Article
Discoveries

The evolution of Sox gene repertoires and regulation of segmentation in arachnids

Luis Baudouin-Gonzalez[1,*], Anna Schoenauer[1,*], Amber Harper[1], Grace Blakeley[1], Michael Seiter[2], Saad Arif[1,3], Lauren Sumner-Rooney[4], Steven Russell[5], Prashant P. Sharma[6] and Alistair P. McGregor[1,3,+]

[1] Department of Biological and Medical Sciences, Faculty of Health and Life Sciences, Oxford Brookes University, Oxford, OX3 0BP, United Kingdom.

[2] Department of Evolutionary Biology, Unit Integrative Zoology, University of Vienna, Althanstrasse 14, 1090 Vienna, Austria.

[3] Centre for Functional Genomics, Oxford Brookes University, Oxford, OX3 0BP, United Kingdom.

[4] Oxford University Museum of Natural History, Parks Road, Oxford OX1 3PW, United Kingdom.

[5] Department of Genetics, University of Cambridge, Downing Street, Cambridge, CB2 3EH, United Kingdom.

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

* Contributed equally

+ Corresponding author

Key words: evolution, development, Sox genes, segmentation, spiders, arachnids, arthropods

Abstract

The Sox family of transcription factors regulate many different processes during metazoan development, including stem cell maintenance, nervous system specification and germline development. In addition, it has recently become apparent that SoxB genes are involved in embryonic segmentation in several arthropod species. Segmentation in arthropods occurs in two main ways: long germ animals form all segments at once, best exemplified in the well-studied *Drosophila melanogaster system*, and short germ animals form anterior segments simultaneously, with posterior segments added sequentially from a segment addition zone. In both *D. melanogaster* and the short germ beetle *Tribolium castaneum,* the SoxB gene *Dichaete* is required for correct segmentation and, more recently, we showed that a close relative of *Dichaete, Sox21b-1,* is required for the simultaneous formation of prosomal segments and sequential addition of opisthosomal segments in the spider *Parasteatoda tepidariorum*. Here we further analysed the function and expression of *Sox21b-1* in *P. tepidariorum*. We found that while this gene regulates the generation of both prosomal and opisthosomal segments, it plays different roles in the formation of these tagma reflecting their contrasting modes of segmentation and deployment of gene regulatory networks with different architectures. To further investigate the evolution of Sox genes and their roles we characterised the repertoire of the gene family across several arachnid species with and without an ancestral whole genome duplication, and compared Sox expression between *P. tepidariorum* and the harvestman *Phalangium opilio*. The results suggest that *Sox21b-1* was likely involved in segmentation ancestrally in arachnids, but that other Sox genes could also regulate this process in these animals. We also found that most Sox families have been retained as duplicates or ohnologs after WGD and evidence for potential subfunctionalisation and/or neofunctionalization events.

Introduction

The subdivision of the anterior-posterior body axis into metameric units, a process termed segmentation, is a characteristic of three major animal phyla: arthropods, annelids and vertebrates (Chipman, 2010; Clark et al., 2019; Couso, 2009; Hannibal and Patel, 2013; Tautz, 2004). It is thought that the remarkable morphological diversity found in these taxa owes much to the modular organization of their body plan, which allowed the flexible adaptation of different segments to fulfil particular functions (Chipman, 2010; Damen, 2007; Tautz, 2004). Thus, the origin and evolution of segmentation has long been one of the main focuses of Evo-Devo studies (Chipman, 2010; Clark et al., 2019; Couso, 2009; Damen, 2007; Hannibal and Patel, 2013; Tautz, 2004).

The study and comparison of segmentation among arthropods has been crucial in the characterisation of these mechanisms and their evolutionary origins (Clark et al., 2019; Damen, 2007; Davis and Patel, 1999). Arthropod segmentation can generally be subdivided into two main modes, long germ and short germ; although these terms originally referred to the number of segments specified before gastrulation in insects (Davis and Patel, 2002; Krause, 1939), they are now used more widely as descriptors of arthropod segmentation. In long germ segmentation all segments are specified more or less simultaneously, whereas in short germ a variable number of anterior segments are first specified simultaneously (depending on the species), followed by the sequential formation of posterior segments from a growth or segment addition zone (SAZ) (Janssen et al., 2010; Liu and Kaufman, 2005).

Segmentation is best understood in *Drosophila melanogaster*, which exhibits long germ segmentation (Damen, 2007; Liu and Kaufman, 2005). In this dipteran, segmentation is regulated by a hierarchical gene cascade, which ultimately subdivides the embryo into individual segments (Clark et al., 2019; Damen, 2007; Liu and Kaufman, 2005). Long germ segmentation appears to be confined to holometabolous insects, while short germ segmentation is extensively represented in other insects and all other major arthropod subphyla (i.e. crustaceans, myriapods and chelicerates) (Clark et al., 2019; Damen, 2007; Liu and Kaufman, 2005). Therefore, long germ segmentation is reconstructed as a derived state, while short germ segmentation appears to be closest to the ancestral state of arthropod segmentation (Damen, 2007; Davis and Patel, 1999; Peel, 2004). It is

therefore essential to investigate the genetic mechanisms controlling short germ segmentation among arthropod taxa to inform our understanding of ancestral mechanisms and how they have evolved in different lineages. Over the last two decades, work on short germ insects, myriapods and chelicerates has provided further insights into the evolution of arthropod segmentation by identifying both novel and conserved developmental genetic mechanisms among these animals (reviewed in Clark et al., 2019). Work on spiders in particular has made substantial contributions towards our understanding of chelicerate segmentation and the evolution of arthropod segmentation in general (Hilbrant et al., 2012; Leite and McGregor, 2016; Oda and Akiyama-Oda, 2020). In contrast, relatively little is known about the regulation of segmentation among other arachnids and given they are represented by over 100,000 named species with diverse body plans, these animals hold great promise to learn more about the evolution of development (Garb et al., 2018; Leite and McGregor, 2016; Schwager et al., 2015b).

The spider *Parasteatoda tepidariorum* has been used extensively as a model for the study of segmentation in chelicerates (Hilbrant et al., 2012; Oda and Akiyama-Oda, 2020). Spider segmentation falls within the short germ mode but encompasses at least three different mechanisms along the anterior-posterior axis (Hilbrant et al., 2012): first in the head, the pre-cheliceral, cheliceral and pedipalpal segments are added through the splitting of stripes of segmentation gene expression that is regulated via anterior-to-posterior travelling waves of genes including *orthodenticle* (*otd)* and *hedgehog* (*hh)* (Kanayama et al., 2011; Pechmann et al., 2009). Second, the leg-bearing segments of the prosoma are also added through splitting of wider gene expression domains, but in this case it happens more or less simultaneously and is regulated by the gap gene-like activity of *hunchback* (*hb*) and *Distal-less* (*Dll*) (Pechmann et al., 2011; Schwager et al., 2009). Finally, the opisthosomal segments are sequentially added from a SAZ, which is in part established by the activity of Wnt and Notch-Delta signalling pathways. The interplay between these pathways directs sequential segment addition from the SAZ by regulating dynamic expression of pair-rule gene orthologues, including *even-skipped* (*eve*), and other segmentation genes such as *caudal* (*cad*) (McGregor et al., 2008; Oda et al., 2007; Schönauer et al., 2016).

It was thought that the simultaneous generation of prosomal leg-bearing segments and sequential addition of opisthosomal segments in *P. tepidariorum* were regulated by largely independent GRNs. However, it has recently been shown that the loss of *arrow* abrogates all segments (Setton and Sharma, 2018) and the SoxB gene *Sox21b-1* is involved in both prosomal and opisthosomal segmentation (Paese et al., 2018a; Paese et al., 2018b). Knockdown of *Pt-Sox21b-1* results in the loss of all leg-bearing prosomal segments and disrupts the formation of the SAZ, leading to loss of all opisthosomal segments (Paese et al., 2018a). *Pt-Sox21b-1* knockdown embryos lack expression of both *Delta* (*Pt-Dl)* and *Wnt8* (*Pt-Wnt8)*, key regulators of SAZ formation, which explains the loss of opisthosomal segments (Paese et al., 2018a). Further analysis of *Pt-Sox21b-1* RNAi knockdown embryos suggests that this gene acts upstream of the posterior GRN by regulating *Wnt8* and *Delta-Notch* signalling during SAZ formation (Paese et al., 2018a).

Interestingly, *Dichaete,* a closely related SoxB group gene, is known to regulate segmentation in both *D. melanogaster* and *Tribolium castaneum* (Clark and Peel, 2018; Russell et al., 1996). This finding, together with our work, suggests that SoxB genes regulated segmentation ancestrally in arthropods and continue to do so in simultaneous and sequential segment formation in different lineages. However, this conclusion requires a more detailed understanding of the role of *Sox21b-1* during prosomal segmentation, compared to its function in the SAZ, as well as a broader understanding of the evolution and expression of Sox genes in other arachnids.

To address this, we examined the expression and role of *Sox21b-1* and *Delta* in more detail during development in *P. tepidariorum*. In addition, we explored the evolution and roles of Sox genes in arachnid development further by characterising the repertoire of Sox genes in five additional arachnids and compared the expression of Sox genes between the harvestman *Phalangium opilio* and the spider *P. tepidariorum*.

Materials and Methods

Embryo collection, fixation and staging

P. tepidariorum embryos were collected from adult females from the laboratory culture (Goettingen strain) at Oxford Brookes University, which is kept at 25°C with a 12 hour light-dark cycle. Embryos were staged according to Mittmann and Wolf (2012) and fixed as described in Akiyama-Oda and Oda (2003). Embryos were collected and stored in RNA later from captive mated females of the amblypygids *Charinus acosta,* at one day, one month and two months after the appearance of the egg sacs, and *Euphrynichus bacillifer,* at approximately 30% of embryonic development (Harper et al., 2020). Mixed stage embryos were collected from a female *Pardosa amentata* (collected in Oxford) and *Marpissa muscosa* (kindly provided by Philip Steinhoff and Gabriele Uhl) and stored in RNA later (Harper et al., 2020). *Ph. opilio* embryos were collected and prepared as previously described (Sharma et al., 2012).

Transcriptomics

RNA was extracted from the embryos of *C. acosta, E. bacillifer, Pa. amentata* and *M. muscosa* using QIAzol according to the manufacturer's guidelines (QIAzol Lysis Reagent, Qiagen) (Harper et al., 2020). Illumina libraries were constructed using a TruSeq RNA sample preparation kit and sequenced using the Illumina NovaSeq platform (100 PE) by Edinburgh Genomics (https://genomics.ed.ac.uk). Raw reads quality was assessed using FastQC v0.11.9 (Andrews, 2010). Erroneous k-mers were removed using rCorrector (Song and Florea, 2015) and unfixable read pairs were discarded using a custom Python script (provided by Adam Freedman, available at https://github.com/harvardinformatics/TranscriptomeAssemblyTools/blob/master/FilterUncorrectabledPEfastq.py). Reads were trimmed for adapter contamination using TrimGalore! (available at https://github.com/FelixKrueger/TrimGalore) before *de novo* transcriptome assembly with Trinity (Haas et al., 2013). Transcriptome completeness was evaluated using BUSCO v4.0.2 (Seppey et al., 2019) along with the arachnid database (arachnida_odb10 created on 2019-11-2; 10 species, 2934 BUSCOs) (Harper et al., 2020). The transcriptomes will be made available via an appropriate repository upon publication.

Identification of Sox genes

To identify the Sox gene repertoires of *C. acosta*, *E. bacillifer*, *M. muscosa*, *Pa. amentata*, *Centruroides sculpturatus*, *Ph. opilio*, *Ixodes scapularis* and *Strigamia maritima*, we performed a BLAST search (e-value 0.05) against the available genomic and transcriptomic resources (*C. acosta* – this study; *E. bacillifer* – this study; *M. muscosa* – this study; *Pa. amentata* – this study; *Ce. sculpturatus* – PRJNA422877; *Ph. opilio* – PRJNA236471; *I. scapularis* – PRJNA357111; *S. maritima* – PRJNA20501), using the HMG domain protein sequences previously identified in *P. tepidariorum*, *Stegodyphus mimosarum* and *D. melanogaster* (Paese et al., 2018b). Predicted protein sequences were obtained using the ORFfinder NCBI online tool (https://www.ncbi.nlm.nih.gov/orffinder/; default settings except in 'ORF start codon to use' setting, where the 'Any sense codon' option was used to retrieve gene fragments lacking a start codon). Longest obtained ORFs were annotated as the best hits obtained from the SMART BLAST NCBI online tool (https://blast.st-va.ncbi.nlm.nih.gov/blast/smartblast/; default settings) and by reciprocal BLAST against the proteome of *P. tepidariorum*. A list of Sox sequences identified in this study and other sequences used in the subsequent phylogenetic analysis is provided in Supplementary File 1.

Sequence identity was further verified through the construction of maximum likelihood trees (Figs S1, S2). Protein sequence alignments of HMG domains from *C. acosta*, *E. bacillifer*, *M. muscosa*, *Pa. amentata*, *St. mimosarum* (Paese et al., 2018b), *P. tepidariorum* (Paese et al., 2018b), *Ce. sculpturatus*, *Ph. opilio*, *I. scapularis*, *S. maritima*, *Glomeris marginata* (Janssen et al., 2018), *T. castaneum* (Janssen et al., 2018), *D. melanogaster* (Paese et al., 2018b) and *Euperipatoides kanangrensis* (Janssen et al., 2018) were generated in MEGA v.7 using the MUSCLE v.3 algorithm (default settings, (Kumar et al., 2018)). Phylogenetic analysis was performed using RAxML v.1.5b3 (Stamatakis, 2014) with an LG + Γ substitution model, with nodal support inferred using the rapid bootstrapping algorithm (1000 replicates; Stamatakis et al., (2008). Alignments (Phylip files) and gene trees (TRE files) are provided in Supplementary Files 2-5. The

resulting trees were visualised and processed in FigTree v1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/).

In situ hybridisation probe synthesis

Total RNA was extracted using QiAzol (Qiagen) from stage 1-14 *P. tepidariorum* embryos according to the manufacturer's protocol. Total RNA was then used to generate cDNA using the QuantiTect reverse transcription kit (Qiagen), according to manufacturer's guidelines. For *P. opilio*, total RNA was extracted from several clutches of stage 9-16 embryos using Trizol TRIreagent, following the manufacturer's protocol. cDNA was generated using the Superscript III First Strand cDNA kit (ThermoFisher) with oligo-dT amplification, following the manufacturer's protocol.

Gene-specific primers were designed using Primer3 (http://primer3.ut.ee), and T7 linker sequences were added to the 5' end of the forward primer (GGCCGCGG) and reverse primer (CCCGGGGC). A list of primer sequences is provided in Supplementary File 6. The template for probe synthesis was generated through two rounds of standard PCR method using One*Taq*® 2x Master Mix (New England Biolabs, NEB): the first PCRs from cDNA used the gene-specific primers including the T7 linker sequence. The resulting PCR product was purified using a standard PCR purification kit (NucleoSpin® Gel and PCR Clean-up kit, Macherey-Nagel) and used as a template for the second PCR that used the gene-specific forward primer and a 3' T7 universal reverse primer targeting the forward linker sequence for the antisense probe, and the gene-specific reverse primer and a 5' T7 universal reverse primer targeting the reverse linker sequence. The resulting PCR products were run on an agarose gel (1-2%), and the band with the expected size excised and purified using the NucleoSpin® Gel and PCR Clean-up kit (Macherey-Nagel). The second PCR products were sent for Sanger sequencing to Eurofins Genomics and checked for quality. RNA probe synthesis was performed using T7 polymerase (Roche) with either DIG RNA labelling mix (Roche) or Fluorescein RNA labelling mix (Roche), according to manufacturer's guidelines.

Colourimetric in situ hybridisation (ISH)

Colourimetric ISH was performed following the whole-mount protocol described in Prpic et al., (2008) with minor modifications: steps 4-8 were replaced by two 10 minute washes in PBS-Tween-20 (0.02%) (PBS-T), and at step 18 the embryos were incubated for 30 minutes. Post-fixation was followed by ethanol treatment to decrease background: embryos were incubated for 10 minutes in inactivation buffer (75 g glycine, 600 µl 1N HCl, 50 µl 10% Tween-20 and dH_2O to 10 mL), followed by three wash steps with PBS-T, washed 5 min in 50% ethanol in PBS-T, washed in 100% ethanol until background decreased, washed for 5 minutes in 50% ethanol in PBS-T and finally washed twice with PBS-T. Embryos were then counterstained with DAPI (1:2000; Roche) for ~20 minutes and stored in 80% glycerol in 1x PBS at 4°C. Imaging was performed using a Zeiss Axio Zoom V.16. DAPI overlays were generated in Photoshop CS6.

Double fluorescent in situ hybridisation

Double fluorescent in situ hybridisation protocol was modified from Clark and Akam (2016): fixed embryos were gradually moved from methanol to PBS-T and washed for 15 minutes. Embryos were then transferred to hybridization buffer, hybridized overnight at 65°C and washed post-hybridization as detailed in Prpic et al. (2008). Embryos were incubated in blocking solution (Roche) for 30 minutes and AP-conjugated anti-DIG (1:2000; Roche) and POD-conjugated anti-FITC (1:2000; Roche) added and incubated for two hours. Tyramide biotin amplification (TSA Plus Biotin Kit, Perkin Elmer) was performed for 10 minutes, followed by incubation for 90 minutes in streptavidin Alexa Fluor 488 conjugate (1:500; ThermoFisher Scientific). AP signal was visualised by Fast Red staining (1/2 Fast Red tablet; Kem En Tec Diagnostics). Counterstaining with DAPI (1:2000; Roche) was carried out for 5-10 minutes. Yolk granules were removed manually in PBS and germ bands were flat mounted on poly-L-lysine coated coverslips in 80% glycerol. Imaging was performed using a Zeiss LSM800 confocal with Airyscan. Images were processed using Adobe Photoshop CS6 and FIJI software.

Double stranded RNA preparation

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

Parental RNAi

Five virgin adult female spiders were injected per gene according to the protocol described in Akiyama-Oda and Oda (2006). Each spider was injected in the opisthosoma with 2 µl of dsRNA every 2 days, to a total of five injections. A male was added to the vial for mating after the second injection. Embryos from injected females were fixed at stages 5 to 8.2 as described above. Embryos from GFP-injected control females were generated and treated as described above. For *Pt-Sox21b-1* knockdown we used the same 549 bp dsRNA (fragment 1) as in our previous study (Paese et al., 2018a). The same range and frequencies of phenotypic classes were observed for all cocoons from injected females (Paese et al., 2018a). Since phenotypic class III embryos do not transition from radial to axial symmetry and display major developmental defects, expression analysis was only carried out on embryos of phenotypic classes I and II (Paese et al., 2018a).

Embryonic RNAi

Embryonic injections were carried out as described in Schönauer et al. (2016) with minor changes. Embryos were injected between the 8- and 16-cell stages with small quantities of injection mix made up of 5 µl of FITC-dextran, 5 µl of biotin-dextran and 2.5 µl of dsRNA. Embryos were subsequently fixed at stages 5 to 8.2 of development. Visualization of eRNAi clones was achieved by detecting the co-injected biotin-dextran with the Vectastain ABC-AP kit (Vector Laboratories) after ISH, according to the manufacturer's protocol.

Results

Regulation of prosomal segmentation genes by *Sox21b-1*

Knockdown of *Pt-Sox21b-1* expression inhibits the formation of all leg-bearing segments (Paese et al., 2018a). To further evaluate the effect of *Pt-Sox21b-1* knockdown on the regulation of anterior segmentation we characterised the expression patterns of *Pt-Dll* and *Pt-hb,* which have known roles in this process (Pechmann et al., 2011; Schwager et al., 2009), as well as homeobox gene *Pt-Msx1,* which is expressed in a segmental pattern in the prosoma and SAZ (Leite et al., 2018).

Pt-Dll is required for the development of L1 and L2 segments, and is also expressed in the SAZ (Pechmann et al., 2011) (Fig. 1). In *Pt-Sox21b-1* pRNAi embryos, at stage 8.1, the L1 stripe of *Pt-Dll* expression is somewhat weaker than in wild type embryos and the L2 expression stripe is not distinctive. This is consistent with the loss of L1 and L2 upon *Pt-Sox21b-1* knockdown and suggests this treatment prevents the splitting of *Pt-Dll* expression necessary for the formation of these segments. *Pt-Sox21b-1* knockdown also results in a reduction in *Pt-Dll* expression at the posterior of the germband at stage 7 (n=3) and subsequently this expression completely disappears (n=8), consistent with perturbed development of the SAZ in *Pt-Sox21b-1* pRNAi embryos (Fig. 1G-J').

Pt-hb regulates the development of the L1, L2 and L4 segments (Schwager et al., 2009). Several aspects of *Pt-hb* expression are disrupted by *Pt-Sox21b-1* pRNAi knockdown (Fig. 2), which is consistent with the loss of segments observed after this treatment (Paese et al., 2018a). In stage 7 *Pt-Sox21b-1* pRNAi embryos, stripes of *Pt-hb* expression corresponding to the presumptive precheliceral/pedipalpal and L1/L2 regions appear as normal, although the L4 domain appears to be perturbed in some embryos (n=7) (Fig. 2). Subsequently, while the pre-cheliceral domain and pedipalpal stripes of *Pt-hb* expression appear to develop normally, the L1/L2 domain does not split into segmental stripes, and the L3 stripe cannot be distinguished from L4, while in some embryos all tissue and expression posterior of L1 is lost (n=4) (Fig. 2). These results suggest that while *Pt-Sox21b-1* is not necessary for the activation of *Pt-hb* expression, it is required for development of stripes of *Pt-hb* expression in L1 to L4 and the formation of these segments.

Pt-Msx1 is still expressed in *Pt-Sox21b-1* RNAi embryos at stage 5, although due to the highly deformed state of the embryos analysed, it is difficult to tell whether the pattern is restricted to the presumptive L2-L4 segments or if it extends anteriorly in to the presumptive head and L1 regions (n=5) (Fig. 3). In early stage 7 *Pt-Sox21b-1* RNAi embryos, the anterior stripe in the presumptive head segments and the broad band of expression across the presumptive L2-L4 segments can still be detected (n=6) (Fig. 3F-H). However, at late stage 7, while the anterior stripe of *Pt-Msx1* expression appears to be unaffected, the band of expression across the presumptive L2-L4 segments does not split into segmental stripes (n=2) (Fig. 3H). It was not possible to determine whether *Pt-Msx1* expression in the SAZ depends on *Pt-Sox21b-1* because the tissue posterior to L4 is lost in these embryos (Fig. 3F-H). Taken together, these results are similar to the effect of *Pt-Sox21b-1* knockdown on *Pt-hb* expression and suggest that while *Pt-Sox21b-1* is not required for *Pt-Msx1* activation, it is necessary for splitting of the broad expression of this homeobox gene into segmental stripes in the prosoma either by directly inhibiting *Pt-Msx1* expression in some cells or perhaps indirectly by organising the prosomal cells into segments.

Expression of *Pt-Sox21b-1* with respect to other spider segmentation genes

To better understand the regulation of prosomal and opisthosomal segmentation in *P. tepidariorum* we mapped the expression of *Pt-Sox21b-1* together with *Delta, caudal* or *hairy* via dual colour ISH. At stage 6, *Pt-Sox21b-1* is expressed throughout the embryo overlapping with *Pt-Dl* expression in the anterior and posterior compartment of the SAZ and in the prosoma (Fig. 4A). *Pt-Dl* and *Pt-Sox21b-1* are also co-expressed in the prosoma at stage 7 (Fig. 4B-B''), but their expression dynamics are out of phase in the SAZ (Fig. 4B',B''). At stage 8.1 *Pt-Sox21b-1* is strongly expressed in the anterior SAZ (Fig. 4C'), ubiquitously in the prosoma and in a stripe domain in the forming head lobes (Fig. 4C,C'). At this stage *Pt-Dl* expression in the anterior SAZ and the cheliceral/pedipalpal region does not overlap with *Pt-Sox21b-1* expression (Fig. 4C). Overall these double in situs show that *Pt-Dl* and *Pt-Sox21b-1* may work together to specify prosomal segments but they appear to be out of phase with each other in the SAZ after stage 6 (Fig. 9).

In the case of *caudal,* at late stage 6, *Pt-Sox21b-1* and *Pt-cad* overlap in the SAZ (Fig. 5A',A'') with *Pt-cad* expression being particularly strong in the anterior region. In contrast to *Pt-Sox21b-1, Pt-cad* is not expressed in the prosoma during early stages. By stage 7 (Fig. 5B-B'') the anterior SAZ expression of *Pt*-cad appears even stronger and, as was the case with *Pt-Dl, Sox21b-*1 expression is absent from this domain. At stage 8.1 *Pt-cad* is expressed in the posterior SAZ, extending into the anterior region where it partially overlaps with *Pt-Sox21b-1* (Fig. 5C,C''). *Pt-cad* is also expressed in lateral regions of the anterior SAZ (Fig. 5C''), as well as in the mesoderm of the third walking leg segment, but these domains do not overlap with *Pt-Sox21b-1* expression (Fig. 5C). Taken together, the relative expression of *Pt-Sox21b-1* and *Pt-cad* suggest these genes may contribute to defining the different regions of the SAZ, perhaps working together in some posterior cells but likely playing different roles in the anterior SAZ. Furthermore, the expression patterns suggest that *Pt-Sox21b-1* regulates prosomal segment formation independently of *Pt-cad*. These observations are also consistent with the relative expression and roles of *Pt-cad* and *Pt-Dl* in prosoma versus opisthosoma and different regions of the SAZ (Figs S3, 9).

We previously suggested that *Pt-hairy* (*Pt-h*) might be regulated by *Sox21b-1* because their expression patterns substantially overlap in the prosoma and *Pt-h* expression is lost in *Pt-Sox21b* pRNAi embryos (Paese et al., 2018a). Therefore, we next characterised the expression of *Pt-h* relative to *Pt-Dl* to help determine whether the regulation of *Pt-h* by *Pt-Sox21b-1* could be indirectly via *Pt-Dl.* At stage 5, *Pt-Dl* is expressed in a circular salt-and-pepper pattern in the centre of the germ disc and faintly along the rim (Fig. S4A'); *Pt-h* is also expressed in the centre of the germ disc, with some cells expressing both genes. At the rim of the germ disc *Pt-h* is more widely expressed and shows little overlap with *Pt-Dl* (Fig. S4A''). At stage 7 *Pt-Dl* and *Pt-h* both exhibit expression in the forming leg-bearing segments of the prosoma, where they partially overlap (Fig. S4B',B''). At stage 8.1 *Pt-Dl* expression can also be detected in the walking leg-bearing segments L2-L4 with some overlap, but mostly anterior of *Pt-h* expression (Fig. S4C-C''). These experiments suggest that there is some overlap in *P-Dl* and *Pt-h* expression in the prosoma, but the expression of these genes is out of phase in the SAZ, and by inference that *Pt-h* is expressed in the same cells as *Pt-Sox21b-1* in the prosoma and SAZ (Fig. 9).

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

Since pRNAi mediated knockdown of *Pt-Sox21b-1* results in severe effects during early embryogenesis it is difficult to ascertain direct effects of *Pt-Sox21b-1* on segmentation gene expression. To address this problem, we used eRNAi to induce *Pt-Sox21b-1* knockdown in small subsets of embryonic cells to analyse more specific and local effects of *Pt-Sox21b-1* knockdown.

We first verified the effectiveness of *Pt-Sox21b-1* eRNAi by performing ISH for *Pt-Sox21b-1* in the injected embryos (Fig. 6). We observed that *Pt-Sox21b-1* expression was reduced in the cells of all clones obtained (n=12) (Fig. 6). In some embryos, the effect of eRNAi on *Pt-Sox21b-1* expression appeared to extend to cells adjacent to the clone region (n=11) (Fig. 6E-G); a non-autonomous effect also previously reported for eRNAi with other genes (Kanayama et al., 2011; Schönauer et al., 2016). At stage 5, the region affected by the *Pt-Sox21b-1* eRNAi clone appears to have a lower cell density when compared to the remaining area of the germ disc, albeit based on a single observation (n=1), but consistent with smaller germ discs of *Pt-Sox21b-1* pRNAi embryos (Paese et al., 2018a). Interestingly, at stages 7 (n=3) and 8.1 (n=7), clones in the presumptive leg segments appeared to cause a distortion of the germ band along the AP axis that extended posteriorly into the opisthosoma as a consequence of fewer cells in the affected area compared to regions adjacent to the clone (Fig. 6E-H). These results again highlight that the effect of *Pt-Sox21b-1* loss on cells should be considered when interpreting the effects of *Pt-Sox21b-1* knockdown on the expression of other segmentation genes.

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

Pt-Sox-21b-1 and *Pt-Dl* are both crucial for many aspects of early embryogenesis and segmentation in *P. tepidariorum* (Oda et al., 2007; Paese et al., 2018a; Schönauer et al., 2016). Our results above suggest that there is overlap in their expression in the prosoma and SAZ before stage 6 (Fig. 4). Furthermore, we previously observed that *Pt-Sox21b-1* pRNAi embryos lack *Pt-Dl* expression, implying that *Pt-Sox21b-1* functions upstream of *Pt-Dl* (Paese et al., 2018a). To understand how segmentation is regulated in *P. tepidariorum* it is important to better understand

the interactions between *Pt-Sox-21b-1* and *Pt-Dl*, therefore, we carried out ISH for *Pt-Dl* in embryos with *Pt-Sox21b-1* eRNAi clones.

In stage 5 embryos, the salt-and-pepper domain of *Pt-Dl* at the centre of the germ disc appears to be normal in *Pt-Sox21b-1* knockdown clones (n=2) (Fig. 7); however, we were unable to obtain any clones that definitely overlapped with this domain. We obtained several clones that overlapped with the domain at the rim of the germ disc (n=6), which consistently showed an increase in cells expressing *Pt-Dl* in this region (Fig. 7D,E). At stage 8.2, *Pt-Dl* expression in *Pt-Sox21b-1* knockdown clones in the head region appeared normal (n=2) (Fig. 7I). However, we recovered a knockdown clone in the anlage of the leg-bearing segments which resulted in tissue deformation and an apparent coalescence of *Pt-Dl*-expressing L2-L4 stripes (n=1) (Fig. 7H). In the same embryo, the anterior-most stripe of *Pt-Dl* expression is still detected and may have been stronger than normal but it appeared warped, probably due to the deformation of the germ band (Fig. 7H).

These results suggest that the early *Pt-Dl* expression in the anterior region of the germ disc expands in the absence of *Pt-Sox21b-1*, and while pedipalpal expression appears unaffected, correct formation of L2-L4 stripes of *Pt-Dl* expression requires *Pt-Sox21b-1*. It remains inconclusive whether the posterior domains of *Pt-Dl* expression are regulated by *Pt-Sox21b-1* because we were unable to obtain clones in this region.

***Delta* regulates *hairy* expression in the developing prosoma**

We next examined effects of *Pt-Dl* eRNAi on *Pt-h* expression to test if regulation of *Pt-h* expression by *Pt-Sox21b-1* is indirectly via *Pt-Dl*. In the stage 5 embryo at the rim of the germ disc (Fig. 8A), where *Pt-Dl* and *Pt-h* expression normally overlaps (Fig. S4A-A''), we observed a loss of *Pt-h* expression in *Pt-Dl* clones (n=2) (Fig. 8A'). At stage 7 embryos with *Pt-Dl* clones extending from the anterior of the germ band into the prosoma (Fig. 8B'), a region where *Pt-Dl* and *Pt-h* expression overlaps in the wild type (Fig. S4B-B''), we observed loss of *Pt-h* expression in the prosoma (n=7) (Fig. 8B') but normal *Pt-h* expression in the head lobes. At stage 8.1, when *Pt-Dl* and *Pt-h* are co-expressed in the prosomal segments L2-L4 (Fig. S4C-C''), *Pt-h* expression is lost in a *Pt-Dl*

knockdown clones that spans all three segments (n=3) (Fig. 8C'). Combined, these results suggest that *Pt-Sox21b-1* regulates *Pt-Dl* in the developing prosoma of the spider and that in turn *Pt-Dl* regulates *Pt-h* (Fig. 9).

Arachnid Sox gene repertoires

To further investigate the evolution of *Sox21b* genes and the Sox gene repertoires of arachnids compared to other arthropods more broadly, we surveyed new and available genomic resources to identify Sox HMG-domain containing sequences from several additional arachnids: the spiders *M. muscosa* and *Pa. amentata*., the amblypygids *C. acosta* and *E. bacillifer*, the scorpion *Ce. sculpturatus*, the harvestmen *Ph. opilio* and the tick *I. scapularis*, and a myriapod, the centipede *S. maritima*. Sequences were annotated as the best hits obtained from the SMART BLAST NCBI online tool and by reciprocal BLAST against the proteome of *P. tepidariorum* and compared to previously identified arthropod Sox genes (Janssen et al., 2018; Paese et al., 2018b).

Our previous analysis of the *P. tepidariorum* genome identified fifteen Sox genes (Paese et al., 2018b) (Fig. 10). We identified fourteen Sox genes in *Pa. amentata*, fourteen in *M. muscosa*, fourteen in *C. acosta,* fourteen in *E. bacillifer*, nineteen in *Ce. sculpturatus*, seven in *Ph. opilio*, eleven in *I. scapularis* and twelve in *S. maritima* (Fig. 10; Supplementary File 1). However, a few of the sequences retrieved had incomplete HMG domains (10/105), probably representing fragments of the full sequence (Supplementary File 2). Therefore, the Sox repertoires described here may be incomplete due to deficient or poor sequence data, because Sox gene loss is very rare, especially in the arthropods (Phochanukul and Russell, 2010).

The sequences we obtained were assigned to particular Sox groups based on the best match from reciprocal BLAST and verified using maximum likelihood trees (Figs S1 and S2). The trees supported the classification of most sequences obtained, forming monophyletic clades corresponding to each Sox group (Figs S1 and S2). When constructing trees with all HMG sequences the bootstrap support values were low, as previously found with this domain (Paese et al., 2018b). However, removing *Ptep-SoxB-like*, *Cscu-SoxB-like*, *Isca-SoxB-like*, *Smar-SoxB-like*, *Smar-SoxE-like* and *Tcas-SoxB5*, as well as *Gmar-Sox21a* (which has an incomplete HMG domain

sequence) gave better support values for Sox Group B (81) and Sox Group E (80), with Sox Groups D and F still being very well supported (100 and 98 respectively) (Fig. S2). The only exception was Sox Group C genes (34), possibly due to considerable sequence divergence within this group. The monophyly of Sox Groups E and F was also well supported (80), and within Sox Group B, there was strong support for the monophyly of *SoxN* sequences (79), as well as reasonable support for the monophyly of arachnid *Dichaete* sequences (61) and arachnid *Sox21b* sequences (63) (Fig. S2). Together, these analyses are concordant with the phylogenetic relationships for the Sox family previously established using vertebrate and invertebrate Sox genes (Bowles et al., 2000; McKimmie et al., 2005; Wilson and Dearden, 2008).

At least one representative of each Sox group was found for all species surveyed, with the exception of *Sox21a* and *SoxF* in *Ph. opilio* (Fig. 10). Consistent with previous surveys (Janssen et al., 2018; Paese et al., 2018b), a single copy of *Dichaete* was found in all species analysed (Fig. 10). A single copy of *SoxN* was also found in most species, with the exception of *Ce. sculpturatus* where two sequences with identical HMG domains were found, although they differ outside this domain (Fig. 10). A single copy of *Sox21a* is present in non-WGD arthropods, whereas two copies were found in all seven arachnopulmonate species analysed (Fig. 10). Additional *SoxB-like* sequences were also found in *Ce. sculpturatus*, *I. scapularis* and *S. maritima*, although, much like other previously identified *SoxB-like* genes (i.e. *Tcas-SoxB5* and *Ekan-SoxB3*) (Janssen et al., 2018; Paese et al., 2018b), these sequences diverge significantly from other Group B genes and therefore cannot be reliably classified (Fig. 10 and Supplementary File 2).

All non-WGD arthropods (insects, myriapod, harvestman and tick) have single copies of *SoxC* and *SoxD*, but in arachnopulmonates, at least two copies for each of these groups were found (Fig. 10). At least two *SoxE* and *SoxF* genes were also found in all arachnopulmonates except for *St. mimosarum,* where *SoxF* is only represented by one gene, and *Ph. opilio,* where *SoxF* is potentially missing (Fig. 10). *SoxE* also appears to be duplicated in *G. marginata*, *S. maritima* and *I. scapularis*, although *Smar-SoxE2* has an incomplete domain, and the two copies in *I. scapularis* contain non-overlapping fragments of the HMG domain (indicating they might be the same gene) (Fig. 10 and Supplementary File 2). An additional *SoxE-like* sequence was found in *S.*

maritima, although this sequence is highly divergent from other *SoxE* genes (Fig. 10 and Supplementary File 2). We note that *SoxE* duplication may be relatively common in invertebrate lineages compared to other Sox families, with duplications previously identified in Hymenoptera (Wilson and Dearden, 2008). Two *SoxF* genes were also found in *S. maritima* but they have incomplete HMG domains, and so may represent fragments of a single gene (Fig. 10 and Supplementary File 2).

Finally, our survey revealed an interesting evolutionary pattern for *Sox21b*. It appears that among the species we have surveyed this Sox gene is only duplicated in *P. tepidariorum* (Fig. 10). This suggests that after the arachnopulmonate WGD both ohnologs of *Sox21b* were only retained in a restricted lineage including *P. tepidariorum* or that this Sox gene was tandemly duplicated later in this spider, which raises many interesting questions regarding the evolution of its role during segmentation in arachnids.

Expression of Sox genes during embryogenesis in *P. tepidariorum* and *P. opilio*

To help decipher whether *Sox21b-1* played an ancestral role in arachnid segmentation or evolved this role only in the lineage leading to *P. tepidariorum*, we compared the expression patterns of Sox genes in this spider with their orthologs in *Ph. opilio*. Although we previously characterised Sox gene expression patterns in *P. tepidariorum*, we only detected expression for six out of 15 Sox genes present in this species (*Pt-SoxN*, *Pt-Sox21b-1*, *Pt-SoxC1*, *Pt-SoxD1*, *Pt-SoxE1* and *Pt-SoxF2*) (Paese et al., 2018b). However, it remained likely that the other genes were expressed at least at low levels during embryogenesis because their transcripts could be detected in RNA-Seq data (Iwasaki-Yokozawa et al., 2018). We therefore performed additional analyses using longer ISH probes at a wider range of embryonic stages.

The expression patterns observed for the *P. tepidariorum* Sox genes were generally the same as those we reported previously (Paese et al., 2018b) (Figs S5-S11). However, we were now able to detect expression of *Pt-Sox21a-1* and *Pt-Sox21a-2* in non-overlapping patterns mainly in the neuroectoderm, which is suggestive of sub- and/or neofunctionalisation (Fig. S7). We also observed *Pt-SoxE2* expression in a similar pattern to *Pt-SoxE1* in the prosomal and opisthosomal

limb buds, albeit with some temporal and spatial differences, as well as unique expression of *Pt-SoxE2* in opisthosomal cells that are possibly the germline progenitors (Schwager et al., 2015a) (Fig. S10). We also found that while *Pt-SoxF1* and *Pt-SoxF2* are both expressed in the developing prosomal and opisthosomal appendages; each has also a specific domain of expression: *Pt-SoxF1* is expressed in the secondary eye primordia (Fig. S11), and *Pt-SoxF2* is expressed along the dorsal border of the prosomal segments and along the edge of the non-neurogenic ectoderm (Fig. S11).

We were able to detect expression for four of the seven identified *Ph. opilio* Sox genes (*SoxN*, *Sox 21b*, *SoxC*, and *SoxD*). For the three other *Ph. opilio* Sox genes we either failed to obtain a signal, presumably because the short sequences obtained made poor probes (*Dichaete*), or we were unable to amplify the fragment using PCR (*SoxE-1* and *SoxE2*).

Po-SoxN and *Pt-SoxN* are strongly expressed in the developing neuroectoderm consistent with our previous analysis of *Pt-SoxN* expression (Paese et al., 2018b), as well as at the tips and base of the prosomal appendages (Fig. S6). *Pt-SoxC1* and *Po-SoxC* are also expressed in the developing neuroectoderm as well as in the prosomal appendages at later stages of development, although *Pt-SoxC1* is also expressed in a few cells where *Po-SoxC* is not detected (Fig. S8). *Po-SoxD* and *Pt-SoxD2* also exhibit similar expression in the pre-cheliceral region, prosomal appendages and as dorsally restricted segmental stripes (Fig. S9), but an additional domain of *Pt-SoxD2* expression was detected in the opisthosomal appendages (Fig. S9).

With respect to understanding the evolution and role of *Sox21b* in segmentation, we found that like *Pt-Sox21b-1, Po-Sox21b* is also expressed in a segmental pattern during early stages as well as later in the neuroectoderm (Fig. 11). We were also able to detect expression for *Pt-Sox21b-2* in the last two opisthosomal segments (Fig. 11) as well as a faint signal in the developing neuroectoderm, (Fig. 11). These results suggest that *Sox21b* played a role in segmentation ancestrally in arachnids and that there may have been subfunctionalisation of *Sox21b* function after duplication in *P. tepidariorum*.

Intriguingly, although we did not detect expression of *Dichaete* or find a *Sox21a* gene in *Ph. opilio*, we found that *Pt-D* and *Pt-Sox21a-1* are also both expressed in patterns consistent with

roles in segmentation in *P. tepidariorum* (Figs S5 and S7). Therefore, it remains possible that other Sox genes in addition to *Sox21b* are involved in segmentation in spiders and other arachnids.

Discussion

Pt-Sox21b-1 is used in different GRNs underlying prosomal and opisthosomal segmentation

We previously showed that *Pt-Sox21b-1* is required for the simultaneous formation of the leg-bearing prosomal segments and the sequential production of opisthosomal segments in *P. tepidariorum* (Paese et al., 2018a). Here we show that while *Pt-Sox21b-1* appears not to be necessary for the activation of *Pt-Dll*, *Pt-hb* and *Pt-Msx1*, it is needed for the separation of the broad domains of their expression into segmental stripes and for the viability of cells contributing to the prosomal segments. This suggests that *Sox21b-1* does not directly regulate *Pt-Dll*, *Pt-hb* and *Pt-Msx1* during prosomal segmentation but perhaps acts in parallel to regulate the division and viability of cells and organise them into segments (Fig. 9). We suggest that these aspects of *Pt-Sox21b-1* function in the formation of the limb-bearing segments likely involves at least indirect regulation of aspects of *Pt-Dl* expression in this tagma, which in turn is required for expression of *Pt-h* in these segments. This is consistent with the loss of prosomal segments observed following *Pt-Sox21b-1* RNAi and the disruption of these segments in *Pt-Dl* RNAi embryos (Oda et al., 2007). However, further work is needed to decipher how *Pt-Dl* and *Pt-h* regulation of prosomal segments is integrated with the gap-like functions of *Pt-hb* and *Pt-Dll,* as well as other factors involved in cell organisation in the germband such as *Toll* genes (Benton et al., 2016; Pechmann et al., 2011; Schwager et al., 2009) (Fig. 9).

Pt-Sox21b-1 is also required for the formation of the SAZ and the subsequent addition of opisthosomal segments. We further characterised the expression of *Pt-Sox21b-1* with respect to other genes involved in the addition of opisthosomal segments and found that *Pt-Sox-21b* expression overlaps with *Pt-cad*, and by inference *Pt-h*, in some SAZ cells, but that its dynamic expression is out of phase with *Pt-Dl* (Fig. 9). The dynamic expression of *Pt-Sox21b-1* in the SAZ and its potential interactions with *Pt-cad* contrasts with the role of *Pt-Sox21b-1* in the prosoma and appears to be more consistent with the role of SoxB genes in insect segmentation (Clark and Peel, 2018). However, we failed to recover clones of *Pt-Sox21b-1* eRNAi knockdown in posterior cells and therefore the function and interactions of this gene in the SAZ requires further investigation.

Comparison of the expression of *Sox21b* gene expression between *P. tepidariorum* and *Ph. opilio* indicates that these genes likely regulated segmentation in the ancestor of arachnids rather than deriving this role after WGD or tandem duplication in this spider. Our results also suggest that *Pt-Sox21a* and *Pt-D* play roles in segmentation in *P. tepidariorum*, which in the case of the latter gene would be consistent and perhaps conserved with insects (Clark and Peel, 2018). Although note that previous RNAi knockdown of *Pt-Sox21a* and *Pt-D* did not appear to perturb segmentation in *P. tepidariorum* (Paese, 2018), further investigation of these genes and their potential role in segmentation is required in this spider and other arachnids.

The expanded Sox repertoire of arachnopulmonates

We previously reported duplication of several Sox gene families in the spider *P. tepidariorum* and our new analysis provides a wider perspective on the evolution of the Sox genes in arachnids. We find that all arachnopulmonates we analysed have at least two copies of each Sox gene family except for the SoxB genes *Dichaete*, *SoxN* (with the exception of *C. sculpturatus*) and *Sox21b* (with the exception of *P. tepidariorum*). This is consistent with the retention of Sox21a, SoxC, SoxD, SoxE and SoxF ohnologs after the WGD in the ancestor of arachnopulmonates that was not shared with harvestmen and ticks. We suggest that there may have been subsequent lineage-specific duplication of *SoxN* in *C. sculpturatus,* and *Sox21b* in *P. tepidariorum* (see above) although we cannot exclude that these are true ohnologs and only one was retained in other lineages. Unfortunately, previous analysis of the synteny of *Pt-Sox21b-1* and *Pt-Sox21b-2* was inconclusive with respect to whether they had arisen by WGD or a more recent tandem duplication (Paese et al., 2018b). Our survey also suggests there have been lineage-specific duplications of *SoxC* and *SoxD* genes in the spiders *P. amantata* and *M. muscosa*, respectively, as well as *SoxE* in Amblypygi and perhaps other families in *C. sculpturatus* (Fig. 9). This Sox gene retention in arachnids after WGD broadly parallels the retention of ohnologs of these genes after WGD in vertebrates (Schepers et al., 2002; Voldoire et al., 2017) and further indicates similar genomic outcomes following these independent WGD events (Leite et al., 2018; Schwager et al., 2017). Interestingly, the main fate of retained duplicated Sox genes in vertebrates appears to be subfunctionalisation, although there are

likely cases of neofunctionalization, for example in teleosts (Cresko et al., 2003; De Martino et al., 2000; Klüver et al., 2005; Voldoire et al., 2017).

To evaluate the fates of retained spider Sox ohnologs we compared their expression during embryogenesis to their single-copy homologs in the harvestman *Ph. opilio*. Our comparison of Sox gene expression patterns provides evidence for broad conservation between the spider *P. tepidariorum* and the harvestman *Ph. opilio*, consistent with the roles of these genes in other arthropods. However, differences in the expression of *P. tepidariorum* duplicates with respect to their expression in *Ph. opilio* could represent cases of sub- and/or neofunctionalization, but this requires verification by comparing the expression of Sox genes among other arachnids with and without an ancestral WGD.

Declaration of Interest

The authors declare no competing interests.

Author contributions

LBG, AS, SR, LSR, PPS and APM designed this project; LBG, AS, AH, GB, MS, SA, and LSR performed the experimental work; LBG, AS, SR, AH, SA, LSR, PPS and APM analysed data; and LBG, AS and APM wrote the paper with the help of all other authors.

Acknowledgements

This study was funded by a Leverhulme Trust grant (RPG-2016-234) to APM and AS, a Nigel Groome Studentship from Oxford Brookes University to LBG, a BBSRC DTP studentship to AH, a John Fell Fund grant (0005632) from the University of Oxford to LSR and a NSF CAREER IOS-1552610 to PPS. We thank Philip Steinhoff and Gabriele Uhl for kindly providing *M. muscosa* embryos.

References

Akiyama-Oda, Y. and 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**, 1735–1747.

Akiyama-Oda, Y. and 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**, 2347–2357.

Andrews, S. (2010). FastQC. *Babraham Bioinforma.*

Benton, M. A., Pechmann, M., Frey, N., Stappert, D., Conrads, K. H., Chen, Y. T., Stamataki, E., Pavlopoulos, A. and Roth, S. (2016). Toll Genes Have an Ancestral Role in Axis Elongation. *Curr. Biol.* **26**, 1609–1615.

Bowles, J., Schepers, G. and Koopman, P. (2000). Phylogeny of the SOX family of developmental transcription factors based on sequence and structural indicators. *Dev. Biol.* **227**, 239–255.

Chipman, A. D. (2010). Parallel evolution of segmentation by co-option of ancestral gene regulatory networks. *BioEssays* **32**, 60–70.

Clark, E. and Akam, M. (2016). Odd-paired controls frequency doubling in Drosophila segmentation by altering the pair-rule gene regulatory network. *Elife* **5**, e18215.

Clark, E. and Peel, A. D. (2018). Evidence for the temporal regulation of insect segmentation by a conserved sequence of transcription factors. *Dev.* **145**, 1–15.

Clark, E., Peel, A. D. and Akam, M. (2019). Arthropod segmentation. *Dev.* **146**, dev170480.

Couso, J. P. (2009). Segmentation, metamerism and the Cambrian explosion. *Int. J. Dev. Biol.* **53**, 1305–1316.

Cresko, W. A., Yan, Y. L., Baltrus, D. A., Amores, A., Singer, A., Rodríguez-Marí, A. and Postlethwait, J. H. (2003). Genome Duplication, Subfunction Partitioning, and Lineage Divergence: Sox9 in Stickleback and Zebrafish. *Dev. Dyn.* **228**, 480–489.

Damen, W. G. M. (2007). Evolutionary conservation and divergence of the segmentation process in arthropods. *Dev. Dyn.* **236**, 1379–1391.

Davis, G. K. and Patel, N. H. (1999). The origin and evolution of segmentation. *Trends Biochem. Sci.* **9**, M68-72.

Davis, G. K. and Patel, N. H. (2002). Short, Long, and Beyond: Molecular and Embryological Approaches to Insect Segmentation. *Annu. Rev. Entomol.* **47**, 669–699.

De Martino, S., Yan, Y. L., Jowett, T., Postlethwait, J. H., Varga, Z. M., Ashworth, A. and Austin, C. A. (2000). Expression of sox11 gene duplicates in zebrafish suggests the reciprocal loss of ancestral gene expression patterns in development. *Dev. Dyn.* **217**, 279–292.

Garb, J. E., Sharma, P. P. and Ayoub, N. A. (2018). Recent progress and prospects for advancing arachnid genomics. *Curr. Opin. Insect Sci.* **25**, 51–57.

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., Couger, M. B., Eccles, D., Li, B., Lieber, M., et al. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* **8**, 1494–1512.

Hannibal, R. L. and Patel, N. H. (2013). What is a segment? *Evodevo* **4**, 35.

Harper, A., Baudouin Gonzalez, L., Schönauer, A., Seiter, M., Holzem, M., Arif., S., McGregor, A.P. and Sumner-Rooney, L. (2020). Widespread retention of ohnologs in key developmental gene families following whole genome duplication in arachnopulmonates. *bioRxiv*.

Hilbrant, M., Damen, W. G. M. and McGregor, A. P. (2012). Evolutionary crossroads in developmental biology: the spider Parasteatoda tepidariorum. *Development* **139**, 2655–2662.

Iwasaki-Yokozawa, S., Akiyama-Oda, Y. and Oda, H. (2018). Genome-scale embryonic developmental profile of gene expression in the common house spider Parasteatoda tepidariorum. *Data Br.* **19**, 865–867.

Janssen, R., Le Gouar, M., Pechmann, M., Poulin, F., Bolognesi, R., Schwager, E. E., Hopfen, C., Colbourne, J. K., Budd, G. E., Brown, S. J., 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., Andersson, E., Betnér, E., Bijl, S., Fowler, W., Höök, L., Leyhr, J., Mannelqvist, A., Panara, V., Smith, K., et al. (2018). Embryonic expression patterns and phylogenetic analysis of panarthropod sox genes: insight into nervous system development, segmentation and gonadogenesis. *BMC Evol. Biol.* **18**, 88.

Kanayama, M., Akiyama-Oda, Y., Nishimura, O., Tarui, H., Agata, K. and Oda, H. (2011). Travelling and splitting of a wave of hedgehog expression involved in spider-head segmentation. *Nat. Commun.* **2**, 1–7.

Klüver, N., Kondo, M., Herpin, A., Mitani, H. and Schartl, M. (2005). Divergent expression patterns of Sox9 duplicates in teleosts indicate a lineage specific subfunctionalization. *Dev. Genes Evol.* **215**, 297–305.

Krause, G. (1939). Die Eitypen der Insekten. *Biol. Zentralbl.* **59**, 4950536.

Kumar, S., Stecher, G., Li, M., Knyaz, C. and Tamura, K. (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **35**, 1547–1549.

Leite, D. J. and McGregor, A. P. (2016). Arthropod evolution and development: recent insights from chelicerates and myriapods. *Curr. Opin. Genet. Dev.* **39**, 93–100.

Leite, D. J., Baudouin-Gonzalez, L., Iwasaki-Yokozawa, S., Lozano-Fernandez, J., Turetzek, N., Akiyama-Oda, Y., Prpic, N. M., Pisani, D., Oda, H., Sharma, P. P., et al. (2018). Homeobox gene duplication and divergence in arachnids. *Mol. Biol. Evol.* **35**, 2240–2253.

Liu, P. Z. and Kaufman, T. C. (2005). Short and long germ segmentation: Unanswered questions in the evolution of a developmental mode. *Evol. Dev.* **7**, 629–646.

McGregor, A. P., Pechmann, M., Schwager, E. E., Feitosa, N. M., Kruck, S., Aranda, M. and Damen, W. G. M. (2008). Wnt8 Is Required for Growth-Zone Establishment and Development of Opisthosomal Segments in a Spider. *Curr. Biol.* **18**, 1619–1623.

McKimmie, C., Woerfel, G. and Russell, S. (2005). Conserved genomic organisation of Group B Sox genes in insects. *BMC Genet.* **6**, 26.

Mittmann, B. and 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**, 189–216.

Oda, H. and Akiyama-Oda, Y. (2020). The common house spider Parasteatoda tepidariorum. *Evodevo* **11**, 6.

Oda, H., Nishimura, O., Hirao, Y., Tarui, H., Agata, K. and Akiyama-Oda, Y. (2007). Progressive activation of Delta-Notch signaling from around the blastopore is required to set up a functional caudal lobe in the spider Achaearanea tepidariorum. *Development* **134**, 2195–2205.

Paese, C. L. B. (2018). Investigating the roles of Hes and Sox genes during embryogenesis of the spider Parasteatoda tepidariorum.

Paese, C. L. B., Schoenauer, A., Leite, D. J., Russell, S. and McGregor, A. P. (2018a). A SoxB gene acts as an anterior gap gene and regulates posterior segment addition in a spider. *Elife* **7**, e37567.

Paese, C. L. B., Leite, D. J., Schönauer, A., McGregor, A. P. and Russell, S. (2018b). Duplication and expression of Sox genes in spiders. *BMC Evol. Biol.* **18**, 205.

Pechmann, M., McGregor, A. P., Schwager, E. E., Feitosa, N. M. and Damen, W. G. M. (2009). Dynamic gene expression is required for anterior regionalization in a spider. *Proc. Natl. Acad. Sci.* **106**, 1468–1472.

Pechmann, M., Khadjeh, S., Turetzek, N., McGregor, A. P., Damen, W. G. M. and Prpic, N. M. (2011). Novel function of Distal-less as a gap gene during spider segmentation. *PLoS Genet.* **7**, e1002342.

Peel, A. (2004). The evolution of arthropod segmentation mechanisms. *BioEssays* **26**, 1108–1116.

Phochanukul, N. and Russell, S. (2010). No backbone but lots of Sox: Invertebrate Sox genes. *Int. J. Biochem. Cell Biol.* **42**, 453–464.

Prpic, N.-M., Schoppmeier, M. and Damen, W. G. M. (2008). Whole-mount in situ hybridization of spider embryos. *CSH Protoc.* **2008**, pdb.prot5068.

Russell, S. R. H., Sanchez-Soriano, N., Wright, C. R. and Ashburner, M. (1996). The Dichaete gene of Drosophila melanogaster encodes a SOX-domain protein required for embryonic segmentation. *Development* **122**, 3669–3676.

Schepers, G. E., Teasdale, R. D. and Koopman, P. (2002). Twenty Pairs of Sox. *Dev. Cell* **3**, 167–170.

Schönauer, A., Paese, C. L. B., Hilbrant, M., Leite, D. J., Schwager, E. E., Feitosa, N. M., Eibner, C., Damen, W. G. M. and McGregor, A. P. (2016). The Wnt and Delta-Notch signalling pathways interact to direct pair-rule gene expression via *caudal* during segment addition in the spider *Parasteatoda tepidariorum*. *Development* **143**, 2455–2463.

Schwager, E. E., Pechmann, M., Feitosa, N. M., McGregor, A. P. and Damen, W. G. M. (2009). hunchback Functions as a Segmentation Gene in the Spider Achaearanea tepidariorum. *Curr. Biol.* **19**, 1333–1340.

Schwager, E. E., Meng, Y. and Extavour, C. G. (2015a). Vasa and piwi are required for mitotic integrity in early embryogenesis in the spider Parasteatoda tepidariorum. *Dev. Biol.* **402**, 276–290.

Schwager, E. E., Schönauer, A., Leite, D. J., Sharma, P. P. and McGregor, A. P. (2015b). Chelicerata. In *Evolutionary Developmental Biology of Invertebrates 3: Ecdysozoa I: Non-Tetraconata*, pp. 99–139. Springer-Verlag Vienna.

Schwager, E. E., Sharma, P. P., Clarke, T., Leite, D. J., Wierschin, T., Pechmann, M., Akiyama-Oda, Y., Esposito, L., Bechsgaard, J., Bilde, T., et al. (2017). The house spider genome reveals an ancient whole-genome duplication during arachnid evolution. *BMC Biol.* **15**, 62.

Seppey, M., Manni, M. and Zdobnov, E. M. (2019). BUSCO: Assessing genome assembly and annotation completeness. In *Methods in Molecular Biology*, pp. 227–245. Humana Press Inc.

Setton, E. V. W. and 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**, E3491–E3500.

Sharma, P. P., Schwager, E. E., Extavour, C. G. and Giribet, G. (2012). Hox gene expression in the harvestman Phalangium opilio reveals divergent patterning of the chelicerate opisthosoma. *Evol. Dev.* **14**, 450–463.

Song, L. and Florea, L. (2015). Rcorrector: Efficient and accurate error correction for Illumina RNA-seq reads. *Gigascience* **4**, 48.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313.

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

Tautz, D. (2004). Segmentation. *Dev. Cell* **7**, 301–312.

Voldoire, E., Brunet, F., Naville, M., Volff, J. N. and Galiana, D. (2017). Expansion by whole genome duplication and evolution of the sox gene family in teleost fish. *PLoS One* **12**, e0180936.

Wilson, M. J. and Dearden, P. K. (2008). Evolution of the insect Sox genes. *BMC Evol. Biol.* **8**, 120.

Figures and Legends

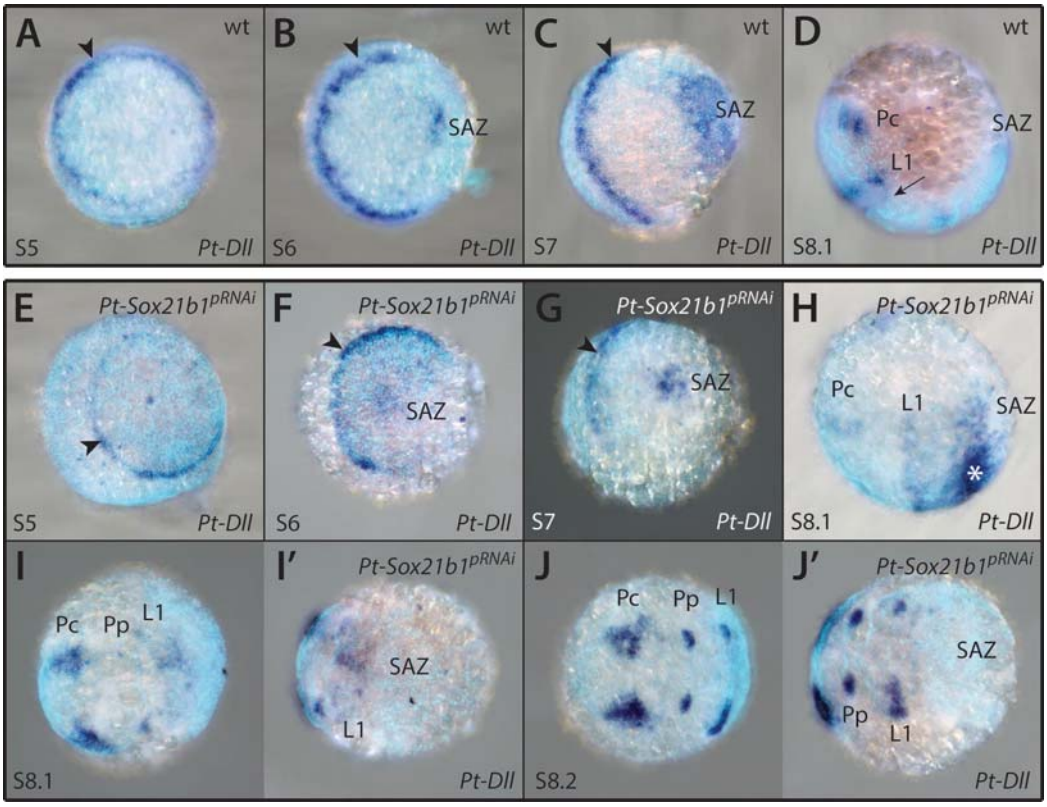

Figure 1. Effect of *Pt-Sox21b-1* pRNAi knockdown on *Pt-Dll* expression

Pt-Dll is expressed in the presumptive region of the L1 segment from as early as stage 5 (A; black arrowhead), first as a ring-like pattern which transforms into a stripe at the germ disc to germ band transition (B; black arrowhead). Additional expression domains in the SAZ and pre-cheliceral region arise at stages 6 (B) and 7 (D) respectively, as well as a faint stripe in the L2 segment (black arrow). (E-J') Expression pattern of *Pt-Dll* in *Pt-Sox21b-1* pRNAi embryos (detailed description in the text). Anterior is to the left in all images. Pc, pre-cheliceral region; Pp, presumptive pedipalpal segment; L1, presumptive L1 segment; SAZ, segment addition zone.

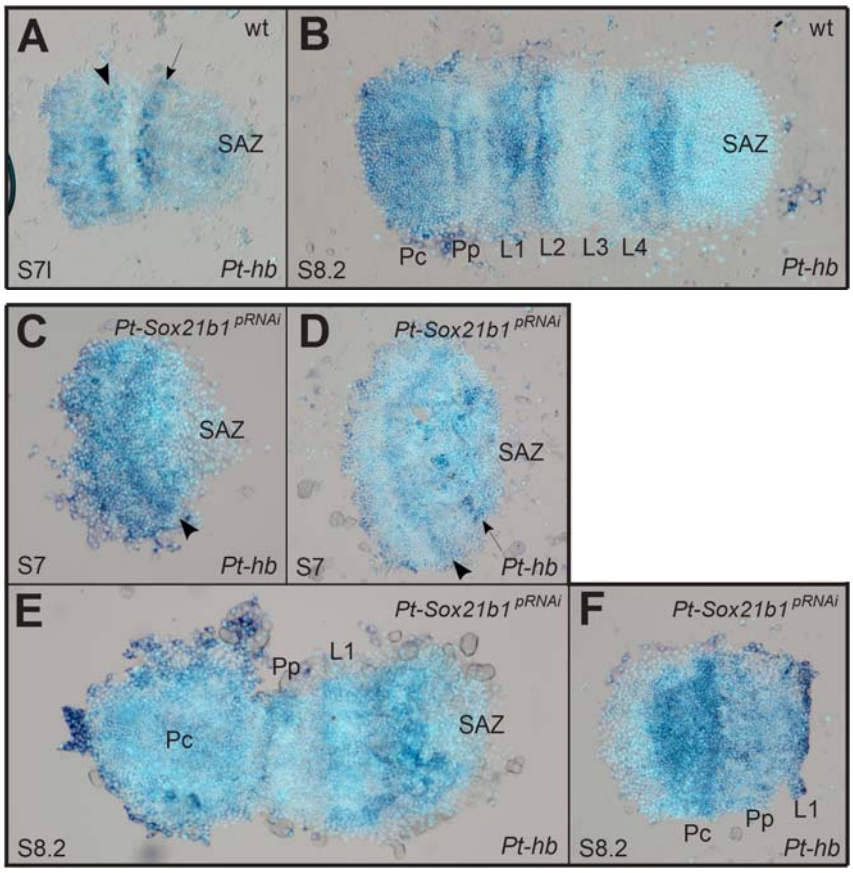

Figure 2. Effect of *Pt-Sox21b-1* pRNAi knockdown on *Pt-hb* expression

At late stage 7 (A), two bands corresponding to the head segments and L1+L2 segments (black arrowhead) are observed. Subsequently, by stage 8.2 (B) *Pt-hb* expression becomes segmentally restricted, with a domain covering the pre-cheliceral and cheliceral segments, and as stripes in all other anterior segments, with the strongest expression in the L1, L2 and L4 segments. Expression of *Pt-hb* in *Pt-Sox21b-1* pRNAi embryos (C-F) (detailed description in the text). Anterior is to the left in all images. Pc, pre-cheliceral region; Pp, presumptive pedipalpal segment; L1-L4, presumptive L1-L4 segments; SAZ, segment addition zone.

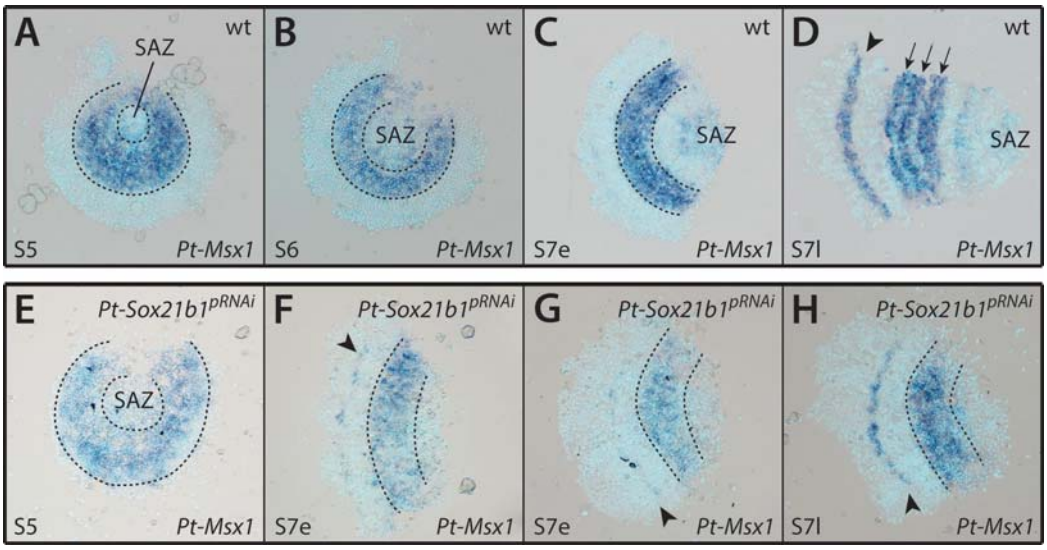

Figure 3. Effect of *Pt-Sox21b-1* pRNAi knockdown on *Pt-Msx1* expression

Pt-Msx1 is first expressed at stage 5 (A) in the presumptive region of the L2-L4 segments (dashed lines), forming a band of expression at stage 6 (B). At early stage 7 (C), two additional domains of expression arise, in the SAZ and in a faint stripe corresponding to the presumptive head segments. This anterior stripe of expression becomes stronger by late stage 7 (D; black arrowhead) and the L2-L4 domain starts splitting into three stripes corresponding to each leg segment (black arrows). By late stage 7, faint expression is still present in the SAZ and a faint stripe can be detected in the first opisthosomal segment. (E-H) Expression pattern of *Pt-Msx1* in *Pt-Sox21b-1* pRNAi embryos (detailed description in the text). Anterior is to the left in all images except A and B. All embryos are flat mounted; SAZ, segment addition zone.

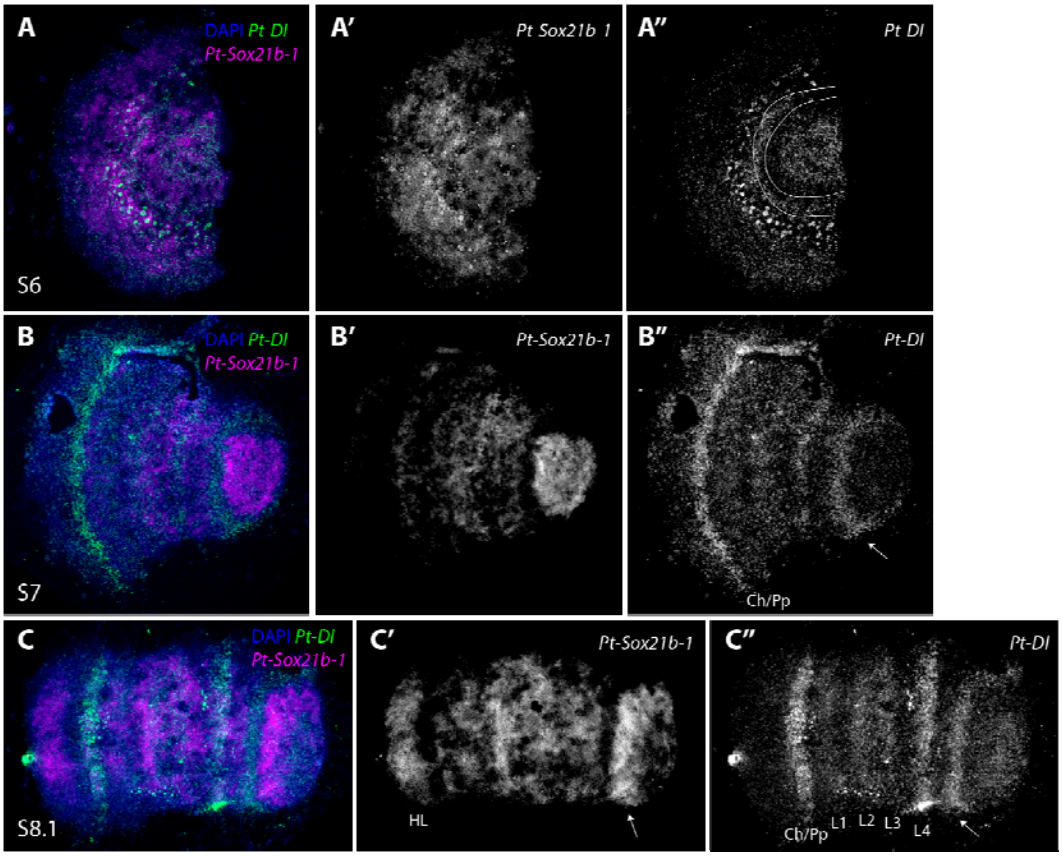

Figure 4. Expression of *P. tepidariorum Sox21b-1* and *Dl*

Double-fluorescent in situ of *Pt-Sox21b-1 and Pt-Dl* in stage 6 (A-A”), 7 (B-B”) and 8.1 (C-C”) embryos. At stage 6, *Pt-Sox21b-1* is expressed almost ubiquitously throughout the embryo (A’). *Pt-Dl* is expressed in a solid expression domain in the SAZ, a faint but solid expression domain in the anterior SAZ (white lines in A’’), adjacent to a salt-and-pepper domain in the forming prosoma (A’’). Expression of both genes at this stage largely overlaps (A). At stage 7 *Pt-Sox21b-1* is expressed strongly in the SAZ and throughout the forming prosoma (B’). *Pt-Dl* is also expressed in the anterior SAZ (white arrow in B’’), but this expression does not overlap with *Pt-Sox21-b1* (white arrow in B’’). *Pt-Dl* and *Pt-Sox21b-1* are co-expressed in the forming prosoma at stage 7 (B). *Pt-Dl* also exhibits a strong expression domain in the developing cheliceral-pedipalpal region (5 B’’). At stage 8.1 *Pt-Sox21b-1* is strongly expressed in the anterior SAZ (arrow in C’), ubiquitously in the prosoma and in a stripe domain in the forming head lobes (C,C’). At this stage *Pt-Dl* is expressed in the anterior SAZ (white arrow in C’’), strongly in L4 and weaker in the remaining leg-bearing segments and again strongly in the cheliceral/pedipalpal region. Anterior is to the left in all images. Ch, chelicerae; Pp, pedipalps; HL, head lobe; L1-L4, walking legs 1-4.

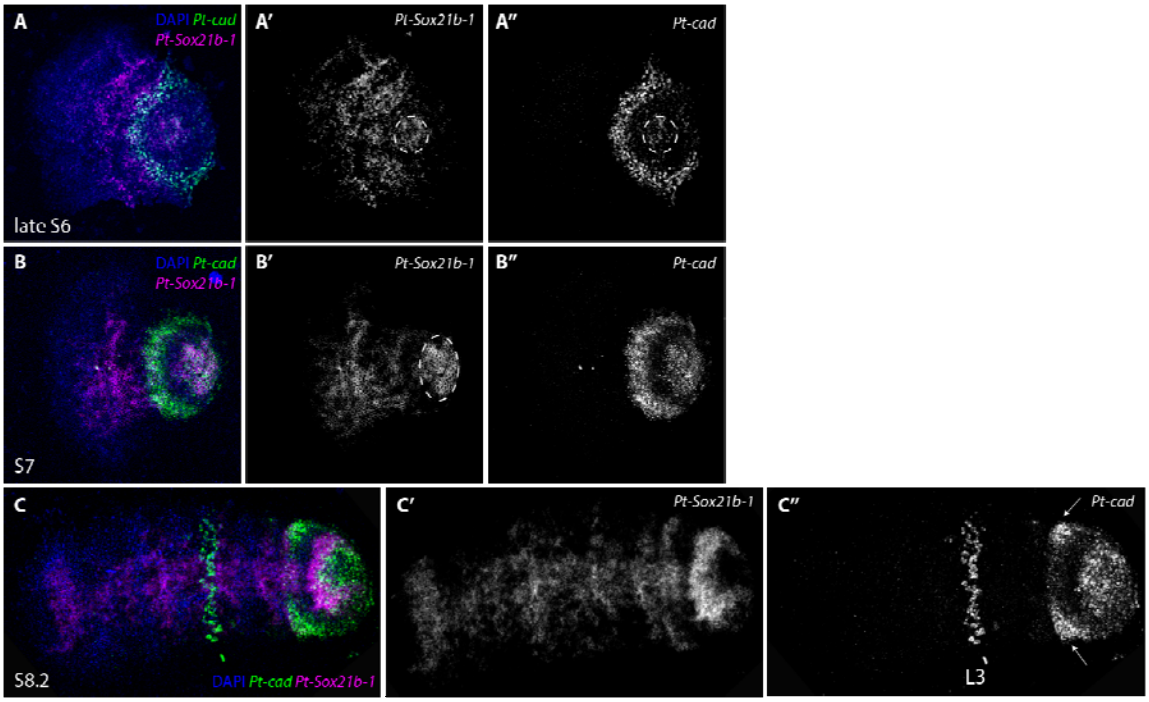

Figure 5. Double-fluorescent in situ of *P. tepidariorum Sox21b-1* and *cad*

Double-fluorescent in situ of *Pt-Sox21b-1 and Pt-cad* in stage 6 (A-A"), 7 (B-B") and 8.1 (C-C") embryos. (D-F"). At late stage 6 *Pt-Sox21b-1* is expressed in a faint circular domain in the posterior SAZ overlapping with *Pt-cad* (dashed circle in A''), although expression of these two genes doesn't overlap in the anterior SAZ (A,A''). *Pt-Sox21-b1* and *Pt-cad* are both expressed in a solid domain the posterior SAZ at stage 7 (B). *Pt-cad* is not expressed in the prosoma at stage 7 (B'), but a second strong expression domain is observed in the anterior SAZ, which does not overlap with *Pt-Sox21b-1* expression (B,B'). At stage 8.2 *Pt-cad* expression can be observed in the anterior and posterior SAZ, where it partial overlaps with *Pt-Sox21b-1* expression (C,C''). *Pt-cad* is also expressed in the lateral parts of the anterior SAZ (white arrows in C''), as well as in the mesoderm of the third walking leg segment but these domains do not overlap with *Pt-Sox21b-1* expression (C). Anterior is to the left in all images. L3, 3rd walking leg.

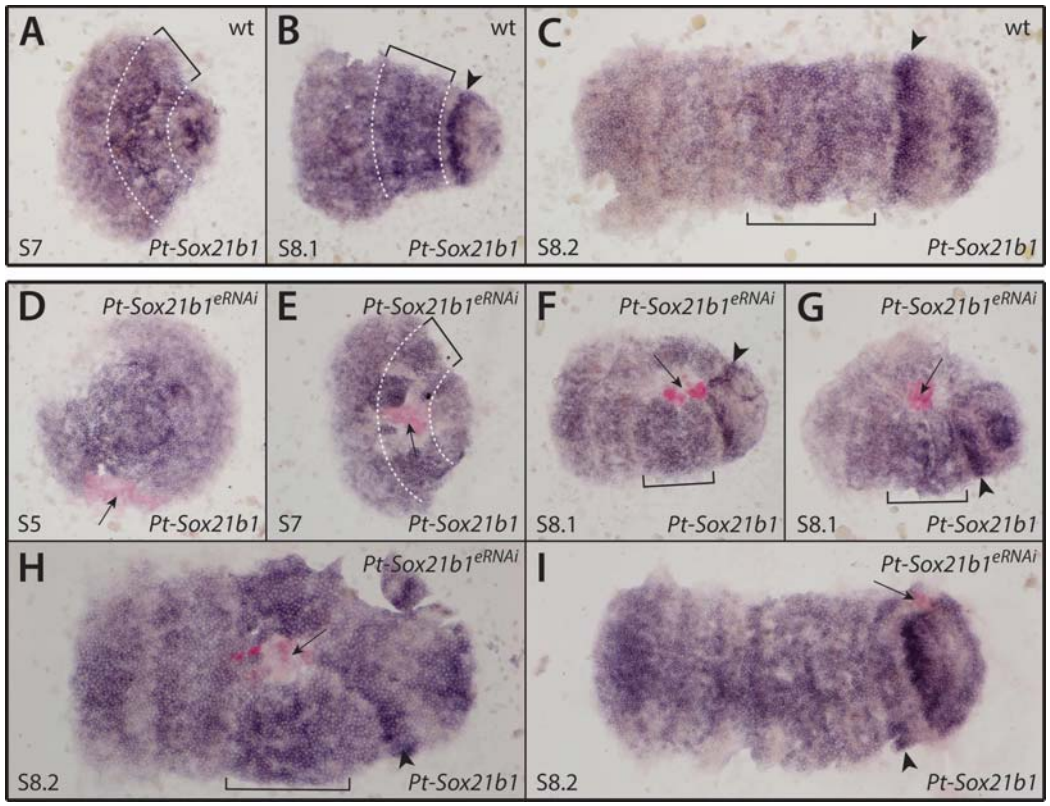

Figure 6. Effect of *Pt-Sox21b-1* eRNAi knockdown in early segmentation

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

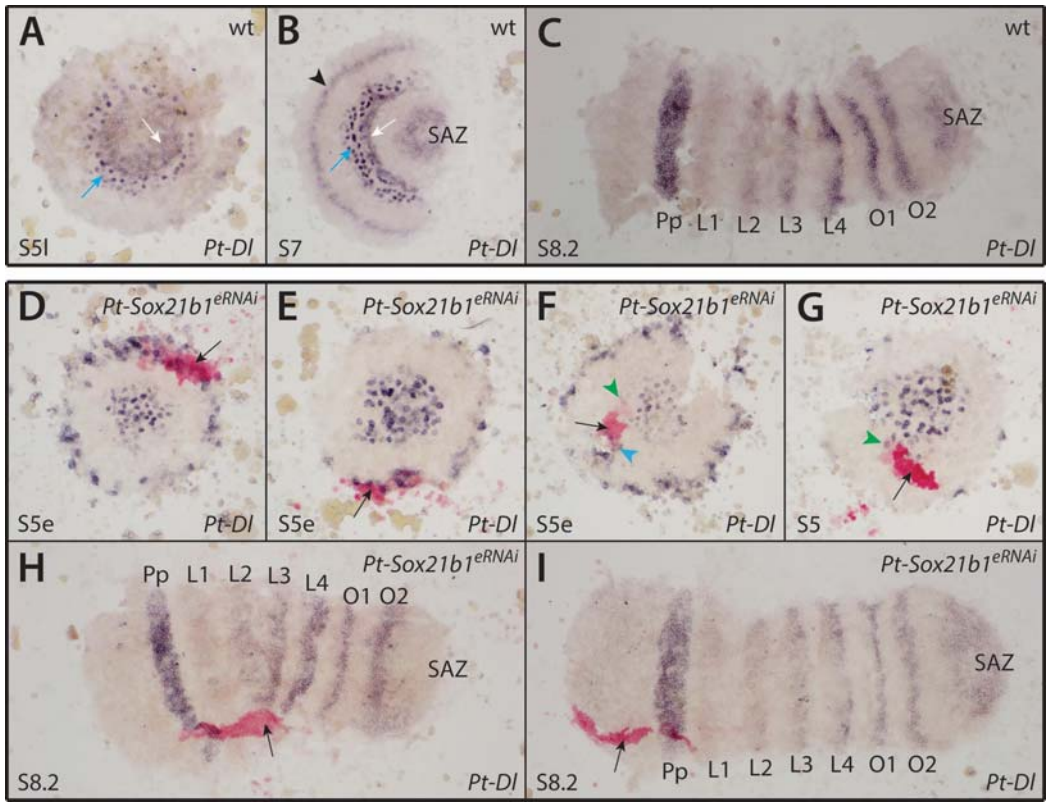

Figure 7. Effect of *Pt-Sox21b-1* eRNAi knockdown on *Pt-Dl* expression

At early stage 5, *Pt-Dl* is expressed in a subset of cells at the rim and in the centre of the germ disc, in a salt-and-pepper pattern (Oda et al., 2007). By late stage 5 (A), the expression at the rim fades and the posterior salt-and-pepper domain clears from the centre (blue arrow), where a new solid domain of expression arises (white arrow). By stage 7 (B), the posterior domain also clears from the centre, with *Pt-Dl* expression being restricted to two adjacent bands of expression, a more anterior salt-and-pepper band (blue arrow) and a more posterior solid band (white arrow). Expression also reappears in the SAZ and an anterior stripe of expression is formed in the presumptive pedipalpal segment (black arrowhead). At stage 8.2 (C), expression is restricted to segmental stripes of varying strength and the SAZ, and the salt-and-pepper domain is no longer present. (D-H) Expression pattern of *Pt-Dl* in *Pt-Sox21b-1* eRNAi embryos (detailed description in the text). Anterior is to the left in B, C, H and I. All embryos are flat mounted. Black arrows point at knockdown clones marked by pink staining. Pp, presumptive pedipalpal segment; L1-L4, presumptive L1-L4 segments; O1, presumptive O1 segment; O2, presumptive O2 segment; SAZ, segment addition zone.

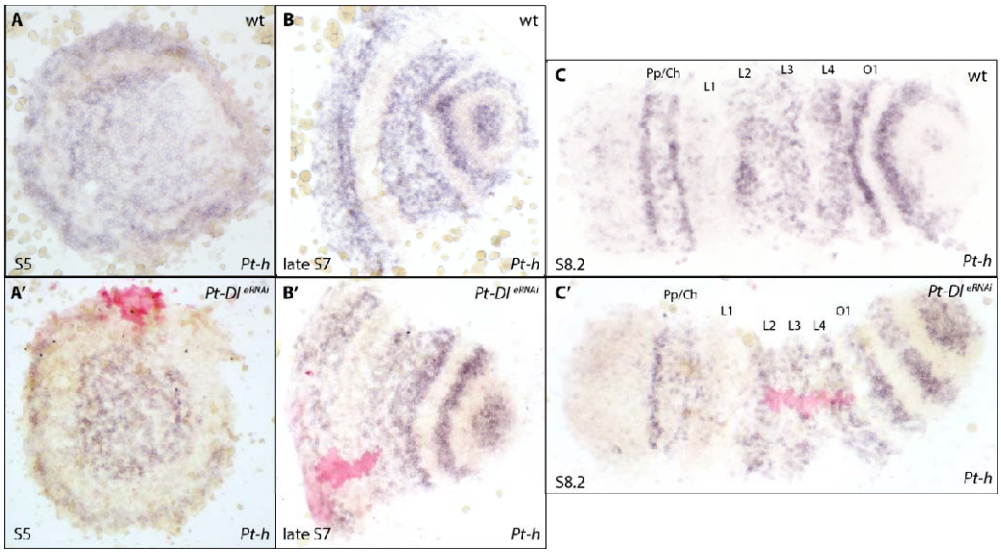

Figure 8. Effect of *Pt-Dl* eRNAi knockdown on *Pt-h* expression

Expression pattern of *Pt-h* at stage 5 (A,A'), late stage 7 (B,B') and stage 8.2 (C,C') in wildtype (A-C) and *Pt-Dl* eRNAi embryos (A'C'). *Pt-Dl* eRNAi clones are shown in pink (A'-C'). Anterior is to the left. Ch, chelicerae; Pp, pedipalps; L1-L4, walking legs 1-4.

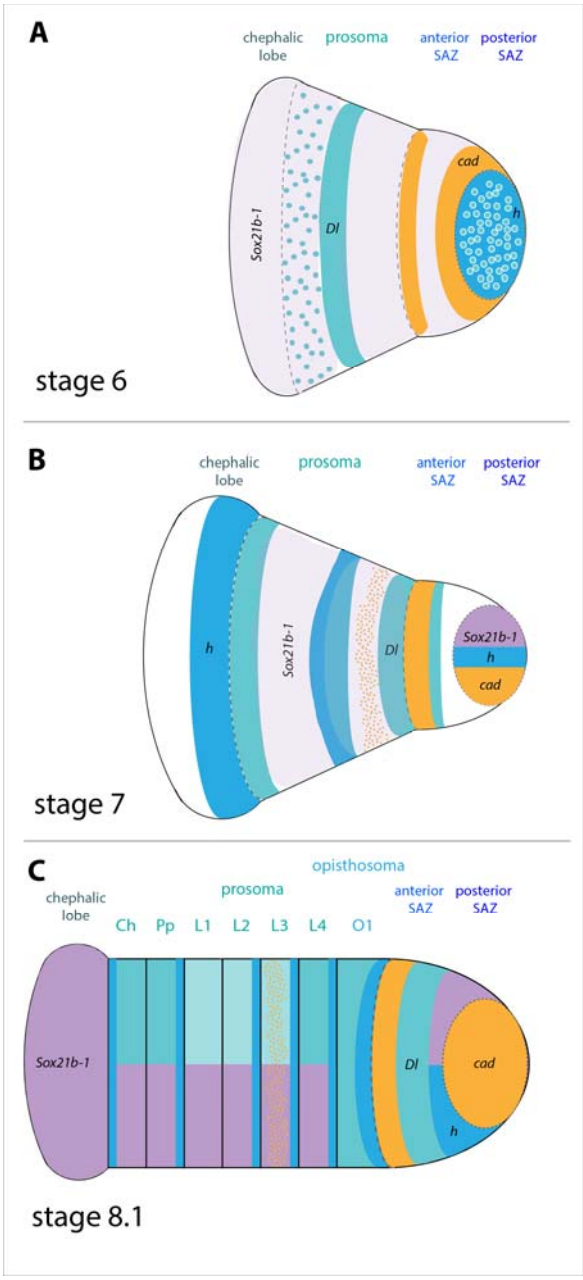

Figure 9: Summary of *Sox21b-1* expression in the GRN prosomal and opisthosomal development

Schematics of a stage 6 (A), stage 7 (B) and stage 8.1 (C) embryos. Embryos are oriented with the anterior to the left and the different parts of the germ band (cephalic lobe, prosoma, anterior SAZ and posterior SAZ) are indicated at the top and separated by dashed lines. Note that expression of *Sox21b-1*, *cad* and *h* overlap in the posterior SAZ (dashed lines), *Dl* in the prosoma overlaps with *h* expression and *Sox21b-1* is expressed in the entire prosoma (B). Solid expression of *Sox21b-1* in lilac and faint expression in light lilac, solid *Dl* expression in green and faint expression in light

green, *cad* expression in orange and *h* in blue. Ch, chelicerae; Pp, pedipalps; L1-L4, walking legs 1-4; O1-5, opisthosomal segments 1-5.

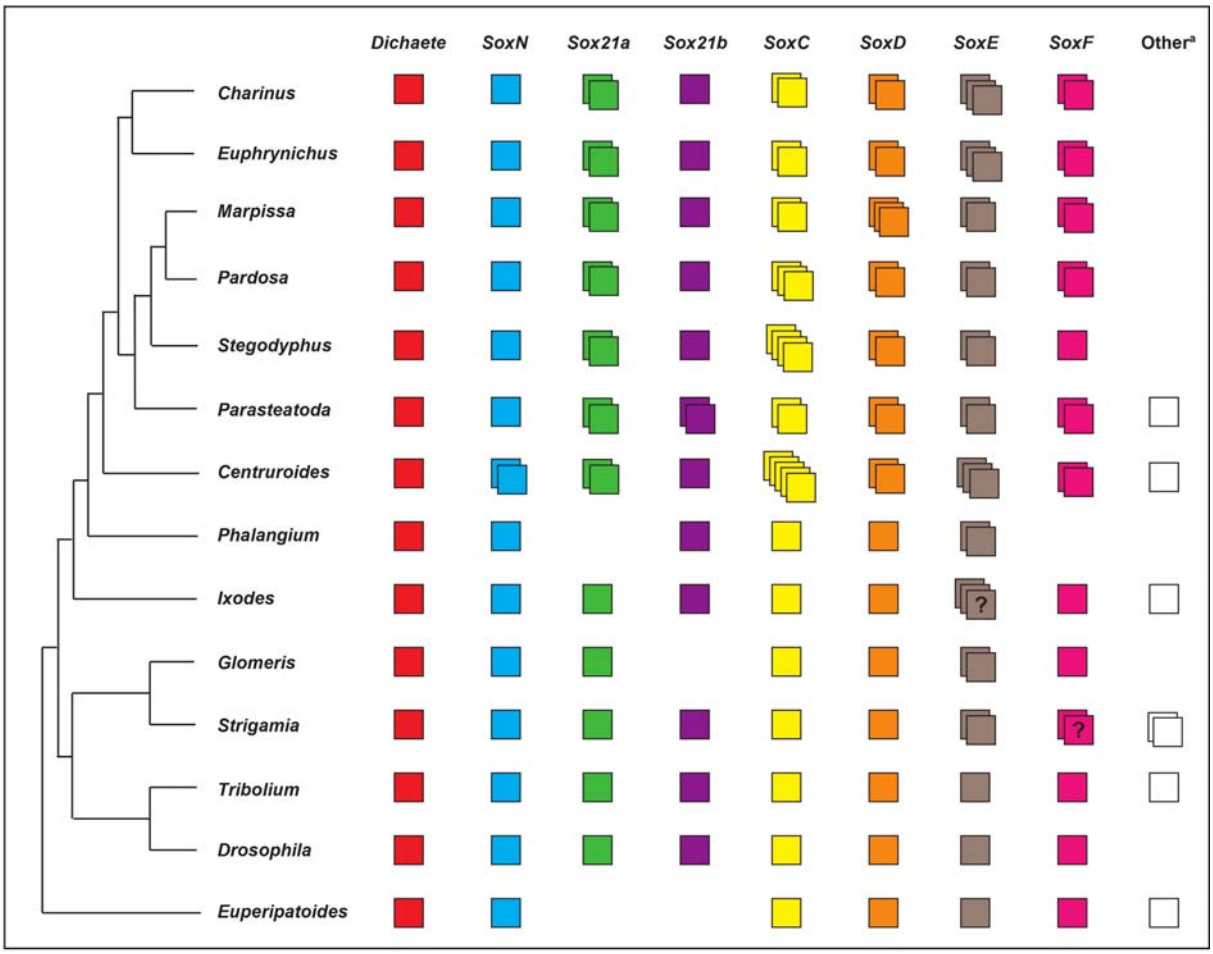

Figure 10. Panarthropod Sox gene repertoires

Sox gene repertoires in the surveyed species (*Ce. sculpturatus*, *Ph. opilio*, *I. scapularis*, *S. maritima*, *M. muscosa*, *Pa. amentata*, *C. acosta* and *E. bacillifer*) and other panarthropods for which the Sox gene repertoires were previously identified (*P. tepidariorum*, *St. mimosarum*, *G. marginata*, *T. castaneum*, *D. melanogaster* and *E. kanangrensis*). Each coloured box represents a different gene. ? = unclear number of copies due to incomplete sequences. Other[a] = unresolved or highly divergent sequences.

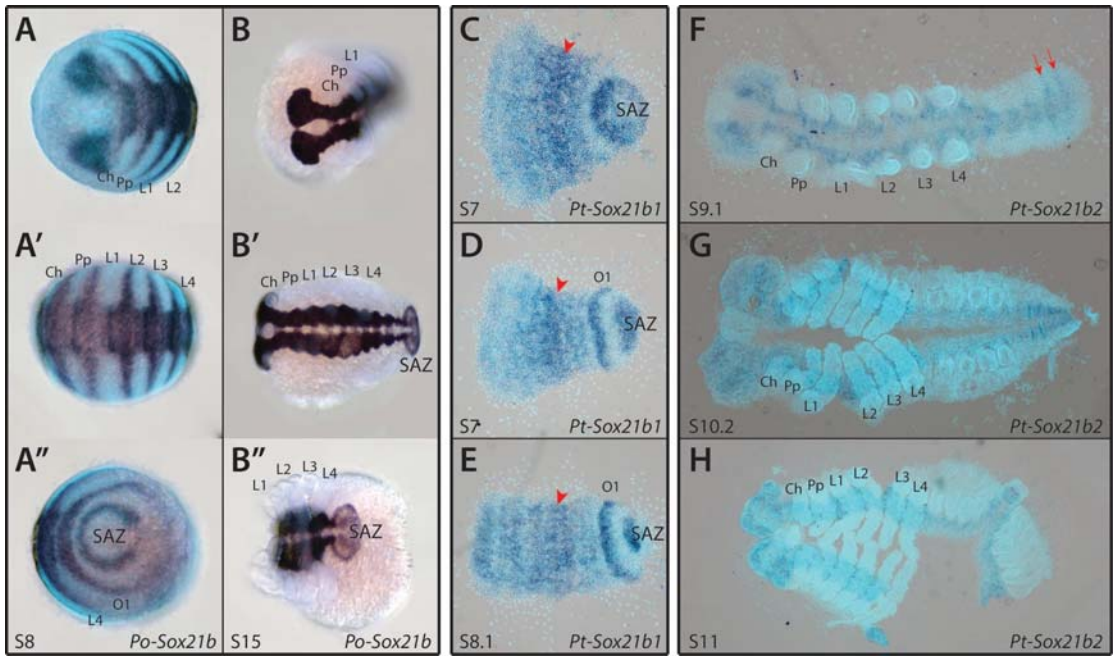

Figure 11. Expression patterns of *P. opilio* and *P. tepidariorum Sox21b* genes

Po-Sox21b has a segmental pattern of expression at stage 8 (A-A''), similar to that previously observed for *Pt-Sox21b-1* (C-E). Expression of *Pt-Sox21b-1* in the prosomal region appears to be stronger in the presumptive L2-L4 segments (red arrowheads) at stages 7 (C and D) and 8.1 (E). Additionally, *Pt-Sox21b-1* is strongly expressed in the SAZ and first opisthosomal segment at these stages (C-E). ISH for *Pt-Sox21b-2* produced a faint signal in the developing neuroectoderm, suggesting this paralog is expressed at low levels in this tissue (F-H). Furthermore, at stage 9.1 (F), expression appears to be segmental in the last two segments in the opisthosoma (red arrows). No signal was detected for *Pt-Sox21b-2* during earlier stages of embryogenesis. Anterior is to the left in all images. Embryos in C-H are flat mounted. SAZ, segment addition zone; Ch, chelicerae; Pp, pedipalps; L1-L4, walking legs 1-4; O1, opisthosomal segment 1.

Supplementary Figures

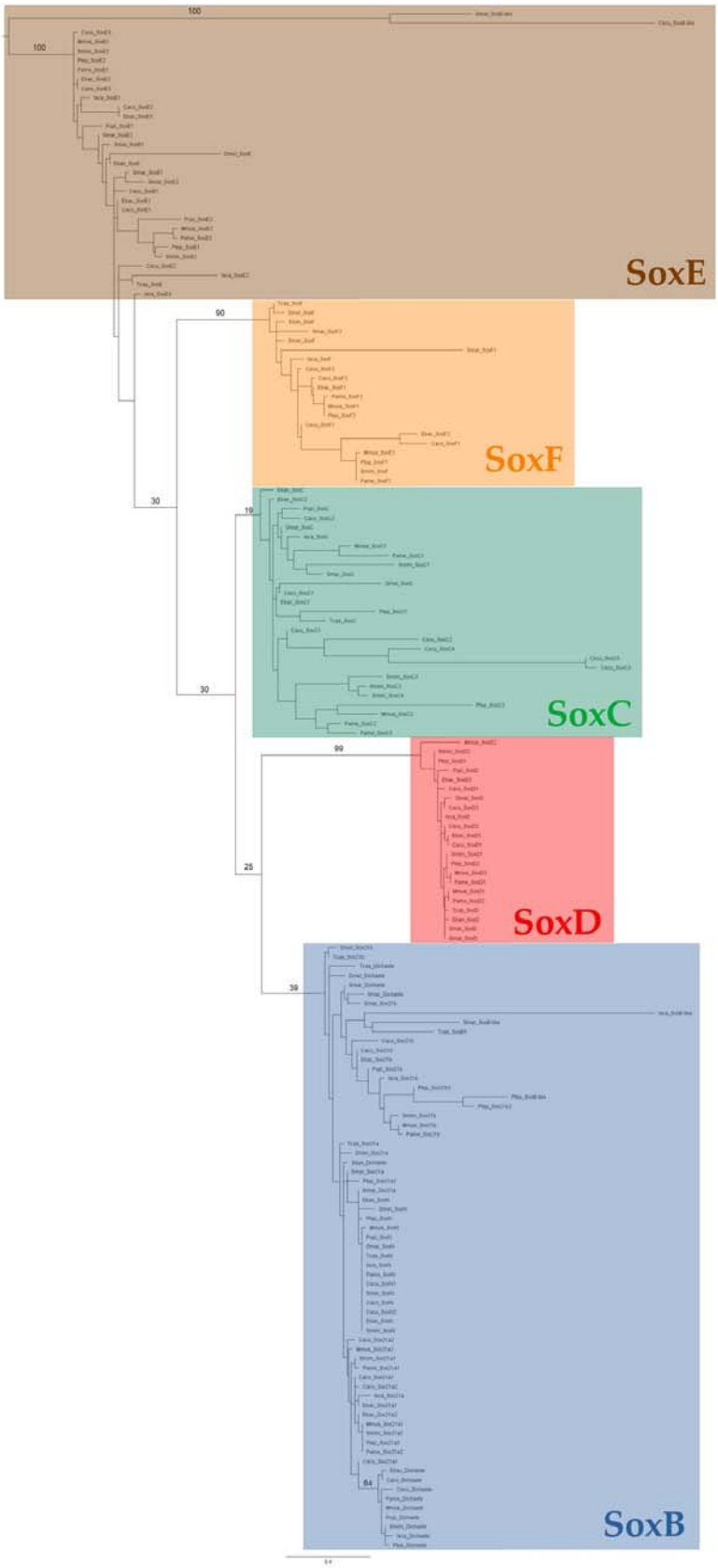

Figure S1: Panarthropod Sox HMG domain phylogeny

Maximum likelihood tree of HMG domain sequences from *E. kanangrensis*, *G. marginata*, *S. maritima*, *T. castaneum*, *D. melanogaster*, *I. scapularis*, *Ph. opilio*, *Ce. sculpturatus*, *St. mimosarum*, *P. tepidariorum*, *M. muscosa*, *Pa. amentata*, *C. acosta* and *E. bacillifer*. Rooted at midpoint. Bootstrap values are shown for main branches. RAxML run using an LG + Γ substitution model (1000 replicates).

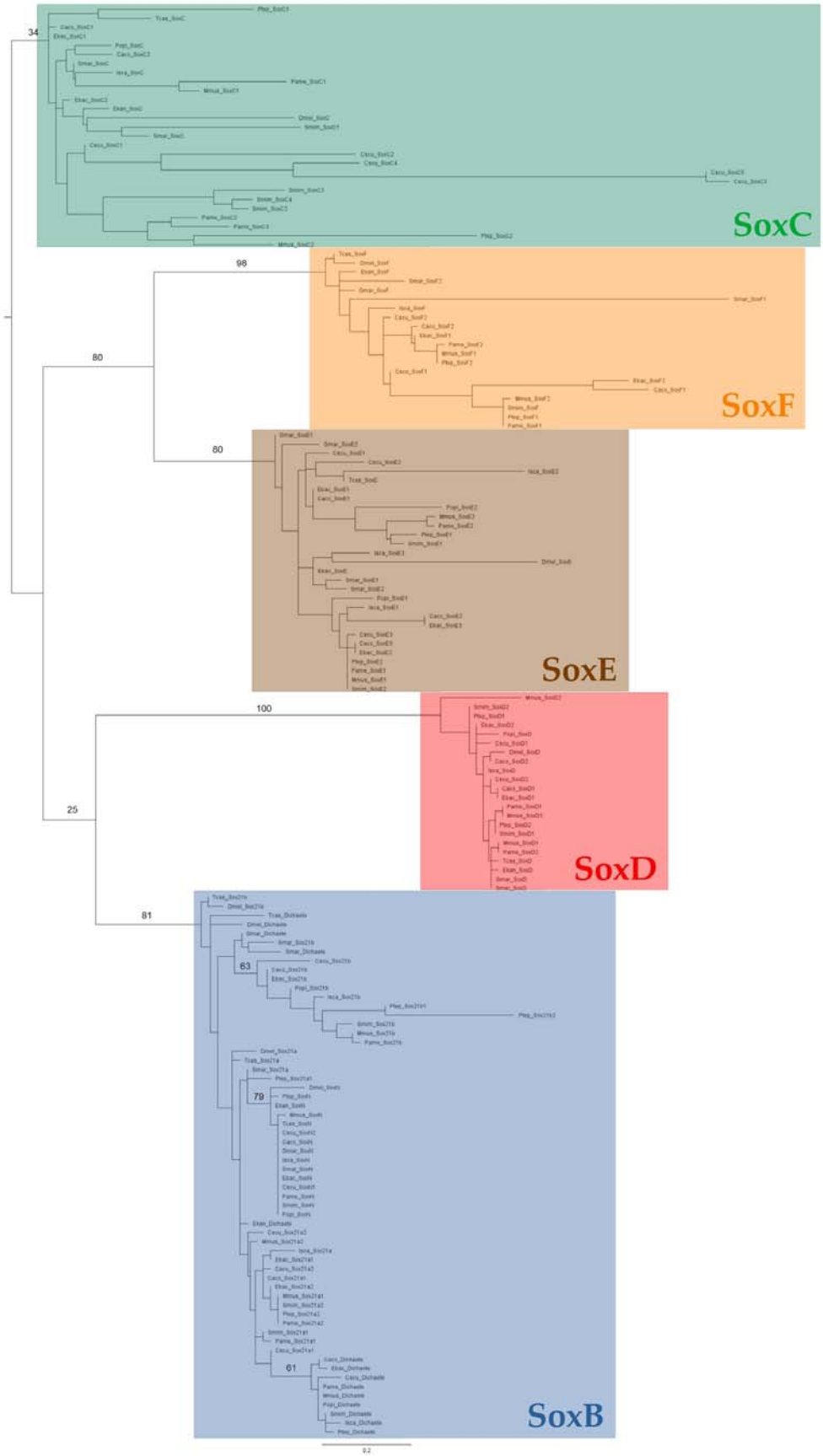

Figure S2: Panarthropod Sox HMG domain phylogeny excluding highly divergent

sequences Maximum likelihood tree of HMG domain sequences excluding all SoxB-like and SoxE-like sequences, and *Gmar-Sox21a*. Rooted at midpoint. Bootstrap values are shown for main branches. RAxML run using an LG + Γsubstitution model (1000 replicates).

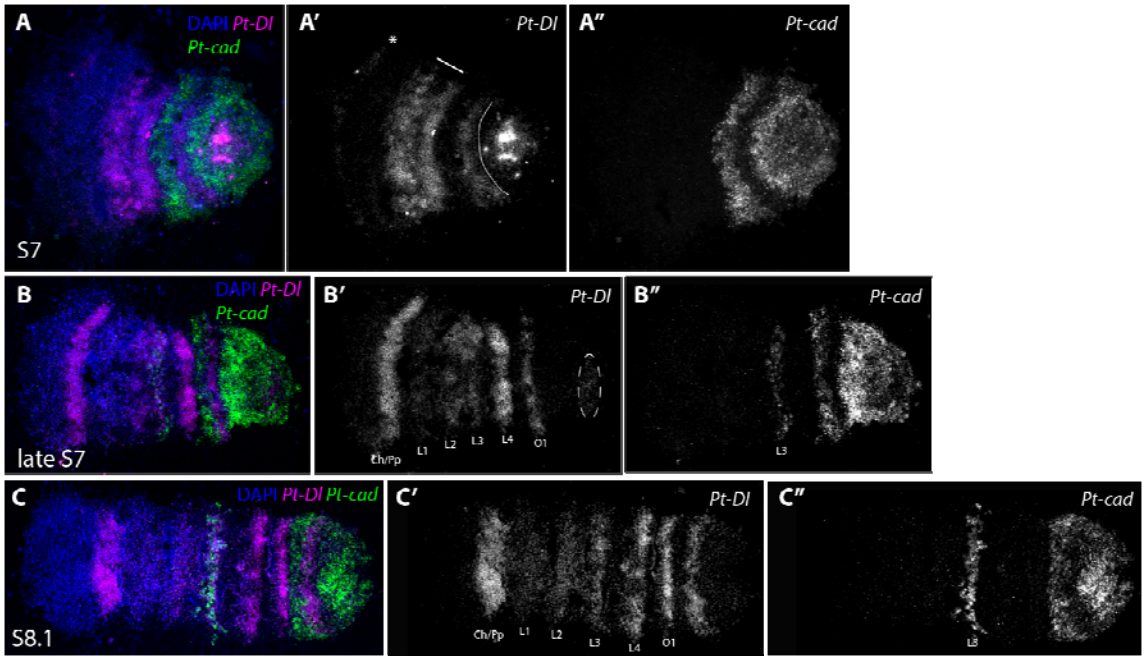

Figure S3. Expression patterns of *P. tepidariorum Dl* and *cad*

Double-fluorescent in situ of *Pt-Dl* and *Pt-cad* in stage 7 stage, late stage 7 and 8.1 embryos (A-C''). White lines in A' indicate the anterior SAZ and prosomal expression of *Pt-Dl* respectively. At stage 5, *Pt-Dl* is expressed in a circular salt-and-pepper pattern in the centre and faintly along the rim of the germ disc (white dashed line and white solid line Fig. S11A') while *Pt-h* is expressed diffusely in the centre and at the rim of the germ disc but in different cells from *Pt-Dl* (white line Fig. S11A''). At stage 7 *Pt-Dl* is expressed in the SAZ (dashed circle Fig. S11B'), in a solid domain in the prosoma (white lines Fig. S11B'), abutting another salt-and-pepper domain (Fig. S11B'). At this stage *Pt-h* shows faint expression in the SAZ, diffuse expression in the prosoma (white line Fig. S11B'') and the developing head lobes (white arrow Fig. S11B''). Note that both genes are expressed throughout the forming leg-bearing segments of the prosoma, where they partially overlap (Fig. S11B'). At stage 8.1, *Pt-Dl* expression can be detected in the anterior SAZ, O1, the walking leg-bearing segments L1-L3 and strongly in the presumptive cheliceral/pedipalpal segments (Fig. S11C'). *Pt-h* is also expressed in the anterior SAZ, O1, the walking leg segments L1-L3 (Fig. S11C''), but ahead of *Pt-Dl* expression (Fig. S11C). *Pt-h* is also expressed in the pedipalpal/cheliceral segment, either side of *Pt-Dl* expression (Fig. S11C).

Anterior is to the left in all images. White dashed lines in B' indicate the SAZ expression of *Pt-Dl*. Ch, chelicerae; Pp, pedipalps; HL, head lobe; L1-L4, walking legs 1-4.

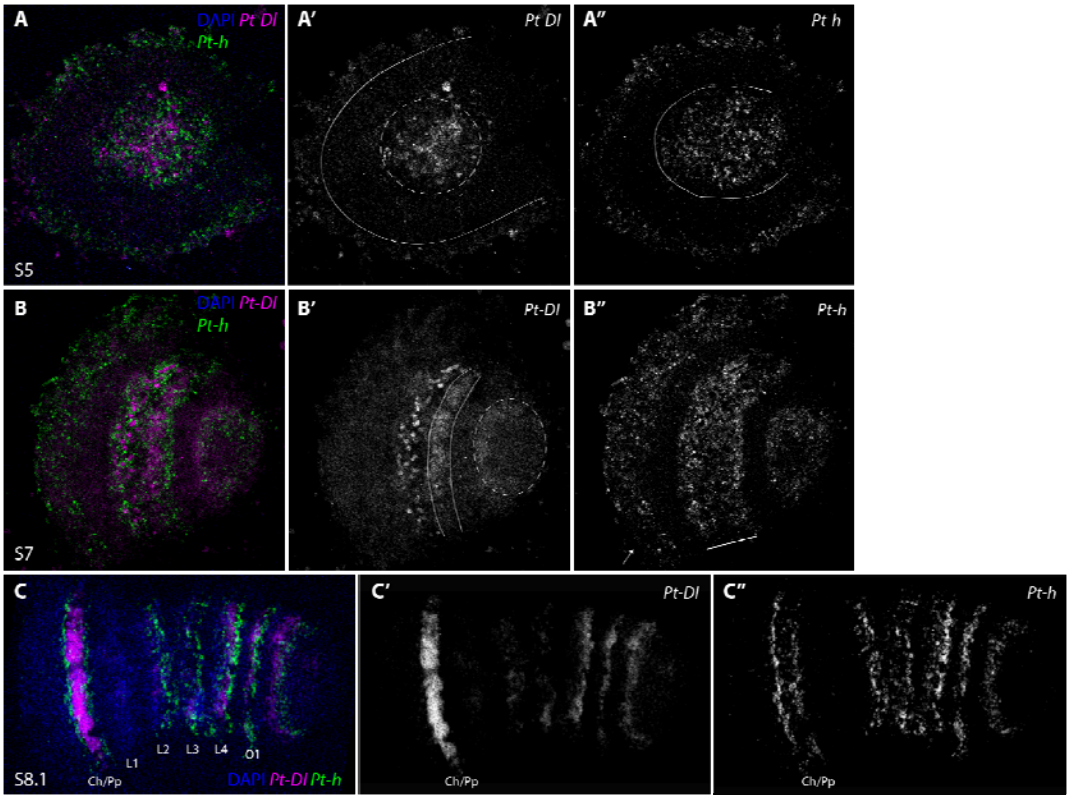

Figure S4. Double-fluorescent in situ of *P. tepidariorum Dl* and *h*

Double-fluorescent in situ of *Pt-Dl* and *Pt-h* in stage 5, stage 7 and 8.1 embryos (A-C''). White dashed circle and white line in A' indicate *Pt-Dl* expression. White line in A'' indicates the posterior *Pt-h* expression. The white dashed circle in B' indicates the faint SAZ expression and the white lines highlight the solid prosomal expression of *Pt-Dl*. The white line in B'' indicates the prosomal expression of *Pt-h*. Anterior is to the left in all images. White dashed lines in B' indicate the SAZ expression of *Pt-Dl*. Ch, chelicerae; Pp, pedipalps; HL, head lobe; L1-L4, walking legs 1-4.

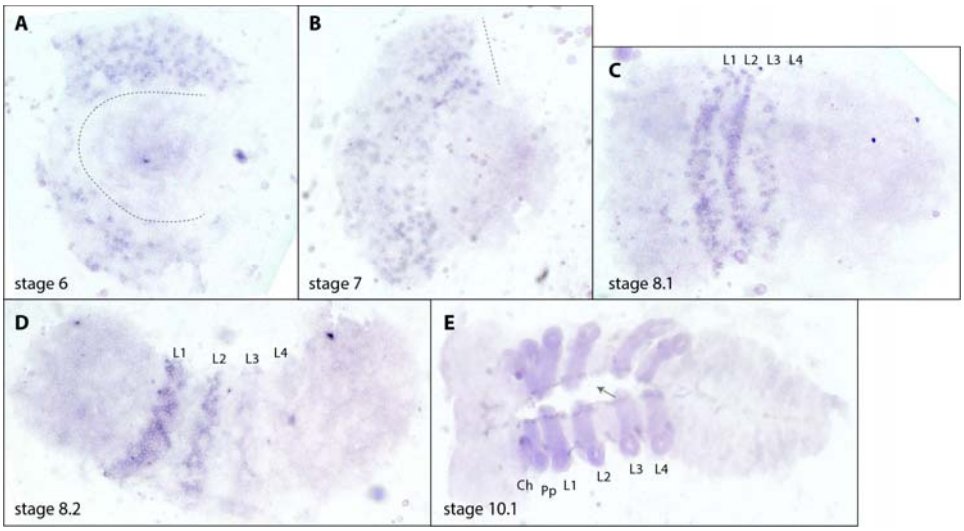

Figure S5: Expression patterns of *P. tepidariorum Dichaete*

Pt-D appears to be expressed in the prosoma during embryogenesis (A-E). At stages 6 and 7, *Pt-D* is expressed in a salt-and-pepper patter throughout the developing prosoma (domain anterior of dashed lines in A; dashed line in B). At stage 8.1 *Pt-D* expression is limited to the leg bearing segments L1-L3 (C), which becomes broader at stage 8.2 (D). Later, at stage 10.1 *Pt-D* is expressed in all prosomal appendages and along the dorsal midline (arrow, E). Anterior is to the left in all images. All embryos are flat mounted. Ch, chelicerae; Pp, pedipalps; L1-L4, walking legs 1-4; O1-O5, opisthosomal segments 1-5.

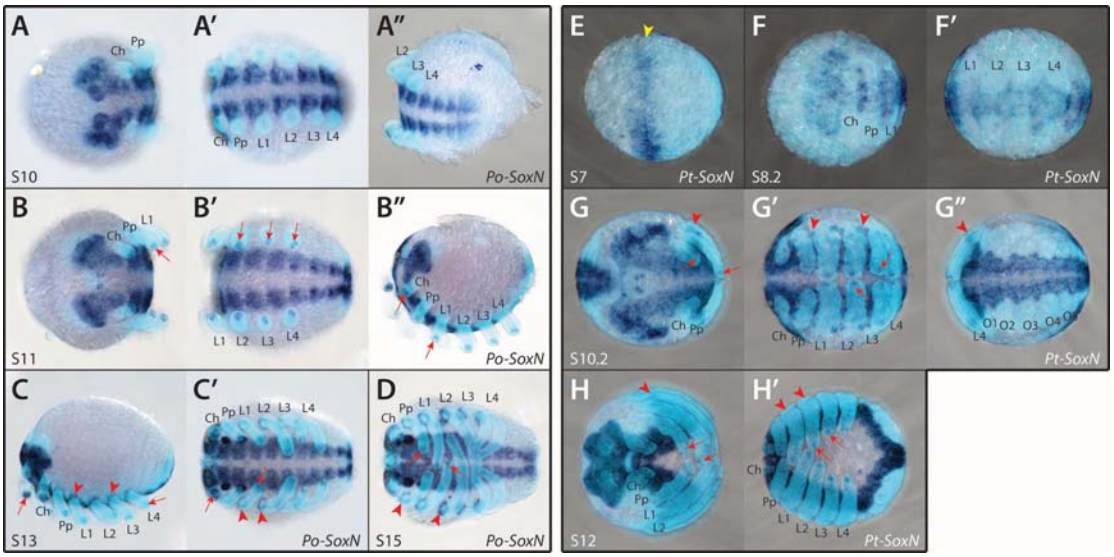

Figure S6: Expression patterns of *P. opilio* and *P. tepidariorum SoxN* genes

Po-SoxN (A-D) and *Pt-SoxN* (E-H') are both strongly expressed in the neuroectoderm throughout embryogenesis, an expression domain detected as early as stage 7 for *P. tepidariorum* in the presumptive region of the developing neuroectoderm (E, yellow arrowhead). Additional expression at the tips (red arrows) and base (red arrowheads) of each prosomal appendage is also present in both species (B-D and G-H'). Anterior is to the left in all images. Ch, chelicerae; Pp, pedipalps; L1-L4, walking legs 1-4; O1-O5, opisthosomal segments 1-5.

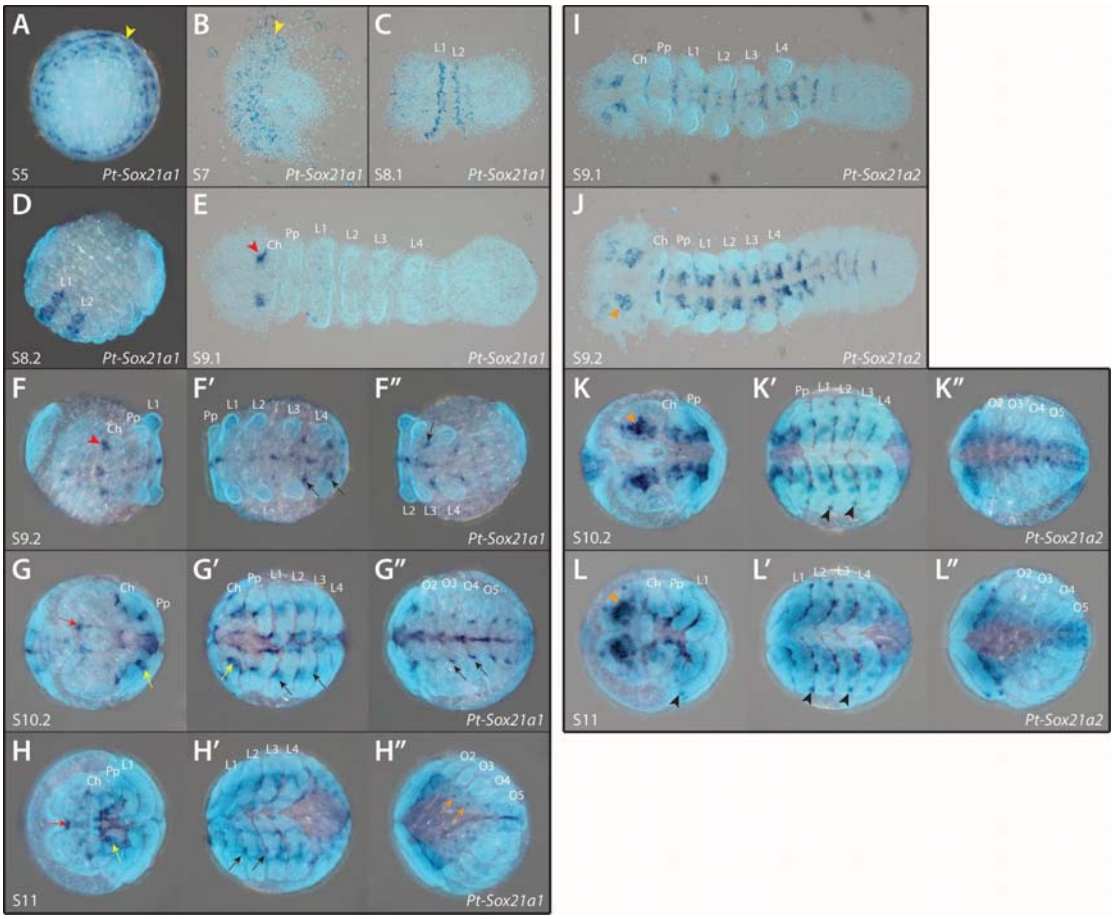

Figure S7. Expression patterns of *P. tepidariorum Sox21a* genes

Pt-Sox21a-1 expression can first be observed at stage 5 (A) as a salt-and-pepper pattern around the edge of the germ disk (yellow arrowhead), which is retained after the transition from radial to axial symmetry (B). At stages 8.1 and 8.2 (C and D), expression is restricted to two bands in the L1 and L2 segments. This pattern fades by stage 9.1 (E) and a new domain of expression arises in the pre-cheliceral region (red arrowhead). At stage 9.2 (F-F''), *Pt-Sox21a-1* starts to be expressed at the midline and small clusters of cells along the dorsal edge of the developing neuroectoderm (black arrows). Expression becomes stronger by stage 10.2 (G-G''), with enlarged domains in the prosomal segments (black arrows). An additional domain can be seen at the tips of the chelicerae (yellow arrow) and surrounding the stomodaeum (red arrow). By stage 11 (H-H''), expression in the pre-cheliceral and prosomal regions increases in complexity and midline expression fades away in most of the opisthosomal segments, replaced by a dot-like pattern (orange arrows). *Pt-Sox21a-2* is mainly expressed in the neuroectoderm, starting at stage 9.1 (I) in the pre-cheliceral region and in segmental clusters of cells. This pattern is maintained throughout embryogenesis (J-L''), with

increasing complexity in the developing brain (orange arrowheads) and, at later stages, an additional domain of expression is found at the base of the prosomal appendages (black arrowheads). The embryos in B, C, E, I and J were flat mounted. Ch, chelicerae; Pp, pedipalps; L1-L4, walking legs 1-4; O1-O5, opisthosomal segments 1-5.

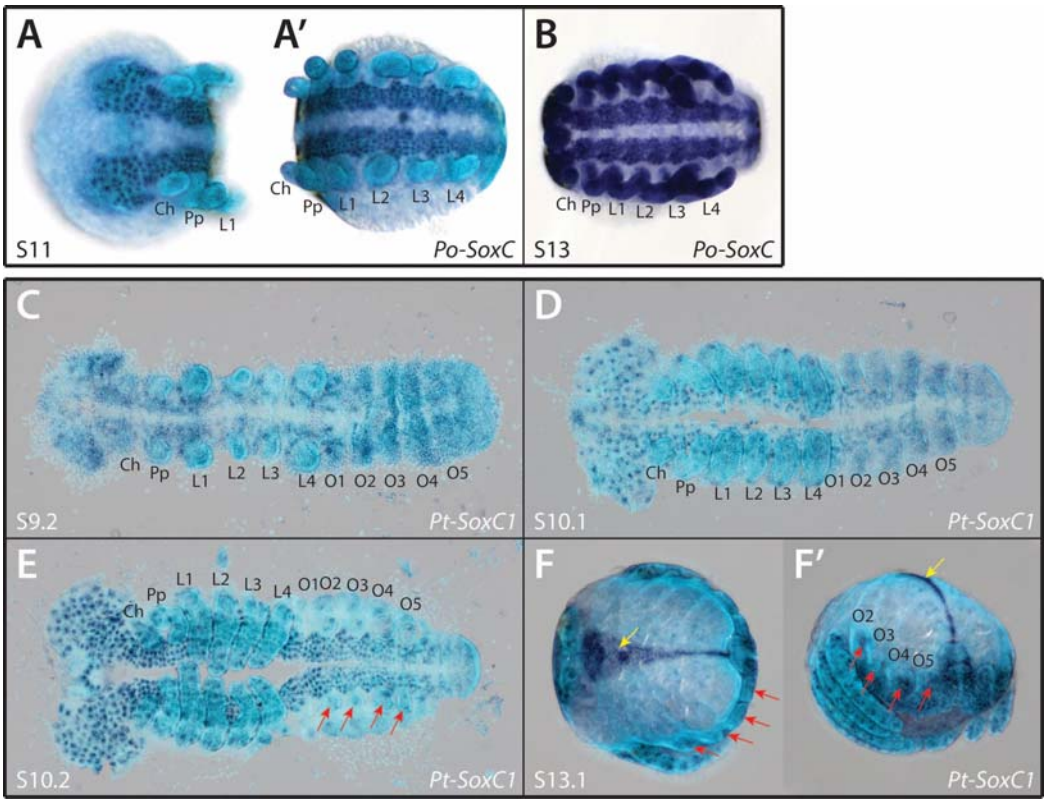

Figure S8. Expression patterns of *P. opilio* and *P. tepidariorum SoxC* genes

Po-SoxC is strongly expressed in the developing neuroectoderm, with more pronounced expression in the clusters of differentiating neuronal progenitors (A-B). At stage 13 (B), *Po-SoxC* expression was also detected in all prosomal appendages. The same neuroectodermal and prosomal appendage expression patterns were detected for *Pt-SoxC1* (C-E). Additional domains of *Pt-SoxC1* expression were observed in the opisthosomal limb buds (A-F', red arrows) and in the remaining extraembryonic tissue during dorsal closure (F and F', yellow arrows). Anterior is to the left in all images. Embryos in C, D and E were flat mounted. Ch, chelicerae; Pp, pedipalps; L1-L4, walking legs 1-4; O1-5, opisthosomal segment 1-5.

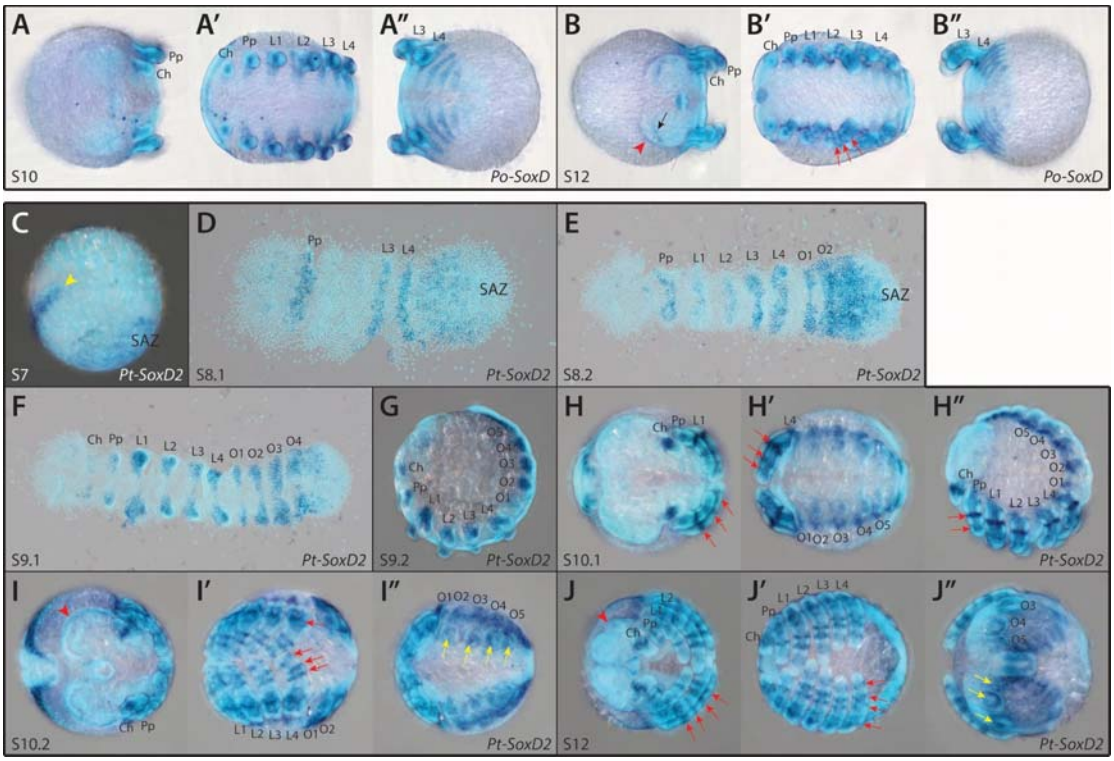

Figure S9. Expression patterns of *P. opilio* and *P. tepidariorum SoxD* genes

Po-SoxD expression is at first detected as dorsally restricted segmental stripes and in the developing prosomal appendages, forming a bifurcated pattern at the tips (A-A”). This pattern is maintained in the chelicerae at stage 12 (B-B’’), developing into 2-4 rings of expression in the pedipalps and walking legs (red arrows). Expression at the edge (red arrowhead) and above the anterior furrows (black arrow) of the pre-cheliceral region is also observed at stage 12 (B), and the segmental stripe expression pattern is still present in the opisthosomal region (B’’). *Pt-SoxD2* expression starts at stage 7 (C) as an anterior stripe (yellow arrowhead) and in the SAZ. At stage 8.1 (D), expression is restricted to the pedipalp, L3 and L4 segments, and anteriorly to the SAZ. Stripes of expression are continuously added to opisthosomal segments and additional stripes are then detected in the L1 and L2 segments (E and F). From stage 9.2 onwards (G-J’’), *Pt-SoxD2* expression is very similar to that of *Po-SoxD*. *Pt-SoxD2* expression develops into 2-4 rings of expression in the pedipalps and walking legs (red arrows) and is also present at the edge of the pre-cheliceral region (red arrowhead), but not above the anterior furrows. Expression in the opisthosoma becomes restricted to the opisthosomal appendages (yellow arrows) and the tissue dorsal to these. Anterior is to the left in all images. The embryos in D, E and F were flat mounted.

SAZ, segment addition zone; Ch, chelicerae; Pp, pedipalps; L1-L4, walking legs 1-4; O1-5, opisthosomal segment 1-5.

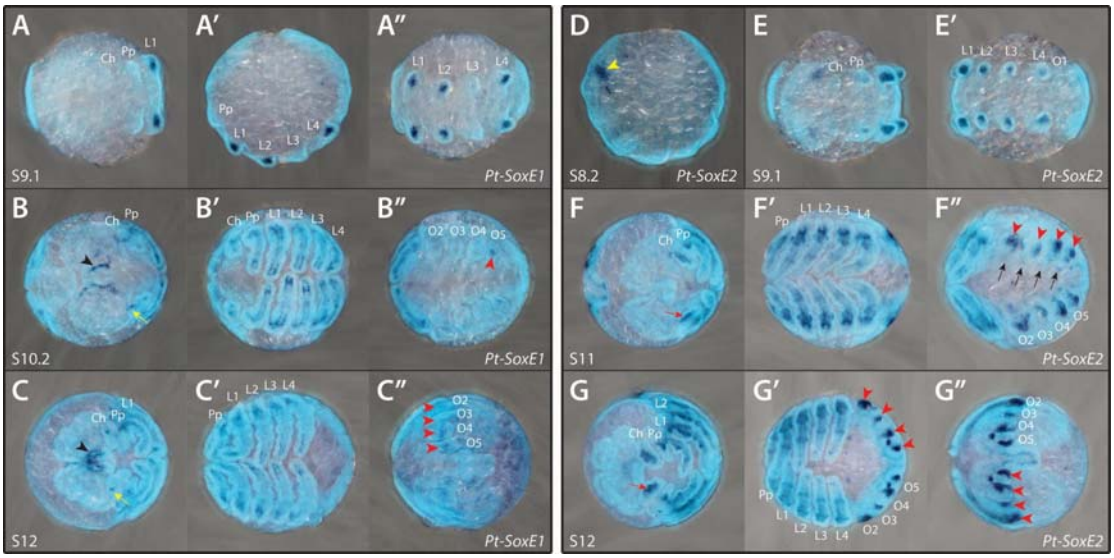

Figure S10. Expression patterns of *P. tepidariorum SoxE* genes

Pt-SoxE1 expression is first visible at stage 9.1 (A-A'') in the developing prosomal appendages of the L1, L2 and L4 segments. Expression extends to the other prosomal appendages, and by stage 10.2 (B-B'') it becomes restricted to the mesodermal tissue, with the exception of the chelicerae, where it is restricted to a small group of cells at the base (yellow arrow). This pattern remains unchanged at stage 12 (C-C''). Two additional expression domains were observed at stage 10.2 (B-B''), in the pre-cheliceral region (black arrowhead) and in the last pair of opisthosomal organs (red arrowhead). At stage 12 (C-C''), expression in the pre-cheliceral region is still present (black arrowhead), and faint expression was observed in all opisthosomal appendages (red arrowheads) and in the expanding dorsal tissue. *Pt-SoxE2* ISH staining (D-G'') revealed expression in the prosomal appendages similar to that of *Pt-SoxE1*, with the exception of the chelicerae, where expression extends to the tip (red arrow). At stage 8.2 (D), *Pt-SoxE2* is expressed in two clusters of cells in the pre-cheliceral region (yellow arrowhead), which starts fading at stage 9.1 (E) and is no longer visible during later stages (F and G). Specific and strong expression can be seen in all opisthosomal appendages (F', G' and G''; red arrowheads). Faint expression is also present in small clusters of ventral cells in the opisthosoma (black arrows). Anterior is to the left in all images. Ch, chelicerae; Pp, pedipalps; L1-L4, walking legs 1-4; O1-5, opisthosomal segment 1-5.

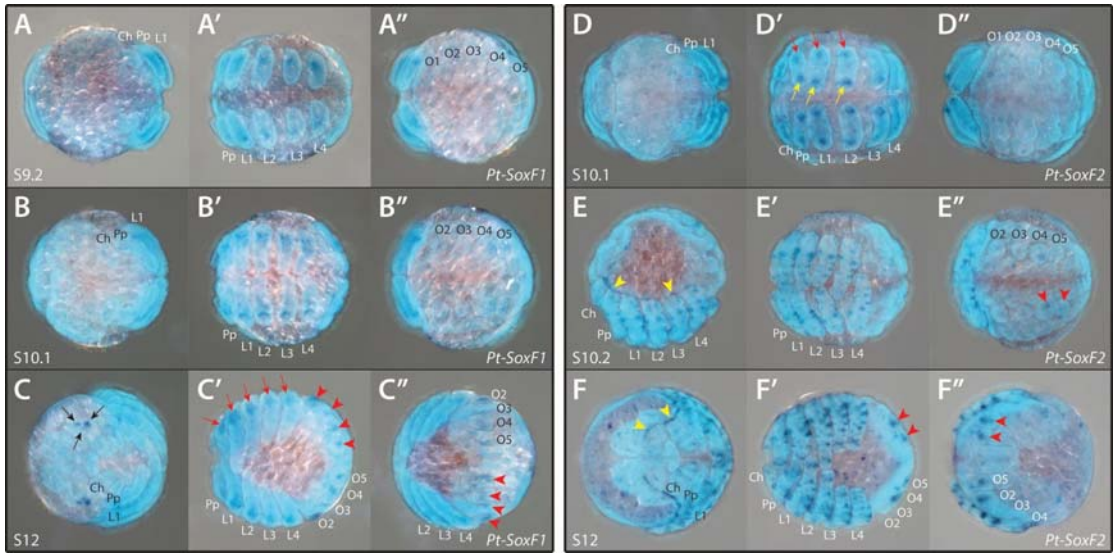

Figure S11. Expression patterns of *P. tepidariorum SoxF* genes

Faint expression of *Pt-SoxF1* is first visible at stage 9.2 (A-A'') in the prosomal appendages. This pattern remains unchanged up to stage 12 (B-B"), when it becomes restricted to the base of the legs (red arrows). At stage 12 (C-C''), additional expression domains arise in the lateral eye primordia (black arrows) and in the opisthosomal appendages (red arrowheads). *Pt-SoxF2* is also expressed at the tips (yellow arrows) and base (red arrows) of the prosomal appendages, starting at stage 10.1 (D-D"). This pattern is more complex at later stages and becomes restricted to several small groups of cells that possibly belong to the developing PNS (E' and F'). At stage 10.2 (E-E''), *Pt-SoxF2* expression appears along the edge of the prosomal region of the germ band, from the cheliceral to the L4 segments (yellow arrowheads). This domain extends to the pre-cheliceral region at stage 12 (F-F''), in a pattern that seems to follow the edge of the growing non-neurogenic ectoderm (yellow arrowheads). Lastly, *Pt-SoxF2* is also expressed in the last two pairs of opisthosomal appendages – the presumptive spinnerets (red arrowheads). Anterior is to the left in all images. Ch, chelicerae; Pp, pedipalps; L1-L4, walking legs 1-4; O1-5, opisthosomal segment 1-5.