in the *Tribolium* embryo. *Dev Biol*, 394(2):367-81.
- Mao, Y. & Freeman, M. (2009). Fasciclin 2, the *Drosophila* orthologue of neural cell-adhesion molecule, inhibits EGF receptor signalling. *Development*, 136(3):473-81.
- Marlow, H., Tosches, M. A., Tomer, R., *et al.* (2014). Larval body patterning and apical organs are conserved in animal evolution. *BMC Biol*, 12:7.

- Mazza, M. E., Pang, K., Reitzel, A. M., *et al.* (2010). A conserved cluster of three PRD-class homeobox genes (homeobrain, rx and orthopedia) in the Cnidaria and Protostomia. *Evodevo*, 1(1):3.
- Medina-Jiménez, B. I., Budd, G. E. & Janssen, R. (2024a). Single-cell RNA sequencing of mid-to-late stage spider embryos: new insights into spider development. *BMC Genomics*, 25(1):150.
- Medina-Jiménez, B. I., Budd, G. E., Pechmann, M., *et al.* (2024b). Single-cell sequencing suggests a conserved function of Hedgehog-signalling in spider eye development. *EvoDevo*, 15(1):11.
- Meyer, A. & Schartl, M. (1999). Gene and genome duplications in vertebrates: the one-to-four (-to-eight in fish) rule and the evolution of novel gene functions. *Curr Opin Cell Biol*, 11(6):699-704.
- 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.
- Miyawaki, K., Mito, T., Sarashina, I., *et al.* (2004). Involvement of Wingless/Armadillo signaling in the posterior sequential segmentation in the cricket, *Gryllus bimaculatus* (Orthoptera), as revealed by RNAi analysis. *Mechanisms of Development*, 121(2):119-130.
- Mohler, J., Mahaffey, J. W., Deutsch, E., *et al.* (1995). Control of *Drosophila* head segment identity by the bZIP homeotic gene cnc. *Development*, 121(1):237-47.
- Morehouse, N. I., Buschbeck, E. K., Zurek, D. B., *et al.* (2017). Molecular Evolution of Spider Vision: New Opportunities, Familiar Players. *Biol Bull*, 233(1):21-38.
- Moreno, N., Bachy, I., Rétaux, S., *et al.* (2003). Pallial origin of mitral cells in the olfactory bulbs of *Xenopus*. *Neuroreport*, 14(18):2355-8.
- Murat, S., Hopfen, C. & Mcgregor, A. P. (2010). The function and evolution of Wnt genes in arthropods. *Arthropod Struct Dev*, 39(6):446-52.
- Nässel, D. R. (2002). Neuropeptides in the nervous system of *Drosophila* and other insects: multiple roles as neuromodulators and neurohormones. *Prog Neurobiol*, 68(1):1-84.
- Nassif, C., Noveen, A. & Hartenstein, V. (1998). Embryonic development of the *Drosophila* brain. I. Pattern of pioneer tracts. *J Comp Neurol*, 402(1):10-31.
- Nguyen, L. T., Schmidt, H. A., Von Haeseler, A., *et al.* (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol*, 32(1):268-74.
- Nong, W., Qu, Z., Li, Y., *et al.* (2021). Horseshoe crab genomes reveal the evolution of genes and microRNAs after three rounds of whole genome duplication. *Commun Biol*, 4(1):83.
- Ntini, E. & Wimmer, E. A. (2011). Second order regulator Collier directly controls intercalary-specific segment polarity gene expression. *Developmental Biology*, 360(2):403-414.
- Oda, H. & Akiyama-Oda, Y. (2020). The common house spider *Parasteatoda tepidariorum*. *Evodevo*, 11:6.
- Ontano, A. Z., Gainett, G., Aharon, S., *et al.* (2021). Taxonomic Sampling and Rare Genomic Changes Overcome Long-Branch Attraction in the Phylogenetic Placement of Pseudoscorpions. *Mol Biol Evol*, 38(6):2446-2467.
- Papadopoulos, J. S. & Agarwala, R. (2007). COBALT: constraint-based alignment tool for multiple protein sequences. *Bioinformatics*, 23(9):1073-9.
- Pascual-Anaya, J., D'aniello, S., Kuratani, S., *et al.* (2013). Evolution of Hox gene clusters in deuterostomes. *BMC Dev Biol*, 13:26.
- Pearson, J. C., Lemons, D. & McGinnis, W. (2005). Modulating Hox gene functions during animal body patterning. *Nat Rev Genet*, 6(12):893-904.
- Pechmann, M., Khadje, S., Turetzek, N., *et al.* (2011). Novel function of Distal-less as a gap gene during spider segmentation. *PLoS Genet*, 7(10):e1002342.
- Pechmann, M., Mcgregor, A. P., Schwager, E. E., *et al.* (2009). Dynamic gene expression is required for anterior regionalization in a spider. *PNAS*, 106(5):1468-72.
- Pechmann, M., Schwager, E. E., Turetzek, N., *et al.* (2015). Regressive evolution of the arthropod tritocerebral segment linked to functional divergence of the Hox gene labial. *Proc Biol Sci*, 282(1814).
- Pignoni, F., Baldarelli, R. M., Steingrímsson, E., *et al.* (1990). The *Drosophila* gene *tailless* is expressed at the embryonic termini and is a member of the steroid receptor superfamily. *Cell*, 62(1):151-63.
- Posnien, N., Hunnekuhl, V. S. & Bucher, G. (2023). Gene expression mapping of the neuroectoderm across phyla - conservation and divergence of early brain anlagen between insects and vertebrates. *Elife*, 12.
- Posnien, N., Koniszewski, N. & Bucher, G. (2011a). Insect Tc-six4 marks a unit with similarity to vertebrate placodes. *Developmental Biology*, 350(1):208-216.

- Posnien, N., Koniszewski, N. D., Hein, H. J., *et al.* (2011b). Candidate gene screen in the red flour beetle *Tribolium* reveals *six3* as ancient regulator of anterior median head and central complex development. *PLoS Genet*, 7(12):e1002416.
- Posnien, N., Schinko, J. B., Kittelmann, S., *et al.* (2010). Genetics, development and composition of the insect head--a beetle's view. *Arthropod Struct Dev*, 39(6):399-410.
- Prpic, N.-M., Schoppmeier, M. & Damen, W. G. M. (2008). Whole-mount in situ hybridization of spider embryos. *Cold Spring Harb. Protoc.*, 3(10).
- Rhinn, M., Lun, K., Luz, M., *et al.* (2005). Positioning of the midbrain-hindbrain boundary organizer through global posteriorization of the neuroectoderm mediated by Wnt8 signaling. *Development*, 132(6):1261-72.
- Rogers, B. T. & Kaufman, T. C. (1997). Structure of the insect head in ontogeny and phylogeny: a view from *Drosophila*. *Int Rev Cytol*, 174:1-84.
- Rosenberg, A. B., Roco, C. M., Muscat, R. A., *et al.* (2018). Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. *Science*, 360(6385):176-182.
- Samadi, L., Schmid, A. & Eriksson, B. J. (2015). Differential expression of retinal determination genes in the principal and secondary eyes of *Cupiennius salei* Keyserling (1877). *Evodevo*, 6:16.
- Schacht, M. I., Schomburg, C. & Bucher, G. (2020). *six3* acts upstream of *foxQ2* in labrum and neural development in the spider *Parasteatoda tepidariorum*. *Dev Genes Evol*, 230(2):95-104.
- Schaeper, N. D., Pechmann, M., Damen, W. G. M., *et al.* (2010). Evolutionary plasticity of collier function in head development of diverse arthropods. *Developmental Biology*, 344(1):363-376.
- Schinko, J. B., Kreuzer, N., Offen, N., *et al.* (2008). Divergent functions of orthodenticle, empty spiracles and buttonhead in early head patterning of the beetle *Tribolium castaneum* (Coleoptera). *Developmental Biology*, 317(2):600-613.
- Schomburg, C., Turetzek, N. & Prpic, N. M. (2020). Candidate gene screen for potential interaction partners and regulatory targets of the Hox gene labial in the spider *Parasteatoda tepidariorum*. *Dev Genes Evol*, 230(2):105-120.
- Schomburg, C., Turetzek, N., Schacht, M. I., *et al.* (2015). Molecular characterization and embryonic origin of the eyes in the common house spider *Parasteatoda tepidariorum*. *EvoDevo*, 6(1):15.
- Schröder, R. (2003). The genes orthodenticle and hunchback substitute for bicoid in the beetle *Tribolium*. *Nature*, 422(6932):621-5.
- Schroder, R., Eckert, C., Wolff, C., *et al.* (2000). Conserved and divergent aspects of terminal patterning in the beetle *Tribolium castaneum*. *Proc Natl Acad Sci U S A*, 97(12):6591-6.
- Schwager, E., Schönauer, A., Leite, D., *et al.* 2015. Chelicerata.
- Schwager, E. E., Sharma, P. P., Clarke, T., *et al.* (2017). The house spider genome reveals an ancient whole-genome duplication during arachnid evolution. *BMC Biol*, 15(1):62.
- Seecoomar, M., Agarwal, S., Vani, K., *et al.* (2000). knot is required for the hypopharyngeal lobe and its derivatives in the *Drosophila* embryo. *Mech Dev*, 91(1-2):209-15.
- Setton, E. V. W. & Sharma, P. P. (2018). Cooption of an appendage-patterning gene cassette in the head segmentation of arachnids. *Proc Natl Acad Sci U S A*, 115(15):E3491-E3500.
- Sharma, P. P., Gupta, T., Schwager, E. E., *et al.* (2014). Subdivision of arthropod cap-n-collar expression domains is restricted to Mandibulata. *Evodevo*, 5(1):3.
- Sharma, P. P., Schwager, E. E., Extavour, C. G., *et al.* (2012). Hox gene expression in the harvestman *Phalangium opilio* reveals divergent patterning of the chelicerate opisthosoma. *Evol Dev*, 14(5):450-63.
- Shimeld, S. M. & Holland, P. W. (2000). Vertebrate innovations. *Proc Natl Acad Sci U S A*, 97(9):4449-52.
- Siegmund, T. & Korge, G. (2001). Innervation of the ring gland of *Drosophila melanogaster*. *J Comp Neurol*, 431(4):481-91.
- Sievers, F., Wilm, A., Dineen, D., *et al.* (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol*, 7:539.
- Simonnet, F., Célérier, M.-L. & Quéinnec, E. (2006). Orthodenticle and empty spiracles genes are expressed in a segmental pattern in chelicerates. *Development Genes and Evolution*, 216(7):467-480.
- Sinakevitch, I., Long, S. M. & Gronenberg, W. (2021). The central nervous system of whip spiders (Amblypygi): Large mushroom bodies receive olfactory and visual input. *Journal of Comparative Neurology*, 529(7):1642-1658.
- Singh, P. P., Arora, J. & Isambert, H. (2015). Identification of Ohnolog Genes Originating from Whole Genome Duplication in Early Vertebrates, Based on Synteny Comparison across Multiple Genomes. *PLoS Comput Biol*, 11(7):e1004394.

- Skoulakis, E. M. & Davis, R. L. (1996). Olfactory learning deficits in mutants for leonardo, a *Drosophila* gene encoding a 14-3-3 protein. *Neuron*, 17(5):931-44.
- Skoulakis, E. M., Kalderon, D. & Davis, R. L. (1993). Preferential expression in mushroom bodies of the catalytic subunit of protein kinase A and its role in learning and memory. *Neuron*, 11(2):197-208.
- Steinmetz, P. R., Urbach, R., Posnien, N., *et al.* (2010). Six3 demarcates the anterior-most developing brain region in bilaterian animals. *Evodevo*, 1(1):14.
- Stollewerk, A., Tautz, D. & Weller, M. (2003). Neurogenesis in the spider: new insights from comparative analysis of morphological processes and gene expression patterns. *Arthropod Structure & Development*, 32(1):5-16.
- Stollewerk, A., Weller, M. & Tautz, D. (2001). Neurogenesis in the spider *Cupiennius salei*. *Development*, 128(14):2673–2688.
- Strausfeld, N. J., Hansen, L., Li, Y., *et al.* (1998). Evolution, discovery, and interpretations of arthropod mushroom bodies. *Learn Mem*, 5(1-2):11-37.
- Strausfeld, N. J., Weltzien, P. & Barth, F. G. (1993). Two visual systems in one brain: Neuropils serving the principal eyes of the spider *Cupiennius salei*. *Journal of Comparative Neurology*, 328(1):63-75.
- Strausfeld, N. J., Wolff, G. H. & Sayre, M. E. (2020). Mushroom body evolution demonstrates homology and divergence across Pancrustacea. *Elife*, 9.
- Take-Uchi, M., Clarke, J. D. & Wilson, S. W. (2003). Hedgehog signalling maintains the optic stalk-retinal interface through the regulation of Vax gene activity. *Development*, 130(5):955-68.
- Tessmar-Raible, K. (2007). The evolution of neurosecretory centers in bilaterian forebrains: insights from protostomes. *Semin Cell Dev Biol*, 18(4):492-501.
- Tessmar-Raible, K., Raible, F., Christodoulou, F., *et al.* (2007). Conserved sensory-neurosecretory cell types in annelid and fish forebrain: insights into hypothalamus evolution. *Cell*, 129(7):1389-400.
- Tessmar-Raible, K. G. K. 2004. *The evolution of sensory and neurosecretory cell types in bilaterian brains*. Doctoral Thesis, Philipps-Universität Marburg. .
- Turetzek, N., Pechmann, M., Janssen, R., *et al.* (2024). Hox genes in spiders: Their significance for development and evolution. *Semin Cell Dev Biol*, 152-153:24-34.
- Turetzek, N., Pechmann, M., Schomburg, C., *et al.* (2015). Neofunctionalization of a Duplicate dachshund Gene Underlies the Evolution of a Novel Leg Segment in Arachnids. *Molecular Biology and Evolution*, 33(1):109-121.
- Urbach, R. & Technau, G. M. (2003a). Early steps in building the insect brain: neuroblast formation and segmental patterning in the developing brain of different insect species. *Arthropod Struct Dev*, 32(1):103-23.
- Urbach, R. & Technau, G. M. (2003b). Molecular markers for identified neuroblasts in the developing brain of *Drosophila*. *Development*, 130(16):3621-37.
- Utting, M., Agricola, H., Sandeman, R., *et al.* (2000). Central complex in the brain of crayfish and its possible homology with that of insects. *J Comp Neurol*, 416(2):245-61.
- Walldorf, U. & Gehring, W. J. (1992). Empty spiracles, a gap gene containing a homeobox involved in *Drosophila* head development. *Embo j*, 11(6):2247-59.
- Walldorf, U., Kiewe, A., Wickert, M., *et al.* (2000). Homeobrain, a novel paired-like homeobox gene is expressed in the *Drosophila* brain. *Mech Dev*, 96(1):141-4.
- Walsh, K. T. & Doe, C. Q. (2017). *Drosophila* embryonic type II neuroblasts: origin, temporal patterning, and contribution to the adult central complex. *Development*, 144(24):4552-4562.
- Wang, W. & Lufkin, T. (2000). The murine Otp homeobox gene plays an essential role in the specification of neuronal cell lineages in the developing hypothalamus. *Dev Biol*, 227(2):432-49.
- Wang, Z., Palmer, G. & Griffith, L. C. (1998). Regulation of *Drosophila* Ca²⁺/calmodulin-dependent protein kinase II by autophosphorylation analyzed by site-directed mutagenesis. *J Neurochem*, 71(1):378-87.
- Wimmer, E. A., Frommer, G., Purnell, B. A., *et al.* (1996). buttonhead and D-Sp1: a novel *Drosophila* gene pair. *Mech Dev*, 59(1):53-62.
- Wolff, C. & Hilbrant, M. (2011). The embryonic development of the central American wandering spider *Cupiennius salei*. *Front Zool*, 8(1):15.
- Wolff, G. H. & Strausfeld, N. J. (2015). Genealogical correspondence of mushroom bodies across invertebrate phyla. *Curr Biol*, 25(1):38-44.
- Wolff, G. H., Thoen, H. H., Marshall, J., *et al.* (2017). An insect-like mushroom body in a crustacean brain. *Elife*, 6.

- Yamazaki, K., Akiyama-Oda, Y. & Oda, H. (2005). Expression patterns of a twist-related gene in embryos of the spider *Achaearanea tepidariorum* reveal divergent aspects of mesoderm development in the fly and spider. *Zoolog Sci*, 22(2):177-85.
- Yoo, S., Nair, S., Kim, H. J., *et al.* (2020). Knock-in mutations of scarecrow, a *Drosophila* homolog of mammalian Nkx2.1, reveal a novel function required for development of the optic lobe in *Drosophila melanogaster*. *Dev Biol*, 461(2):145-159.
- Yu, D., Ren, Y., Uesaka, M., *et al.* (2024). Hagfish genome elucidates vertebrate whole-genome duplication events and their evolutionary consequences. *Nature Ecology & Evolution*, 8(3):519-535.
- Zaffran, S., Das, G. & Frasch, M. (2000). The NK-2 homeobox gene scarecrow (*scro*) is expressed in pharynx, ventral nerve cord and brain of *Drosophila* embryos. *Mechanisms of Development*, 94(1):237-241.

8 Supplementary Materials

8.1 Chapter Two Supplementary Materials

Supplementary Table 2. List of primer sequences.

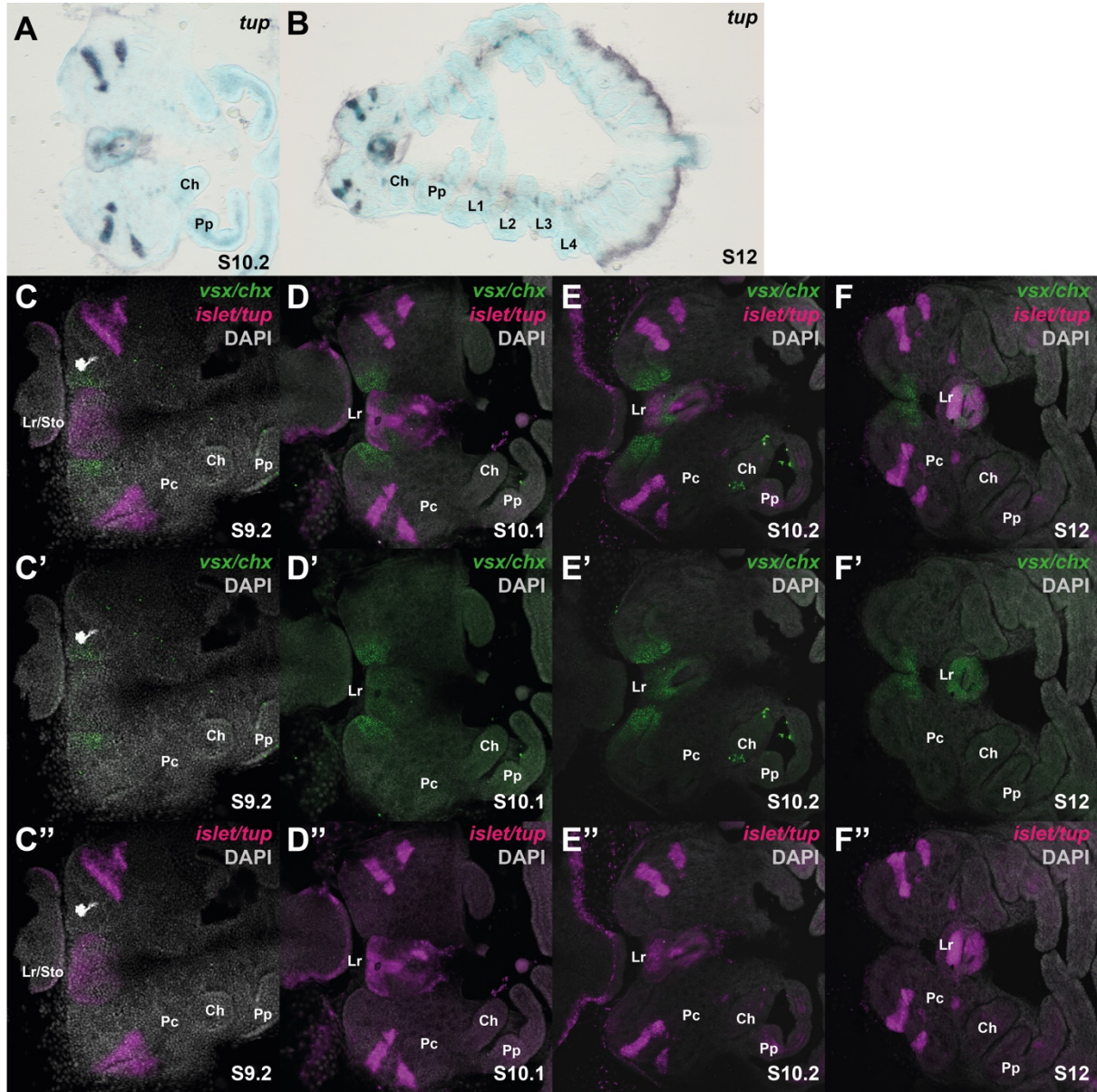
Gene	Gene ID	Forward primer (5'→3')	Reverse primer (5'→3')
<i>Pax6.2</i>	g12868	GGCCGCGGAGACAGGCTCGATCAAACCC	CCCGGGGCTGGCACTAGAATCACTACCGT
<i>six3.1</i>	g1245	GGCCGCGGCCCTACCTGTTGCTCATCCA	CCCGGGGCGCGTATGATGATGGTGCCTG
<i>six3.2</i>	g25543	GGCCGCGGACAGTCCCATGTTTCGTCCTT	CCCGGGGCGGGCTGCTACCGTGTAATG
<i>pax6.1</i>	g12873	GGCCGCGGGCTGCATCGTCTCTGTTTCAG	CCCGGGGCGGTGTTGATGCCTCCGTTAC
<i>omb3</i>	aug3.g3790	GGCCGCGGATGACTCCAGCGACGATGAT	CCCGGGGCGGGGTGTCGATGGAGTTGTA
<i>hbn</i>	g8762	GGCCGCGGAGGAGCGCTCTTCCCTTATCC	CCCGGGGCGACGCTTTGCCTGAAGACAA
<i>lim1a</i>	g12191	GGCCGCGGCAAGTTCCCTTCAGACGGT	CCCGGGGCGCGTTTCTTGTGCCAACTGT
<i>vsx/chx</i>	aug3.g27186	GGCCGCGGCGACCAGCAATCCAAACAGT	CCCGGGGCGTACCATTGCACCGTACAG
<i>tlf</i>	g18090	GGCCGCGGCATTCTTCAACCACTGCCCC	CCCGGGGCTTTTCTCCAGCCTCACCAGT
<i>raxA</i>	g8760	GGCCGCGGTGGATGTCGAAGCTCTCACT	CCCGGGGCTGTTGGAGGAGAGCAGGATC
<i>otp</i>	g8767	GGCCGCGGATTCCGGCCTAGTGAAGAGC	CCCGGGGCTGAAAGTGACTGCGCCAAAG
<i>nkx2.1/scro</i>	g16180	GGCCGCGGGTCCAGTTTGTAGTGCTTGCA	CCCGGGGCGCACATAGCGGCATTGGTAA
<i>so1</i>	g6476	GGCCGCGGTGTCCTGAAAGCAAAAGCCC	CCCGGGGCTGCGTGGGACTGAGTTTAGT
<i>pax2.1</i>	g1094	GGCCGCGGTGTTTTCGTGGGAAATCAGG	CCCGGGGCTGATGGGTGTGAGTGGAGT
<i>pax2.2</i>	g3432	GGCCGCGGCGAAACAAGTTGGTGCCTCT	CCCGGGGCTGAAAGCTTGTTCGGGTGTG
<i>tailup</i>	g20435	GGCCGCGGAGACAATTGCTACCGGGTGA	CCCGGGGCTATCTGCTGGAAATGTGGTGC
<i>hedgehog</i>	g23071	GGCCGCGGGTGCTGGCCGCATTAGTG	CCCGGGGCTGAGTCACCATCGAAACATC
<i>otd2</i>	g5047	GGCCGCGGGACAGTTCCACTCAGCTCCT	CCCGGGGCAATCGGTACTTCTGGTGGG
<i>IrreC-rst</i>	g992	GGCCGCGGGCGACTGTTGTTGGTTTTGC	CCCGGGGCTGCCCGGTCACTATTTTCCT
<i>PC2</i>	g24083	GGCCGCGGCTGCAAGCATGTGGACCATT	CCCGGGGCGCTGTCTGTTGAAGTTGGCC
<i>col1</i>	g10955	GGCCGCGGAGGCAATCAGTCCTAGCGAG	CCCGGGGCTGATCTTGGTGCAGGTAGGG
<i>fkf</i>	g3001	GGCCGCGGGTGAAAGTGGCTCGAACTCC	CCCGGGGCGACAGGTTGGGGCACAGAATA
<i>foxQ2</i>	g224	GGCCGCGGGCGCAAATGGTAAGGGACAT	CCCGGGGCTTGCATGGGTGTTGAAGAGC
<i>croc</i>	g24984	GGCCGCGGGTGGAGCCTATGGACGATGA	CCCGGGGCGAGATCCAGCACGAGTTCCTG
<i>fas2A</i>	g31369	GGCCGCGGAAGAACTGGACGCAGGAGAA	CCCGGGGCGCTCACCACCAACATTGCGA
<i>latrophilin</i>	g14495	GGCCGCGGAAATGGTATTCTGCCGTCGC	CCCGGGGCGGACGGTCGCAGCAAATATT
<i>integrin-a-8</i>	g23098	GGCCGCGGATCAACTGCCAACTGTGCTG	CCCGGGGCTGAGCCCATCTGATCTCCTG
<i>neuroactin</i>	g29023	GGCCGCGGTGCTAGACCTCTCCCTTCT	CCCGGGGCGCCGGAATGTAAACGACCAC
<i>Tbx2/3</i>	aug3.g3745	GGCCGCGGACCACCTGAAGACGGAAGAG	CCCGGGGCGTCTTTTGGCGACGGTGAAT

8.2 Chapter Three Supplementary Materials

Supplementary Table 3. Differentially expressed genes between *so1*-expressing and *so1*-negative cells.

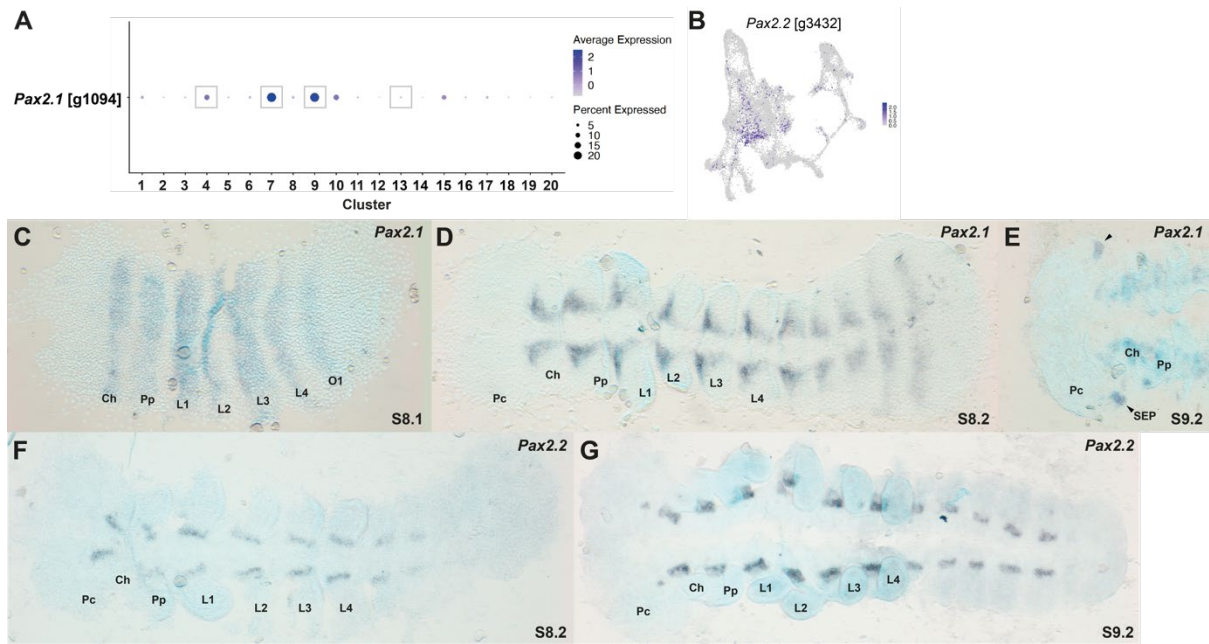
Gene	ID	p_val	avg_log2FC	pct.1	pct.2	p_val_adj
<i>so1</i>	g6476	0.00E+00	0.8526391	1	0	0.00E+00
<i>Awh</i>	g21739	1.13E-70	0.2604808	0.291	0.09	1.63E-66

IrreC-rst	g992	9.48E-54	0.2630529	0.338	0.136	1.36E-49
sog	g13327	3.57E-27	-0.2632098	0.085	0.266	5.13E-23



Supplementary Figure 42. Expression of *islet/tup* and *vsx/chx* in *Parasteatoda tepidariorum* embryos.

Embryos are flat mounted with the anterior to the left. Stages prior to the onset of expression are omitted. **(A-B)** Embryos stained for *islet/tup* transcripts at stages 10.2 **(A)** and 11 **(B)**. **(C-F'')** Embryos stained for *vsx* (green) and *islet/tup* (magenta) transcripts and DNA (grey) at stages 9.1 **(C)**, 10.1 **(D)**, 10.2 **(E)**, and 12 **(F)**. **Other abbreviations:** Ch: chelicerae; Lr: labrum; Pp: pedipalps; L1-4: walking legs 1-4.



Supplementary Figure 43. Expression of *Pax2.1* and *Pax2.2* in *Parasteatoda tepidariorum* embryos.

(A) Dotplot of *Pax2.1*. *Pax2.1* was a marker for several clusters (grey box): 4 [Leg], 7 [PNS], 9 [PNS]; and 13 [Pc posterior]. (B) UMAP of *Pax2.2*. (C-G) Embryos were flat mounted with the anterior to the left. Stages prior to the onset of expression were omitted. Embryos stained for *Pax2.1* transcripts at stages 8.1 (C), 8.2 (D), and 9.2 (E) and *Pax2.2* transcripts at 8.2 (F) and 9.2 (G). Other abbreviations: Ch: chelicerae; Pp: pedipalp; Pc: pre-chelicer; SEP: secondary eye primordia; L1-4: walking legs 1-4.

8.3 Chapter Five Supplementary Materials

Supplementary Table 4. *otp*, *nkx2.1/scro*, *rax*, *vsx/chx*, *vax* gene IDs and homeodomain sequence.

Species	Gene	ID	Homeodomain
<i>Limulus polyphemus</i>	<i>nkx2.1</i>	XP_013775139.1	RRKRRVLFTQAQVYELERRFKQQKYL SAPEREHLAGMIHLTPN QVKIWFQNHRYKCKRQA
	<i>nkx2.1</i>	XP_013789774.1	RRKRRVLFTQAQVYELERRFKQQKYL SAPEREHLANLIHLTPT QVKIWFQNHRYKCKRQA
	<i>nkx2.1</i>	XP_022243894.1	RRKRRVLFTQAQVYELERRFKQQKYL SAPEREHLASLIHLTPT QVKIWFQNHRYKCKRQA
	<i>nkx2.1</i>	XP_013794562.1	RRKRRVLFTQAQVYELERRFKQQKYL SAPEREHLASLIHLTPT QVKIWFQNHRYKCKRQA
	<i>vsx</i>	XP_022257594.1	KRRHRTIFTSYQLEELEKAFKEAHYPDVYAREMLSLKTDLPED RIQVWFQNRRAKWRKTE
	<i>vsx</i>	XP_022249270.1	KRRHRTIFTSYQLEELEKAFKEAHYPDVYAREMLSLKTDLPED RIQVWFQNRRAKWRKTE
	<i>vsx</i>	XP_022244652.1	KRRHRTIFTSSQLDELENAFKEAHYPDVYAREMLSLKTDLAED RIQVWFQNRRAKWRKTE

	rax	XP_013777739.1	HRRNRTTFTTYQLHELERAFAEKSHYPDVYSREELAMKVNLP RVQVWFQNNRAKWRQE
	rax	XP_013789393.1	HRRSRTTFTTYQLHELERAFAEKTHYPDVYSREELAMKVNLP RVQVWFQNNRAKWRQE
	otp	XP_013791515.2	QKRHRTRFTPTQLQELERSFSKTHYPDIFMREELAMRIGL RVQVWFQNNRAKWKRRK
	otp	XP_022245320.1	PKRHRTRFTPAQLQELERSFAKTHYPDIFMREELAMRIGL RVQVWFQNNRAKWKRRK
	vax	XP_022254995.1	PKRARTTFSSAQLQSLEEEFRNNQYLVGKERTKLAQRLGL QVKVWFQNNRTKHKREK
<i>Charinus acosta</i>	nkx2.1	TRINITY_DN2637 0_c1_g1_i2	RRKRRVLFTQAQVYELERRFKQKYL QVKIWFQNNHRYKCKRQA
	otp	TRINITY_DN7882 _c2_g1_i2	QKRHRTRFTPAQLQELERSFSKTHYPDIFMREELAMRIGL RVQVWFQNNRAKWKRRK
	rax-A	TRINITY_DN1705 9_c0_g1_i1	HRRNRTTFTTYQLHELERAFAEKSHYPDVYSREELAMKVNLP RVQVWFQNNRAKWRQE
	vsx	TRINITY_DN3746 4_c0_g1_i1	----- MLSLKTDLPEDRIQVWFQNNRAKWRKTE
	vax	TRINITY_DN1224 5_c0_g1_i2	PKRARTMFSGHQLARLEEA QVKVWFQNNRTKYKREK
<i>Euphrynichus bacillifer</i>	nkx2.1	TRINITY_DN662_ c0_g1_i1	RRKRRVLFTQAQVYELERRFKQKYL QVKIWFQNN-----
	rax-A	TRINITY_DN1823 _c0_g1_i1	HRRNRTTFTTYQLHELERAFAEKSHYPDVYSREELAMKVNLP RVQVWFQNNRAKWRQE
	vax	TRINITY_DN3291 9_c0_g1_i2	PKRARTMFSQYLACLED QVKVWFQNNRTKFKREK
	vsx?	TRINITY_DN7987 _c0_g1_i1	IRRSRTTFTTYQLHQLERA RVQVWFQNNRAKWRKRE
<i>Parasteatoda tepidariorum</i>	nkx2.1	g16180	RRKRRVLFTQAQVYELERRFKQKYL QVKIWFQNNHRYKCKRQA
	Rax-B	g23262	HRRNRTTFTTYQLHELERAFAERSHYPDVYSREELAVKVNLP RVQVWFQNNRAKWRQE
	Rax-A	g8760	HRRNRTTFTTYQLHELERAFAEKSHYPDVYSREELAMKVNLP RVQVWFQNNRAKWRQE
	vsx	g22189	KRRHRTIFTSYQLEELEKAFKDAHYPDVYAREMLSLKTDLP RIQVWFQNNRAKWRKTE
	Otp	g8767	QKRHRTRFTPAQLQELERSFSKTHYPDIFMREELAMRIGL RVQVWFQNNRAKWKRRK
<i>Pholcus phalangoides</i>	nkx2.1	c101836_g1_i1	RRKRRVLFTQAQVYELERRFKQKYL QVKIWFQNNHRYKCKRQA
	rax-A	c102000_g1_i1	HRRNRTTFTTYQLHELERAFAEKSHYPDVYSREELAMKVNLP RVQVWFQNNRAKWRQE
	otp	c103695_g1_i1	QKRHRTRFTPAQLQELERSFSKTHYPDIFMREELAMRIGL RVQVWFQNNRAKWKRRK

	vsx	c99447_g1_i3	KRRHRTIFTSYQLEELEKAFKDAHYPDVYAREMLSLKTDLPED RIQVWFQNRRRAKWRKTE
	rax	c110545_g4_i3	HRRNRTTFTTYQLHELERAFAERSHYPDVYSREELAIKVNLP RVQVWFQNRRRAKWRKTE
<i>Acanthoscurria geniculata</i>	rax-A	TRINITY_DN3407 9_c1_g1_i1	----- MKVNLPEVRVQVWFQNRRRAKWRKTE
	vsx?	TRINITY_DN9849 6_c0_g1_i1	----- MRLDLSEARVQVWFQNRRRAKWRKTE
	nkx2.1	TRINITY_DN4356 6_c1_g1_i1	RRKRRVLFTQAQVYELERRFKQKYLAPEREHLASLIHLTP QVKIWFQNHRYKCKRQA
	otp	TRINITY_DN7689 9_c8_g3_i1	QKRHRTRFTPAQLQELERSFSKTHYPDIFMREELAMRIGL TES RVQVWFQNRRRAKWRKTE
	vax	TRINITY_DN9229 8_c0_g1_i1	PKRTRTTFSEQQLSLENEFRQKQYLIGRERNRLALTLGL SET QVKVWFQNRRRAKWRKTE
<i>Marpissa muscosa</i>	nkx2.1	TRINITY_DN8697 _c0_g1_i1	RRKRRVLFTQAQVYELERRFKQKYLAPEREHLASLIHLTP QVKIWFQNHRYKCKRQA
	otp	TRINITY_DN6896 0_c0_g1_i2	QKRHRTRFTPAQLQELERSFSKTHYPDIFMREELAMRIGL TES RVQVWFQNRRRAKWRKTE
	rax-B	TRINITY_DN5295 2_c0_g1_i2	HRRNRTTFTTYQLHELERAFAERSHYPDVYSREELAVKVN LP RVQVWFQNRRRAKWRKTE
	rax-A	TRINITY_DN1406 6_c0_g1_i1	HRRNRTTFTTYQLHELERAFAEKSHYPDVYSREELAMKVN LP RVQVWFQNRRRAKWRKTE
	vsx	TRINITY_DN1007 6_c0_g1_i1	KRRHRTIFTSYQLEELEKAFKDAHYPDVYAREMLSLKTD LPED RIQVWFQNRRRAKWRKTE
<i>Pardosa amentata</i>	nkx2.1	TRINITY_DN1483 4_c0_g2_i1	RRKRRVLFTQAQVYELERRFKQKYLAPEREHLASLIHLTP QVKIWFQNHRYKCKRQA
	otp	TRINITY_DN1658 52_c0_g1_i1	QKRHRTRFTPAQLQELERSFSKTHYPDIFMREELAMRIGL TES RVQVWFQNRRRAKWRKTE
	rax-A	TRINITY_DN3059 6_c0_g1_i3	HRRNRTTFTTYQLHELERAFAEKSHYPDVYSREELAMKVN LP RVQVWFQNRRRAKWRKTE
	vsx	TRINITY_DN1390 7_c1_g1_i8	KRRHRTIFTSYQLEELEKAFKDAHYPDVYAREMLSLKTD LPED RIQVWFQNRRRAKWRKTE
<i>Mesobuthus martensii</i>	nkx2.1	MMA31928	RRKRRVLFTQAQVYELERRFKQKYLAPEREHLASLIHLTP QVKIWFQNHRYKCKRQA
	rax-B	MMA22329	HRRNRTTFTTYQLHELERAFAEKSHYPDVYCREELAIKVN LP RVQVWFQNRRRAKWRKTE
	rax	MMA31508	HRRNRTTFTTYQLHELERAFAEKSHYPDVYSREELAMKVN LP RVEVWFQNRRRAKWRKTE
	vsx	MMA29966	IDHLATIFTSYQLEELEKAFKEAHYPDVYAREMLSLKTD LPED RIQVWFQNRRRAKWRKTE
	vsx	MMA49053	WMEKLTIFTSYQLEELEKAFKEAHYPDVYAREMLSLKTD LPED RIQVWFQNRRRAKWRKTE
	otp-A	MMA23872_otp-A	QKRHRTRFTPAQLQELERSFAKTHYPDIFMREELAMRIGL TES RVQ-----

	otp	MMA43771_otp	QKRHRTRFTTPAQLQELERSFVKTHYPDIFMREELAMRIGLTES RVQNYVYG-----
<i>Stegodyphus dumicola</i>	nkx2.1	XP_035217516.1	RRKRRVLFTQAQVYELERRFKQQKYLAPEREHLASLIHLTPT QVKIWFQNHRYKCKRQA
	otp	XP_035228352.1	QKRHRTRFTTPAQLQELERSFSKTHYPDIFMREELAMRIGLTES RVQVWFQNRRAKWKRRK
	rax-A	XP_035208675.1	HRRNRTTFTTYQLHELERAFESHYPDVYSREELAMKVNLP RVQVWFQNRRAKWKRRQ
	rax-B	XP_035223039.1	HRRNRTTFTTYQLHELERAFESHYPDVYSREELAMKVNLP RVQVWFQNRRAKWKRRQ
	vsx	XP_035219442.1	KRRHRTIFTTSYQLEELEKAFKDAHYPDVYAREMLSLKTDLP RIQVWFQNRRAKWKRRQ
<i>Phoxichildium femoratum</i>	nkx2.1	TRINITY_DN14493_c0_g2_i1	RRKRRVLFTQAQVYELERRFKQQKYLAPEREHLANVIHLTPT QVKIWFQNHRYKCKRQA
	otp	TRINITY_DN16730_c1_g2_i2	QKRHRTRFTTPAQLNELERSFSKTHYPDIFMREELAMRIGLTES RVQVWFQNRRAKWKRRK
	rax	TRINITY_DN8076_c0_g1_i1	HRRSRTTFTTFQLHELERAFESHYPDVYSREELAMKVNLP RVQVWFQNRRAKWKRRQ
	vsx	TRINITY_DN16317_c3_g4_i3	KRRHRTIFTTSYQLEELEKAFKDAHYPDVYAREMLSLKTDLP RIQVWFQNRRAKWKRRQ
<i>Tetranychus urticae</i>	vsx	XP_025017939.1	KRRHRTIFTTSYQLEELEKAFKDAHYPDVYAREMLSVKTDLP RIQVWFQNRRAKWKRRQ
	otp	XP_025017915.1	QKRHRTRFTTPAQLQELERSFSKTHYPDIFMREELAMRIGLTES RVQVWFQNRRAKWKRRK
	nkx2.1	XP_015792838.1	RRKRRVLFTQAQVYELERRFKQQKYLAPEREHLASLIHLTPT QVKIWFQNHRYKCKRQA
	rax	XP_015789649.1	HRRNRTTFTTYQLHQLERAFESHYPDVYSREELAMKVNLP RVQVWFQNRRAKWKRRQ
	vax?	XP_015794010.1	PKRVRTAFSPSOLMELERAFDKNYVIGNERKELAKKLNLP QVKVWFQNRRTKAKRQK
<i>Daphnia pulex</i>	nkx2.1	XP_046442679.1	RRKRRVLFTQAQVYELERRFKQQKYLAPEREHLASLIHLTPT QVKIWFQNHRYKCKRQA
	otp	XP_046443834.1	QKRHRTRFTTPAQLNELERSFSKTHYPDIFMREELAMRIGLTES RVQVWFQNRRAKWKRRK
	rax	XP_046443190.1	HRRNRTTFTTFQLHELERAFESHYPDVYSREELAMKVNLP RVQVWFQNRRAKWKRRQ
	vsx	XP_046462950.1	KRRHRTIFTTSYQLEELEKAFKDAHYPDVYAREMLSLKTDLP RIQVWFQNRRAKWKRRQ
<i>Bombyx mori</i>	nkx2.1	XP_012545466.1	RRKRRVLFTQAQVYELERRFKQQKYLAPEREHLASLIHLTPT QVKIWFQNHRYKCKRQA
	otp	XP_037877149.1	QKRHRTRFTTPAQLNELERSFSKTHYPDIFMREELAMRIGLTES RVQVWFQNRRAKWKRRK
	rax	XP_004933582.2	HRRNRTTFTTYQLHELERAFESHYPDVYSREELAMKVNLP RVQVWFQNRRAKWKRRQ

	vsx	XP_062529738.1	RRHSRTIFTSYQLDELEKAFKDAHYPDVYAREMLSLKTDLPED RIQVWFQNRRAKWRKTE
<i>Euperipatoides kanangrensis</i>	vsx	SAP35452.1	KRRHRTIFAGYQLEEELEKAFKEAHYPDVYAREMLSLKTDLPED RIQVWFQNRRAKWRKTE
	rax	SAP35450.1	HRRNRRTFTTFFQLHELERAFEEKSHYPDVYSREELAVKVNLP RVQVWFQNRRAKWRKTE
	nkx2.1	SAP35451.1	RRKRRVLFTQAQVYELERRFKQKYLAPEREHLASLIHLTP QVKIWFQNHRYKCKRQA
<i>Strigamia maritima</i>	vax-2	SMAR005036	PKRSRTTFTSTEQLNHLEDEFQRNQYLAGRDRSQLADKLG QIKVWFQNRRTKHKREK
	vax-1	none	PKRSRTTFTSTEQLNHLEDEFQRNQYLAGRDRSQLADTLG QIKVWFQNRRTKHKREK
	otp	SMAR009836	QKRHRTRFTPAQLNELERSFSKTHYPDIFMREELAMRIGL RVQVWFQNRRAKWKRRK
	rax	SMAR009838	HRRNRRTFTTFFQLHELERAFEEKSHYPDVYSREELAMKVN RVQVWFQNRRAKWRKTE
	vsx	SMAR010943	KRRHRTIFTSYQLEEELEKAFKEAHYPDVYAREMLSLKTD RIQVWFQNRRAKWRKTE
	nkx2.1	SMAR011900	RRKRRVLFTQAQVYELERRFKQKYLAPEREHLASLIHLTP QVKIWFQNHRYKCKRQA
<i>Drosophila melanogaster</i>	vsx1	NP_001284898.1	RRHGRTIFTSSQLEEELEKAFKEAHYPDVSAARELLSMKT RIQVWFQNRRAKWRKTE
	vsx2	NP_001033832.1	RRHSRTIFTSYQLEKLEEFKEAHYPDVYAREMLSLKTELP RIQVWFQNRRAKWRKTE
	scro	NP_001015473.1	RRKRRVLFTQAQVYELERRFKQRYLSAPEREHLASLIHLTP QVKIWFQNHRYKCKRQA
	otp	NP_523799.3	QKRHRTRFTPAQLNELERCFSKTHYPDIFMREEIAMRIGL RVQVWFQNRRAKWKRRK
	rax	NP_726006.3	HRRNRRTFTTFFQLHELERAFEEKSHYPDVYSREELAMKVN RVQVWFQNRRAKWRKTE
<i>Tribolium castaneum</i>	scro	XP_973795.3	RRKRRVLFTQAQVYELERRFKQKYLAPEREHLASLIHLTP QVKIWFQNHRYKCKRQA
	otp	NP_001163995.1	QKRHRTRFTPAQLNELERCFSKTHYPDIFMREEIAMRIGL RVQVWFQNRRAKWKRRK
	rax	XP_973468.3	HRRNRRTFTTFFQLHELERAFEEKSHYPDVYSREELAMKVN RVQVWFQNRRAKWRKTE
	vsx	CAX64460.1	KRRHRTIFTSYQLEEELEKAFKDAHYPDVYAREMLSLKTD RIQVWFQNRRAKWRKTE
<i>Ixodes scapularis</i>	otp	EEC19355	QKRHRTRFTPAQLNELERSFSKTHYPDIFMREELAMRIGL RVQVWFQNRRAKWKRRK
	nkx2.1	XP_042144674.1	RRKRRVLFTQAQVYELERRFKQKYLAPEREHLASLIHLTP QVKIWFQNHRYKCKRQA
	vsx	XP_040065896.1	KKKRTIFTTSFQLEEELEKAFKEAHYPDVYAREMLSLKTD RIQVWFQNRRAKWRKTE

	Rax	XP_029829045.2	HRRNRTTFTTFFQLHELERAFEKSHYPDVYSREELAMKVNLP RVQVWFQNRRAKWRQ
<i>Phalangium opilio</i>	vsx	comp35928_c0_s eq2	KKKRRTIFTTSYQLEEELEKAFKEAHYPDVYAREMLSLKTDLP RIQVWFQNRRAKWRKTE
	rax	comp173922_c0_ seq1	RRNRTTFTTYYQLHELERAFEKSHYPDVYSREELAMKVNLP VQVWFQNRRAKWRQEK
	nkx2.1	comp171484_c0_ seq1	RKRRLVFTQAQVYELERRFKQKYLAPEREHLASLIHLTP VKIWFQNHRYKCKRQA
	Vax	none	PKRSRTSFNQWQLDALEREFLRCHGYVVGKDRNLSSKSLGLTE TQVKVWFQNRRTKFKKD
	otp	none	QKRHRTRFTPAQLQELERSFSKTHYPDIFMREELAMRIGL RVQ-----
<i>Centruroides sculpturatus</i>	rax-B	CSCU022535	HRRNRTTFTTYYQLHELERAFEKSHYPDVYCREELAIKVNLP RVQVWFQNRRAKWRQ
	rax-A	XP_023232524.1	HRRNRTTFTTYYQLHELERAFEKSHYPDVYSREELAMKVNLP RVQVWFQNRRAKWRQ
	vsx2	XP_023227221.1	KRRHRTIFTTSYQLEEELEKAFKEAHYPDVYAREMLSLKTDLP RIQVWFQNRRAKWRKTE
	vsx1	CSCU022240	YTLFITIFTTSYQLEEELEKAFKEAHYPDVYAREMLSLKTDLP RIQVWFQNRRAKWRKTE
	otp-B	XP_023237088.1	QKRHRTRFTPTQLQELERSFVKTHYPDIFMREELAVRIGL RVQVWFQNRRAKWKRRK
	otp-A	XP_023232517.1	QKRHRTRFTPAQLQELERSFAKTHYPDIFMREELAMRIGL RVQVWFQNRRAKWKRRK
	nkx2.1	XP_023232030.1	RRKRRVLFTQAQVYELERRFKQKYLAPEREHLASLIHLTP QVKIWFQNHRYKCKRQA
<i>Pardosa pseudoannulata</i>	nkx2.1	CM063505.1:1082 90984-108291163	RRKRRVLFTQAQVYELERRFKQKYLAPEREHLASLIHLTP QVKIWFQNHRYKCKRQA
	otp	CM063504.1:9783 1354-97831494	QKRHRTRFTPAQLQELERSFSKTHYPDIFMREELAMRIGL RVQV-----
	rax-A	CM063504.1:9823 8249-98238389	HRRNRTTFTTYYQLHELERAFEKSHYPDVYSREELAMKVNLP RVQV-----
	rax-B	CM063506.1:8668 7706 to 86687858	HRRNRTTFTTYYQLHELERAFERSHYPDVYSREELAVKVNLP RVQV-----
	vsx	CM063506.1: 85791203- 85791343	----- RTIFTTSYQLEEELEKAFKDAHYPDVYAREMLSLKTDLPEDRIQ -----

Supplementary Figure 44. otp, vsx/chx, rax alignment.

>Ca_TRINITY_DN37464_c0_g1_i1_vsx_partial_HD

HYPDVYAREMLSLKTDLAEDRIQVWFQNRRAKWRKTEKCWGKSTIMAEYGL-----YGA
 M-VRH----SLPLPDSITKKSEEG-----EL-S
 CAPW---LLVFNHEDRNLGLTLTSSGHL-SQTGMYSQSQ-----EAAEKLKDSOSS
 MDNN-EQE-----EKG-----IETTTEEA-----RS-----
 -----KQAVAGSCEEPSTLV-----
 -----S-MSI-----KHNSNNPVSNR--EELRCNSIDTLGL-RAKALQFSTKMFHV
 IEEQMENNKPQDSNQAPAGQDMAYHEIRDPNQPF--GHNMAQ--YRRTSPAPSTTSKSSL
 QSVTSLF--Q-----

 >Dm_NP_001033832.1_vsx2

 -----MPQSRPFQAIQELLGLAVAAATEAPDITTTAG-----A
 TV-----AK-ERQTPTPPK--TT--NATMATAATS-----AATAATPT
 NAA--EGNL-----TSVSEPQQPQQQQQEQHHQP-HH--HQYREHHQM-----
 -----TMA-----AASR-M-----
 ---AYFN-----A-----HAAVAAAFMPHQLAAVHH-----
 -----HHQHQQHHPHHPHHPHGA VGGPPPPPMQHHPHHPH
 HPLLHAQGFPQLKSFA-----A
 GAG-----
 -----TC--LPGSLAPKDFGMESLNGF--
 -----GVGPNSKKKKKKRRHSRTIFTSYQLEKLEEFKFA
 HYPDVYAREMLSLKTELPEDRIQVWFQNRRAKWRKTEKVGWGSTIMAEYGL-----YGA
 M-VRH----SLPLPDTILKSAKDN-----D--A
 VAPW---LLGMH-----RKSI-----EAQSALKDDSGV
 SDHE-DSA-----GSK-----SAHSEDLRSRCHALSS-STESLNVVSPAPSS-----
 CPTS-ASTSAPPT--STSG-----YAGAASSAATTPTG-----ASTNSSSS
 PHIELGSPSPQQQHLQLQQQQQQQASL-----YLGAGAVVTGCAPPSYHPLLDAAANA
 -----AGAGAAASSKDFHMMNTAVAAAAAAQ-----QQHQQQQQ-----QQQQ
 -ATATAPG-LHA-----HNLAALMEHDP--DAFRNNSIACL--RAKAQEHQARLLNN
 GGLFLQVRR-----FAQ--GQAQIQDPSDMLKLSEEH-NNNSPIPIPPAHSTVNTQVN
 VKMELTA--NGGSMKNYEEA-----

 >Ek_SAP35452.1_vsx

 -----QSNPSAAMHQSRPFQAIQELLGLGHNDSHIHRGSSSLL-----
 -----PG-----ASAY
 SARPL-----VSQPQCFTDS-----
 -----LSSS-R-----
 ---MYFN-----S-----P-----AFLPNVSVPTNLN-----
 -PGH-----S-----
 -----LLSIEASQ-----
 -----GI-RQDNSSDWTf-----SG-----
 -----DSRRL-----DGELT-----
 -----GKNCFQAHVNAASSKKKKKKRRHRTIFAGYQLEELEKAFKFA
 HYPDVYAREMLSLKTDLPEDRIQVWFQNRRAKWRKTERCWGRSTIMAEYGL-----YGA
 M-VRH----SLPLPETILKTAKEG-----IMES
 CAPW---LLAGM-----HKKSM-----EAAEQLKEQSGK
 EDDE-----

 -----PELVPKDK--EELRSNSIAAL--RAKAQEHSAKMLSE
 VANKIVEHT-----NG--ET-----TKPCP-----

```

>Me_MMA49053_vsx
-----
-----
-----
-----MVSVMNLLNGN--RLSQGS-----
-----AACDAGTIQTQ RSPFAIQELLGLGSTPVDPKPNPNRQSAG-----
-----GC-----SSSLCL-----PMTH-----Y
-----LPSTTQSYQGRSTDGH CYADA-----
-----G-R-----
---VYFG---P---P---TTFVSGIT---
---AL---
--TPSVFGFDQSSPAVH---QT---
-----DARTDYDWST-----ERF-----
-----QDDYSSQD-----
-----LSCSE-QH-CYREMQQKQWMEKLTIFTSYQLEELEKAFKEA
HYPDVYAREMLSLKTDLPEDRIQVWFQNRRAKWRKTEKCWGKSTIMAEYGL-----YGA
M-VRH-----SLPLPETILKN-KDG-----NATS
CAPW---LLGRSRIF-----

```

```
>Me_MMA29966_vsx  
-----  
-----  
-----  
-----  
-----  
-----  
-----  
-----  
  
-----MI-ERVRSRNTAIDHLATIFTSYQLEELEKAFKEA  
HYPDVYAREMLSLKTDLPEDRIQVWFQNRRAKWRKTEKCWGKSTIMAEYGL-----YGAM-VRH----SLPLPDSILKSAKEG-----DMSSCAPW---LLGMH-----RKSL-----EAAEKLKDVSDSEDQQ-HSM---DEG-----SR-----  
-----  
-----KMEQ  
-AKPA-----GEEQQTAQVSNK--EELRSNSIASL---RAKALEHSARVFSEVGRCGPGE-----AD-----V-NVQGRKKDDGLPDRRPS--LHSIVSLSF-
```

```
>Is_XP_040065896.1_vsx
-----
-----
-----
-----
-----
-----
```



```

M-VRH---SLPLPETILNSAKEN-----D--S
VAPW---LLGMH-----RKSI-----EAARHLKESGSE
REGS-EPE-----ADR-----A-----TS-----TSP-----
H-----
-----PHEN-----Q-AQSSPA-SST-----
-----TAS
-SSAR-----GRASPDPARDDP--DAFRNNSIACL--RAKAQEHQARILSN
NLLMQVRGLTRTP--IST--PNAEDDGPGDI-----DV-GTEQPLD-----TTT-----
-----HPAHL-----
-----
>Dp_XP_046462950.1_vsx
-----
-----
-----
-----MMNLLQHP--DRNIHM-SAPGLLGGLLAASS--
-----GVSGVSANSLPQRSPFAIQELLGLNTSGSSVGHSSSHHQGDHG---
-----V---GGG-NSQQHVYS-----HAMIAAAHHHHHHSHQHPAAAMFA-HPA
HQHQ-----PHHH-SHHGSPAS-VTGGSQSGQSGELAG-HHHHHNSLQAHQALS---
-----ALS-----AAAA-A-----
---SYAS-----ATAAA-----AAAASGSR-----
---C---S-AAAAYNFTH--HHQNF--AAA-----AAAAAAMVGDPSSASRM---YFNP
--AAAAAAFLHQPTAA-----AVAAGSIGSVSASAPGASTAV
GNGTSGGGVFPSPDSAAMRS-----DTN-----
-----ASVGLEGRGVG-----GD--GSHCNSKLDELNVSGGVG-
-----S-NSNGINGA--NNPNRKKKKKRRHRTIFTSYQLEELEKAFKEA
HYPDVYAREMLSLKTDLPEDRIQVWFQNRRAKWRKTEKCWGRSTIMAEYGL-----YGA
M-VRH---SLPLPDTIVKSAKEN-----E--C
VAPW---LLGMH-----RKSL-----EAAHQLTSGDES
NKED-DDE-----NGG-----GVGGGS-G-----ERTRD-----SGN-----
GPND-G-----GDHSS-----G-----DEQQQSEKS
G-----HI-IRESIGGNHHN-----KRNLSPS-GIQSAGAVGQHH-
-----HAAHLSHLHPQH-TSTSS-----SSEA-----STYM
-ASYA-----AAQLHHANRSPV--DDFRMNSVAEL--RAKAAEHQARIAGG
FSVSHHHQQQLQH--LQQ-QQQQQQQQQQQ--QQ-A-----QQRDG--E
FLISSDS--SSSSF-----
-----
>Pf_TRINITY_DN16317_c3_g4_i3_vsx
-----
-----
-----MNATQAS--SNTPHH-HHHHQHPFVH----
---SSFPN---FTGGSSGSTMPQRSPFAIQELLGLGNCHDTGSSAAQNSCIPASSA
VISRPSL---NVG-SSS--VQN-----HQTAYVGS-----SVAAAAHLSSAY
HPRPL-----GGQSQCFSDP-----
-----S-R-----
---MYFN-----TAAAAAFMANVSVNSVN-----
-----TAGT---TMP-
-----VLGLDQSMAAAA-----GAVV-----
---R---GESNPDSFWRY-----DRY-----
-----GGNI-----DQSGLDSEMLEKSGF--
-----GSSG-LGINKKKKKKRRHRTIFTSYQLEELEKAFKEA
HYPDVYAREMLSLKTDLPEDRIQVWFQNRRAKWRKTEKCWGRSTIMAEYGL-----YGA
M-VRH---SLPLPETILKTNKDE-----N-EP
CAPW---LLGMH-----KSKI-----EAAEQFNHHETD
EKED-G-----
-----
-----GSH-----S-----SD-SSS--PNL-----
-----IS-----QQ-----
-----ENQQQIQGSHCK--EDFRSSSIAAL--RAKAQEHSKMLLH
EASNRAM---DI---MSG--VTAAEASQKAQ-----DF-NNV-----LKTSS--A
AVVNSLM--SSGAFHPNATDNSSQQHRINLEKNNSCDHMAMF-----

```



```

-----G-R-----
--VYFG-----TAFMH-A---NMPVHGMG-----
-P-----VP---SMP-
----LLGYEHQALGAR-----AA---E-----
---A---GAAAASADYFWHY-----DR-----
-----FSSGSSSQDFI-----SP--HPSNSGMVQGEFGSKSSLG-
-----GPMDEMGGPA-GIGAHCKKKKKRRRTIFTSYQLEEELEKAFKEA
HYPDVYAREMLSLKTDLPEDRIQVWFQNRRAKWRKTEKCWGKSTIMAEYGL-----YGA
M-VRH---SLPLPESILKSA-----

```

>Cs_XP_023227221.1_vsx2

```

-----MVSVMNLLHGD--CPTTLSHS-----QLT--
-QP-----STSTNSGQGVQQRSPFAIQELLGLSNHSDSPAAAVP-----
-----S-HRPPTTVTS-----TVGMAVGLAG-----PP-TGAAPQPCTPY
QPRP-----IVT-----OPTASQCFPDP-----
-----G-R-----
--MYFG-----TAAAA-FMPNMSVPAAMS-----
-S-----VTGMTGITPSMS---TMP-
----MLGFEQPTSA-----
-----PRQDATDSFWHY-----EQR-----
-----FPVNGLSH-CL---SS--Q---STGLVTDYSSKNL--
-----SDDV-GSLNKKKKKKRRHRTIFTSYQLEEELEKAFKEA
HYPDVYAREMLSLKTDLPEDRIQVWFQNRRAKWRKTEKCWGKSTIMAEYGL-----YGA
M-VRH---SLPLPDSILKSAKEG-----DMSS
CAPW---LLGTS-----SVNGMHRKSL-----EAAEKLKDVSDS
EDQQ-HSM---D-----
-----EG-----
-----T-R---KVEQ
-PKTS-----GEEQQTQVSNK--EELRSNSIASL--RAKALEHSARVFSE
VGRCGP---E---AEV--NVQ---G-----RKKEDGVPDRRPS--L
HSIVSLF-----

```

>Lp_XP_022249270.1_vsx

```

-----MVSVMNLLHGD--RRTIHV-SAPNTLNPYPTSLT--
-SITQH-----VGPGLGSPTIQRTPFQELLGLDNQRENSPSSHRSAS-----
---ARL---TA---P--T-----PVTGPTC-PTF
QPRG---LGT-----QPGPPHCFPDP-----
-----G-R-----
--MYFG-----PAA---FMPTMSVPSGMT-----
--GV--T--G--LH-----GVTGVPRVTPISS---CMP-
-----MLTLDQPTTG-----
-----TRQHSTTDSF-----
-----GS-----TT--F---PPLVFMN-QKNR--
-----STYC-PYWSKEEKKRRHRTIFTSYQLEEELEKAFKEA
HYPDVYAREMLSLKTDLPEDRIQVWFQNRRAKWRKTEKCWGKSTIMAEYGL-----YGA
M-VRH---SLPLPESILKGTKEG-----EL-S

```

[illegible]

>Sd_XP_035219442.1_vsx

```
-----
-----
-----
-----
-----
-----
-----
-----MLD--RRIADRSM-----
--R--KG-T--FHEFRH--ALVQDWALLPQKAIND---TIASLPRQCQASS---SAR-
--VYHTCHYLDISTA-----
-----YSSYSFDLLRCE-----
-----GLLG-----AD--F--S-----KNGF--
-----GDDI--NSMNKKKKKKRRHRTIFTSYQLEEELEKAFKDA
HYPDVYAREMLSLKTDLPEDRIQVWFQNRRAKWRKTEKCWGKSTIMAEYGL-----YGA
M--VRH---SLPLPESILKSAKEG-----DVSS
CAPW---LLGMH-----RKSQ-----EAAEKLKSDSLS
SEEH-HSQ---QQL-----ALPQSSSS-----NNNTS-----TT-----
A-----A-----MTVQQPSFH
HHH-----HHHHS-HQQRHHHQGSS-----SSTSSSSR-GSS--PILGGP--
-----S-TSS-----GSHH---QQQQ
-P-----PNK--EDLRSSSIASL--RAKAIEHSAKVFE
GRDVPP---S--DPS--AS---PRAK---D-----VVLMERPS--I
HSIGSLF-----
-----
```

>Pa_TRINITY_DN13907_c1_g1_i8_vsx

```
-----
-----
-----
-----MVSVMNLLHGD--PHVGR--G--GPHHPHPGPLS--
-VP---Q---STASSAAATSHQSPFAIQELLGLGNTSDGARSSPA-----
-----AP-----NVS-----AVSMGHHHPY
HPRA-----PQQGFDP-----
-----G-R-----
---MYFG-----AGFM---PGAMHA-----
-----MPAPP---HMP-
-----MLGFEQPT-----
-----SHPRND---A-----
-----NVLG-----AD--F--T-----KNGL--
-----GEDI--NSSNKKKKKKRRHRTIFTSYQLEEELEKAFKDA
HYPDVYAREMLSLKTDLPEDRIQVWFQNRRAKWRKTEKCWGKSTIMAEYGL-----YGA
M--VRH---SLPLPESILKSAKEG-----DVSS
CAPW---LLDNP-----WKLT-----PGCSSVI-----
-----
-----
-----
-----
-----
-----
```

>Mm_TRINITY_DN10076_c0_g1_i1_vsx

```
-----
-----
-----
-----MVSVMNLLHGD--PHGSGRPGAPLSHHPHLSV-S--
-GPGP-----QSTAASHQSPFAIQELLGLGNNDNNNTTARSPP-----
---P-----NSP-----HVS-----T--ATAAVSMGHPY
HPRA-----PQQGFDP-----
-----G-R-----
```

```

---MYFG-----AGFM---PGAMHG-----
-----M-----P---APP---HMP-
---MLGFDQPSQA-----
-----LSHPRTD---A-----
-----NVLG-----AD--F--T-----KNGL--
-----GEDI-NSSNKKKKKKRRHRTIFTSYQLEEELEKAFKDA
HYPDVYAREMLSLKTDLPEDRIQVWFQNRRAKWRKTEKCWGKSTIMAEYGL-----YGA
M-VRH---SLPLPESILKSAKEG-----DVGS
CAPW---LLENP-----WKLT-----PGCSSVI-----

```

>Pp_c99447_g1_i3_vsx

```

-----MVSVMNLLHDS--HHRQ--PG--SHHAPHLSVLS--
-NPGPSPG-----IGSHQNPAPTHQRSFPAIQELLGLGNSTDNNNTGSNNSS-----
---PAT---SPG-QRPNPQPS-----HLS-----VV-TSAGTHCGAPY
QPRP---LTSH---HHHQPTAQCFPDP-----S-R-----
---MYFG-----T--A-FMPNM-----
-----SMGMSAMN-----PATSMCAA---AAA-
-----MLGFDQTPPG-----
-----HHRPDPTA-----
-----AVLG-----TE--F--T-----KNGI--
-----SEEQ-SSVSKKKKKRRHRTIFTSYQLEEELEKAFKDA
HYPDVYAREMLSLKTDLPEDRIQVWFQNRRAKWRKTEKCWGKSTIMAEYGL-----YGA
M-VRH---SLPLPESILKAAKEG-----DVTS
CAPW---LLAPM-----HWQAL-----VSGTANAT-----
---Q-SSP---LSI-----YRDASKVT-----G--SS-----GE-----
A-----
-----QG-----SRI--
-----Q-FPG-----GDDE---QQQR
-P-----QLDDNP--AAAVR-----

```

>Pan_CM063506.1_85791203-85791343_vsx_partial

```

-----RTIFTSYQLEEELEKAFKDA
HYPDVYAREMLSLKTDLPEDRIQV-

```


>Lp_XP_022245320.1_otp

-----M-----NIA
KN--DGTG--ESLKPN--G-----GADHHYTPGGA---TLQVLESTGI
HQRPMVVPGN-----GSGVG-SGSGPTHCSDSKPLKPKRHRTRFTPAQLQEELERSFAKT
HYPDIFMREELAMRIGLTESRVQVWFQNRRAKWKKRKKTN--VFRNP-----GV
LLPSH---SLPPLGFMS-DG-----L
CSFG-----PGPDTRWSMAA-----GMTQM--GSPPLTL-----
-----GPP-----SLSRQSG-LGQS---L-----
-----STPSMTGINQSGVP-----LNHGP-----LA-----GSV--
-----T-----TTGYQTHY-GLN----TLNS--PSGVAAPHGLGPVTAGMAM
A-----AASPPNSSSSTPPQLSAACTMSM-----
-----QGLD-----PGACNGEIDVWRETSITTLL--RRKA-----
-----IEHTATMSVFR-----

>Me_MMA23872_otp-A_partial_HD

-----MLNNL---H-GATV--
PLVAKEDSSGDV-----
-----LKPN--GGVDPT-----QVSHGGTI
QN-----LDN-TSHSGGPSSDGEKPVKQKRHRTRFTPAQLQEELERSFAKT
HYPDIFMREELAMRIGLTESRVQ-----

>Pf_TRINITY_DN16730_c1_g2_i2_otp


```

PLAFSLLGS-----
-----KED---PGLVK-----S--NGGDTGPL-----QVLDGGAL
HSVGH-----GPPSH-NEPPSNSSDNNKPVKQKRHRTRFTPAQLQEELERSFSKT
HYPDIFMREELAMRIGLTESRVQVWFQNRRAKWKKRKKTN--VFRSP-----GA
LLPSH---SLPPFGSMG-DG-----F
CSFG-----AEARWPM-----SQM--GPGPTLP-----
-----LGP---TSLARQPA-LAQS---L-----SQH
SAAA---AAMGGMNQYMAV-----QVSTQLATNSLPSI-----TTS--
-AC-G-----SPMYQSPYGALT---ASP-----TVAPPGLT-----
AGMPT----ASGGTPPGASPSPPSPQQPQP-----
-----CAVESSDMWRGTSIASL---RRKA-----
-----LEHTASMSVFR-----
-----
-----

```

>Pp_c103695_g1_i1_otp

```

-----MLNNL---HGAA---
PLGSKEDGH-----
-----SLK--SANDGIGL-----SVLDGGGI
HGSGGGSIPGGTGSAPSND-GGVQGNVGGGGDKPVKQKRHRTRFTPAQLQEELERSFSKT
HYPDIFMREELAMRIGLTESRVQVWFQNRRAKWKKRKKTN--VFRTF-----GA
LLPSH---SLPPFGSMG-DG-----F
CSFS-----PAADTRWPM-----AQM--GSPGLGP-----
-----LGP---SLSRQPS-LGQS---L-----SQH
SAVA---AAMGGMN-----QVSSQLVSNLSLPSI-----TST--
-SA-C---S-----ASVYQSPYGSLN---SLNA--GAAVAPPPGLSPVTAGITA
AGMPGVASGAATGTTTPDSSSSPQLGCGGAGMGL-----
-----GGTSMVAG-----SHHPGADPGDMWRGTSIASL---RRKA-----
-----LEHTASMSVFR-----
-----
-----

```

>Ag_TRINITY_DN76899_c8_g3_i1_otp

```

-----MLNNL---HGG-ST--
PLVTKDEGGGGE-----
---GL-----KAN---GGVDGG-----GGTL-----QVLDGGGG
GGVH---GAAGHGGGGGGG-GHGVPQSAGDADKPVKQKRHRTRFTPAQLQEELERSFSKT
HYPDIFMREELAMRIGLTESRVQVWFQNRRAKWKKRKKTN--VFRSP-----GA
LLPSH---SLPPFGSMG-DG-----F
CSFG-----PAADTRWPMTA---SMTQM--GSPGL-P-----
-----LGP---SLSRQPG-LGQS---L-----SQH
SAAAAAAAAAAMGGM-----SQNLAPSSLPSI-----TTS--
-SC-S---S-----AAVYQSPYGGLN---SLTS--PS-VAPPPGLAPVTAGMAA

```



```

-----
-----
-----
-----
-----
-----
-----
-----MSRR-----L-----LPMD-VRRAPDTLMRNPMADI---HEGRQR--
ALEFS--LHLLK-----
---TY-----SYN---AGLQQS-HEAV--KASDQDSL-----KPVPEPSP
QGHP---TLQGS-----AQ-VTHHSSHGGDADKPAKQKRHRTRFTPAQLQELERSFSKT
HYPDIFMREELAMRIGLTESRVQVWFQNRRAKWKKRKKTSG--VFRSP-----GA
LLPSH---SLPPFGSVG-AAAAA--AE-----AA---AGL
ASFSG---ADGGAPPRWAPHA---AMQQM--GH-----
-----LGP---ALARQPP-LGPP--S-----SMP
AIAQV--SCVYGEQ-----VDRNQ--HSLG-R-----GLS--
-KC-----FG-----
-----
-----
-----
-----

```

>Lp_XP_013789393.1_rax

```

-----
-----
-----
-----
-----
-----
-----
-----MSVEVNTNHRRYLQAKV-----SASDN---NSRGCLSLHA
HSIDALLGERRQKPPEKRLDLSELN---E-----NQTNLLQ---IR-----
-----TF---PD-----LNP-----DEDLV-----PFQ
C--L-----THY-----KGPDENHK---EQQNR-EPGFSNN-----AL--
-----A-----AEEEFDVRSETHEDSPKRPHRRSRTTFTTYQLHELERAFEKT
HYPDVYSREELAMKVNLPPELRVQVWFQNRRAKWRQERLENENTLRSLRPTKSSDCPLAG
AGMT---TNSMNYVSGPC---PSLGLSS-----SS--MLTG
ASMQ-----RLTLDLSDW-MTPPRMGPPLS-----
-----DTLPGF---LAHSPCMCP---SYV-----TPSTLPTASVS
N-----STAFSG-----P-----
-----
-----PTGNISI-CSGMLGSSSPNLNLSVNEMDF-----
-----TFADSKPTDLRSSSICAL---RLKAKEHVNLIKR
LTSV-----
-----
-----
-----
-----
-----
-----

```

>Is_XP_029829045.2_rax

```

-----
-----
-----MKVNLPEVRVQVWFQNRRAKWRRQEKMESQSALQHRLG-----GGG
-----QSP-----SS-----PG-SGSNA
MVGP-----GLPLDPW-LAPPLLSGAPF-----
-----MG-----TAAAPFYA-----TSTAAAAFLQG
R-----ADT-----
-----
-----YQRLAPPSSPLNLALSNEPK-----
-----SQSQPQDGGDDPTNSDRTSSIVAL---RLRAREHVQLLCKG
V-----
-----
-----

```

>Ek_SAP35450.1_rax

```

-----
-----
-----
-----
-----
-----
-----
-----
-----TAE-----KRDSKSVN---NSSVVNCSPS
HTIDAILGLKERKSASSETLTPVD-----SD-----
-----A-----KSP-----RGA-----
-SES-----QKL-----LESKEPLT---NDENDSFSSNNASS-----CN-K
RD-----ED-----DLMEDDGDCDFDGPKKKHRRNRRTFTTFLHELERAFEKS
HYPDVYSREELAVKVNLPVVRVQVWFQNRRAKWRRQEKMESNALKLNTEYPMS-----S
LAR-----TTTN
GLSS-----TLPLDPW-LTPPIANSTPI-----
-----QTLPGF---LSHAPATYP---SYL-----TPPTSTTFSSS
C---IIG-----
-----
-----QEKFDDNDPRSSSIVSL---RMKAKE-----
-----
-----

```

>Pf_TRINITY_DN8076_c0_g1_i1_rax

```

-----
-----
-----
-----
-----
-----
-----
-----
-----MAKAQSR
HDLDCFLKSQ-----
-----RP-----HTSNL-----
---I-----RCPNTAQI-----SDKDPEK-----S-----GH-D
KYKA---EME-----LNEENMHEDLGNDENGQKKKHRRSRTFTTFLHELERAFEKS
HYPDVYSREELAMKVNLPVVRVQVWFQNRRAKWRRQEKLENSASLRIHDLERHS---P---
-----DSGIR---N-----DLAR
SAVT-----FLPMDPW-LTPPVFNSTLG-----
-----HAMPGF---ITSSYPGFIPQPSIL-----ASSASRIR---
-----SH-----
-----
-----TISPPIVLSQSPRDL-----
-----

```



```

-----
-----
-----
-----MQVDRSPLAPL-----AFSESNG
CRSSHLASTLGNRTI-----
-----S-----DERKSTPI
HSIDAILGLKTWKMRNSPSPPPSK---SPC-----DRELMG---EC-----
---SSPHHV---KI-----FRP-----TE-----
-DLE-----DHTRE---NSRSPDDNHM-----SSNSHTDSNG-----
-----GGNKADGDSGSEEGPKKKHRRNRRTTFTTYQLHELERAFEKS
HYPDVYSREELAMKVNLPVVRVQVWFQNRRAKWRRQEKLESQNALRTLGHCSGDYSLTGG
-----RPQSPMTSL-----ALGSCS-----PPTSVMVTG
TGPS-----QISLDPWSLASPLIGSSLA-----
-----AGLPGF---LNRHPSVYP---SYL-----TPAAPATTLTV
P-----CSMA-----
-----
-----ELASGA-PSIPLEHGPVNLSLGPTDT-----
-----KPDDNGLDPRSSSIVSL---RMKAKEHVEYLSKG
LVIM-----
-----

```

>Bm_XP_004933582.2_rax

```

-----
-----
-----
-----
-----
-----
-----
-----MDAPDE-RFFADRS
PDRS-----DTNSPRSAVSTP-----NSN-----
-----SSINVTQSI-----
---SLSQQNLETL-NRMGMFFHAQQMQLNQSF---DVKRLGLT---QVPGASQGPR
HSIDAILGLSSQRQARLSEFEQR-----REECDPVP---VS-----
-----PG-----
--AV-----ESAGE---GSCNSNDGYN---QRT-----D
DKSP--P-----ASDDDTPOQSTDNKKKHRRNRRTTFTTYQLHELERAFEKS
HYPDVYSREELAMKVNLPVVRVQVWFQNRRAKWRRQEKMEAAARLGLSEYGC-----GGGG
-----MPR
LSAL-----GLPVDPW-LAPPLLS-----
-----ALPGF---LSHPPSGYP---SYL-----TAAGPAPAP---
-----
-----
-----PAAPDPRSSSIAAL---RMKAKEHVENISKG
LQMV-----
-----

```

>Sd_XP_035208675.1_rax-A

```

-----
-----
-----
-----
-----
-----
-----
-----MQVEVNVDRSPLATL-----TFPESNG
CRGSHLPAG-----H--K-----AVSDDRKTDLPLHSID-----
-----AILGLRTWRKKRRSPSPSPPN-----KSPCDDLQ---GE-----
---PSHPRV---KI-----FRP-----AE-----
-DID-----DHTRE---NSSSPTENQH---S---SP-VHADLSG-----GS---

```

```

-----SKPDPDSGSEEGPKKKHRRNRRTTFTTYQLHELERAFEKS
HYPDVYSREELAMKVNLPVVRVQVWFQNRRAKWRRQEKLESQNALRSLGHVPDYPVSG--
-----GRPQSPITSLA---TALGSCS-----PPTSAVTP
SGPS-----QLPLDPWSLTAPLLGSSLA-----
-----AGLPGF---LNRHPSVYP---SYL-----TPAAPTAAALTV
P-----CSIA-----
-----ELAPGT-PSGPLEHSGPVNLSLGPTDT-----
-----KSEESGLDPRSSSIVSL---RMKAKEHVEYLSKG
LAIM-----

```

>Pa_TRINITY_DN30596_c0_g1_i3_rax-A

```

-----MQVEINMDRSPLAPL-----TFPESNG
CRTGNK-----GGAEDRKSELPLHSID-----
-----AILGLRTWRKKRRSPSPTPTP-----KSPHAILQ--GE-----
---TNHSHT---SL-----FRP-----AE-----
-DID-----DHTRE---NSGSPTDNQ-----H-SSSQDLDSGG-----CT--
-----KQQDPDSGSEEGPKKKHRRNRRTTFTTYQLHELERAFEKS
HYPDVYSREELAMKVNLPVVRVQVWFQNRRAKWRRQEKLESQNALRTLGHVPGDFTGSGG
-----RPQSPITSLA---AALGPCS-----SPTSAVTP
TGPN-----QLSLDPWSLTAPLLGSSLA-----
-----AGLPGF---LNRHPSVYP---SYL-----TPAAPAPALTV
P-----CS--MGEL-----
-----APGT-PSGPLDLPGPVNLSLGPTDS-----
-----KPEDGLDARSSSIVSL---RMKAKEHVEHLNKG
LTIM-----

```

>Mm_TRINITY_DN14066_c0_g1_i1_rax-A

```

-----MIFMQRLCIFIILKHIF-DQKKIKTWHE-IFLIVNL
EIRS---FEEIYQLTVLSCYRKRTLPLVLYNLLTMQVEISMDRSPLAPL-----TFPESNG
CRTGN-----K-----GGPDDRKLDRPPLHSIDA-----
-----ILGLRTWRKKRRSPSPTPTP-----NS---PH-TVLQ---EE-----
---ANHTHT---NI-----FRP-----AE-----
-DID-----DHTRE---N-SGSPSDI-----QH-SSSQDLDPGG-----GN-K
Q-----QDPDSGSEEGQGPKKKHRRNRRTTFTTYQLHELERAFEKS
HYPDVYSREELAMKVNLPVVRVQVWFQNRRAKWRRQEKLESQNALRTLGHVPGDYGGSGG
-----RPQSPITSLA---TALGPCS-----SPTSVVTP
TGPN-----QLSLDPWSLTAPLLGSSLA-----
-----AGLPGF---LNRHPSVYP---SYL-----TPAAPAAALTV
P-----CS--MGELAT-----
-----GT-PSGPLDHPGPVNLSLGPTDS-----
-----KAEEGGLDARSSSIVSL---RMKAKEHVEYLNKG

```

```
LAIM-----  
-----  
-----  
>Pp_c102000_g1_i1_rax-A  
-----  
-----  
-----  
-----  
-----MHSES  
-----SEKMLDATTRSP-----LAPLACSDFLT-----SFALSEK  
CRLQATTKTVKRQED-DIVSN--T-----RNNNNINSSLPLHSIDAIL-----  
---GLKSSKKSR-----RHHDNIYSRQQ-----QHQBH  
ESSSASSRYRSHHHSVKNTRNGED-----ELLM---EN-----  
---SG-----NVR-----  
-----RE---ASTEKDDGSH---HQHESSSSSGI-----  
-----KCD-----GDGSEDGGNGGSSDSNKKKHRRNRRTFTTTYQLHELERAFAEKS  
HYPDVYSREELAMKVNLPEVRVQVWFQNRAKWRRQEKLSQLALRSLGHVTEFSGNGSS  
LTRSHSPSMS-----CT---GGLS--P-----LGSP-TS  
TTTP-----TLSLDPW-FTASPLLGSSL-----  
I---VPCSP-----SYL-----TPPGPTPTMAG  
-----KST---V-----  
-----TSVDSh-CS-----PINLSL-----  
-----TTESKDNTCFDPRSSSIVSL---RMKAKEHV DYLSKG  
LAIM-----
```

-----MKVNLPEVRVQVWFQNRRAKWRRQEKLESQNA

```

-----
-----MRMEVD-----KNV
CD-----PRRAD-----SLPK-----ARF-----
--S--SG-KPEDLK-----GCGASLPVNIS-----LYS--
--IDAILGLKSKRRHHGSTGKSYDV---TPSDSQPRTNNPIVR---GE-----
--TT--SQ--TT-----LTP-----AD-----
-TSR-----SPLSS---SSSAASSC-----E--QASKEYIDGCS-----GA-A
SISD--DARED-----GTRFDRGVGDGSPDDSPKKKHRRNRRTFTTYQLHELERAFEKS
HYPDVYSREELAMKVNLPVVRVQVWFQNRRAKWRRQEKLENQNALRTLGHVTSDFRSG--
---SSVLP RSQSPSSCLT---PTLGSCS-----PPT-S-AS
SPSG-----PLTLDPW-MTSPLLGSSTL-----
-----GLPGF---LNHHPAVYP---SYL-----TPATPSSMPVP
C---PAGLNGIVA---GTP-----
-----V-----
-----TSAAL-----DHSGPINLTLASPD-----
-----CKVGEEGVMDLQSSSIVAL---RMRAKEHVEFIKKG
LTII-----

```

>Ca_TRINITY_DN17059_c0_g1_i1_rax-A

```

-----
-----MRMEQGPLIER-----ENAS-----
-----SDPRRGDISKS-----RLS-----
-P-----SNKQ-----
-----DPSNNEDFKKSCAP---SFSLNVS-----LYS--
--IDAILGLKSKRRPHVSGRSADV---SSRENHSRTTNPQVR---SD-----
-----P--AS-----LTP-----AD-----
-TSQ-----SPLSV---SSSGRTSPF---EHSTKGYGEDRNCS-----GP-Q
VVSD--DGRED-----PLRFDI---DGSQEDCPKKKHRRNRRTFTTYQLHELERAFEKS
HYPDVYSREELAMKVNLPVVRVQVWFQNRRAKWRRQEKLENQNALRTLGHVTS-DYRGGG
---SSL-PRSQSPSSALT---PTLGSCS-----PST-S-AS
SPSG-----PLTLDPW-MTTPLLGSSTL-----
-----GLPGF---LNHHPAVYP---SYL-----TPATPSAMPVP
C---PAGLNG---VVAGT-----
-----PV-----
-----TSAAL-----DHSGPINLTLGNAEY-----
-----KLDDEGVMDLQSSSIVAL---RMRAKEHVEFIKKS
LTII-----

```

>Sm_SMAR009838_rax

```

-----
-----MIDW-----
-----SHIQKITQRNVY---KSY-----
--AVCPRTNVQLLSVVIPQ---AL--VI-----SRD--
--STGFMGFCKISRDFTRFREFFY-----EIH-EIL---TK-----
-----Y--PS-----
-----EWLTKEI---HENDSAHSNGVASN-----SA-D
EKSP--DSGVN-----GGEGLA---MGLQEDGPKKKKHRRNRRTFTTYQLHELERAFEKS

```


>Pp_c110545_g4_i3_rax

>Pt_g23262_rax-B

```
>Sd XP 035223039.1 rax-B
```


GSNT-----PLL FDPW-LTPSLLGLRLP-----
-----GLPGF-----QLHTVYP---SYL-----TPANLSS-----

-----AAM-----
-----SSE--SSS KIKSSPLNLSLDNDDHTN-----
-----DEHQDPRSTSIASL---RHKAKEHLQVISKG
QIT-----

