






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}

¹Department of Biological and Medical Sciences, Faculty of Health and Life Sciences, Oxford Brookes University, Oxford, United Kingdom

²Unit Integrative Zoology, Department of Evolutionary Biology, University of Vienna, Vienna, Austria

³Centre for Functional Genomics, Oxford Brookes University, Oxford, United Kingdom

⁴Oxford University Museum of Natural History, Oxford, United Kingdom

⁵Department of Genetics, University of Cambridge, Cambridge, United Kingdom

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

[†]These authors contributed equally to this work.

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

Associate editor: John Parsch

Abstract

The Sox family of transcription factors regulates many processes during metazoan development, including stem cell maintenance and nervous system specification. Characterizing the repertoires and roles of these genes can therefore provide important insights into animal evolution and development. We further characterized the Sox repertoires of several arachnid species with and without an ancestral whole-genome duplication and compared their expression between the spider *Parasteatoda tepidariorum* and the harvestman *Phalangium opilio*. We found that most Sox families have been retained as ohnologs after whole-genome duplication and evidence for potential subfunctionalization and/or neofunctionalization events. Our results also suggest that *Sox21b-1* likely regulated segmentation ancestrally in arachnids, playing a similar role to the closely related *SoxB* gene, *Dichaete*, in insects. We previously showed that *Sox21b-1* is required for the simultaneous formation of prosomal segments and sequential addition of opisthosomal segments in *P. tepidariorum*. We studied the expression and function of *Sox21b-1* further in this spider and found that although this gene regulates the generation of both prosomal and opisthosomal segments, it plays different roles in the formation of these tagmata reflecting their contrasting modes of segmentation and deployment of gene regulatory networks with different architectures.

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

Introduction

The Sox (Sry-related high-mobility group (HMG) box) genes encode an ancient family of transcription factors that play important roles in the regulation of many aspects of animal development (Kamachi and Kondoh 2013). There are eight known groups (A–H) of Sox genes defined by the sequence of their HMG domains (Bowles et al. 2000; Heenan et al. 2016). Although group A (represented by Sry) is restricted to eutherian mammals and groups G and H are also thought to be lineage specific, representatives of groups B–F are generally found in all metazoan lineages, although there have been lineage-specific losses (Bowles et al. 2000). Our work, and that of others, indicate that groups B–F were all represented in the common ancestor of arthropods with each group represented in *Drosophila melanogaster* (Phochanukul and

Russell 2010; Janssen et al. 2018; Paese, Leite, et al. 2018). However, there has been duplication of some groups in other arthropod lineages (Paese, Leite, et al. 2018) and we have shown that the group B genes *Sox21a* and *Sox21b* as well as the groups C, D, E, and F genes are represented by at least two genes in the spider *Parasteatoda tepidariorum*. This is consistent with the whole-genome duplication (WGD) that took place in an ancestor of arachnospulmonates (Schwager et al. 2017; Leite et al. 2018) (fig. 1) and patterns of gene retention observed after independent WGD in vertebrates (Wegner 1999; Bowles et al. 2000). Further characterization of the repertoire and role of Sox genes among arachnids has great potential to better understand the evolution of these genes, and their expression and function in the development of arthropods, as well as gene fate and the outcomes of WGD more generally.

We previously found that one of the *P. tepidariorum* SoxB genes, *Sox21b-1*, regulates the formation of segments in this spider (Paese, Leite, et al. 2018; Paese, Schönauer, et al. 2018). In spiders, there are at least two modes of trunk segmentation: the leg-bearing segments of the prosoma (cephalothorax) are generated more or less simultaneously like the segments of long-germ insects. Opisthosomal (abdominal) segments are added sequentially like most other arthropods, although one at a time instead of two, in contrast to short-germ insects (reviewed in Hilbrant et al. 2012; Clark et al. 2019; Oda and Akiyama-Oda 2020). It also appears that different gene regulatory networks underlie the simultaneous formation of anterior versus sequential addition of posterior segments (Oda et al. 2007; McGregor et al. 2008; Schwager et al. 2009; Pechmann et al. 2011; Schönauer et al. 2016; Paese, Schönauer, et al. 2018; Akiyama-Oda and Oda 2020). The leg-bearing segments of the prosoma are added through splitting of wider gene expression domains regulated by the gap gene-like activity of *hunchback* (*hb*) and *Distal-less* (*Dll*) (Schwager et al. 2009; Pechmann et al. 2011). The opisthosomal segments are sequentially added from a posterior segment addition zone (SAZ), which is in part established by the activity of Wnt and Notch-Delta signaling pathways (Oda et al. 2007; Schönauer et al. 2016). The interplay between these pathways directs sequential segment addition from the SAZ by regulating dynamic expression of pair-rule gene orthologs, including *even-skipped* (*eve*), and other segmentation genes such as *caudal* (*cad*) (Oda et al. 2007; McGregor et al. 2008; Schönauer et al. 2016). Knockdown of *Pt-Sox21b-1* expression 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, Schönauer, et al. 2018). Therefore, *Sox21b-1*, along with *arrow* and *Msx1*, is one of the few genes known to regulate both simultaneous and sequential segmentation in *P. tepidariorum* (Paese, Schönauer, et al. 2018; Setton and Sharma 2018; Akiyama-Oda and Oda 2020).

Interestingly, *Dichaete*, a closely related SoxB group gene, is found to regulate segmentation in both *D. melanogaster* and *Tribolium castaneum*, long-germ and short-germ insects, respectively (Russell et al. 1996; Clark and Peel 2018). 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.

This conclusion requires a broader understanding of the evolution and expression of Sox genes in other arachnids, as well as a more detailed understanding of the role of these genes, including *Sox21b-1*, in *P. tepidariorum*. To address this, we explored the evolution and roles of Sox genes in arachnid development further by characterizing the repertoire of Sox genes in five additional arachnids with and without an ancestral WGD and compared the expression of Sox genes between the harvestman *Phalangium opilio* and the spider *P. tepidariorum*. Furthermore, we assessed the potential roles of *Dichaete*, *Sox21a-1*, *Sox21b-2*, and *SoxD-2* in segmentation in *P. tepidariorum* and examined the expression and function of *Sox21b-1* in more detail. We found similar patterns of retention of duplicated Sox genes after WGD in spiders,

scorpions, and whip spiders and evidence for potential sub and/or neofunctionalization during embryogenesis in the spider *P. tepidariorum*. Our results also suggest that although other Sox genes, including *Dichaete*, may not be required for segmentation in *P. tepidariorum*, *Sox21b-1* regulated prosomal and opisthosomal segmentation ancestrally in arachnids as a key component of the gene regulatory networks for their simultaneous versus sequential production, respectively.

Results

Arachnid Sox Gene Repertoires

To further investigate the evolution of the Sox gene repertoires of arachnids compared with other arthropods more broadly, we surveyed new and available genomic resources to identify Sox HMG domain-containing sequences from several additional arachnids: the spiders, *Marpissa muscosa* and *Pardosa amentata*; the amblypygids, *Charinus acosta* and *Euphrynichus bacillifer*; the scorpion, *Centruroides sculpturatus*; the harvestmen, *Ph. opilio* and the tick *Ixodes scapularis*; and a myriapod, the centipede *Strigamia 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 with previously identified arthropod Sox genes (Janssen et al. 2018; Paese, Leite, et al. 2018).

Our previous analysis of the *P. tepidariorum* genome identified 15 Sox genes (Paese, Schönauer, et al., 2018) (fig. 1). We identified 14 Sox genes in *P. amentata*, 14 in *M. muscosa*, 14 in *C. acosta*, 14 in *E. bacillifer*, 19 in *C. sculpturatus*, 7 in *Ph. opilio*, 11 in *I. scapularis*, and 12 in *S. maritima* (fig. 1 and supplementary file 1, Supplementary Material online). However, some of the sequences retrieved had incomplete HMG domains (10/105), probably representing fragments of the full sequence (supplementary file 2, Supplementary Material online). Therefore, the Sox repertoires described here may be incomplete due to the quality of genomic or transcriptomic assemblies or some genes were potentially not captured in the transcriptomes sequenced, as we found previously for homeobox genes (Leite et al. 2018), 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 (supplementary figs. S1 and S2, Supplementary Material online). The trees supported the classification of most sequences obtained, forming monophyletic clades corresponding to each Sox group (supplementary figs. S1 and S2, Supplementary Material online). When constructing trees with all HMG sequences, the bootstrap support values were low, as previously found with this domain (Paese, Leite, et al. 2018). However, removing *Pt-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) (supplementary fig.

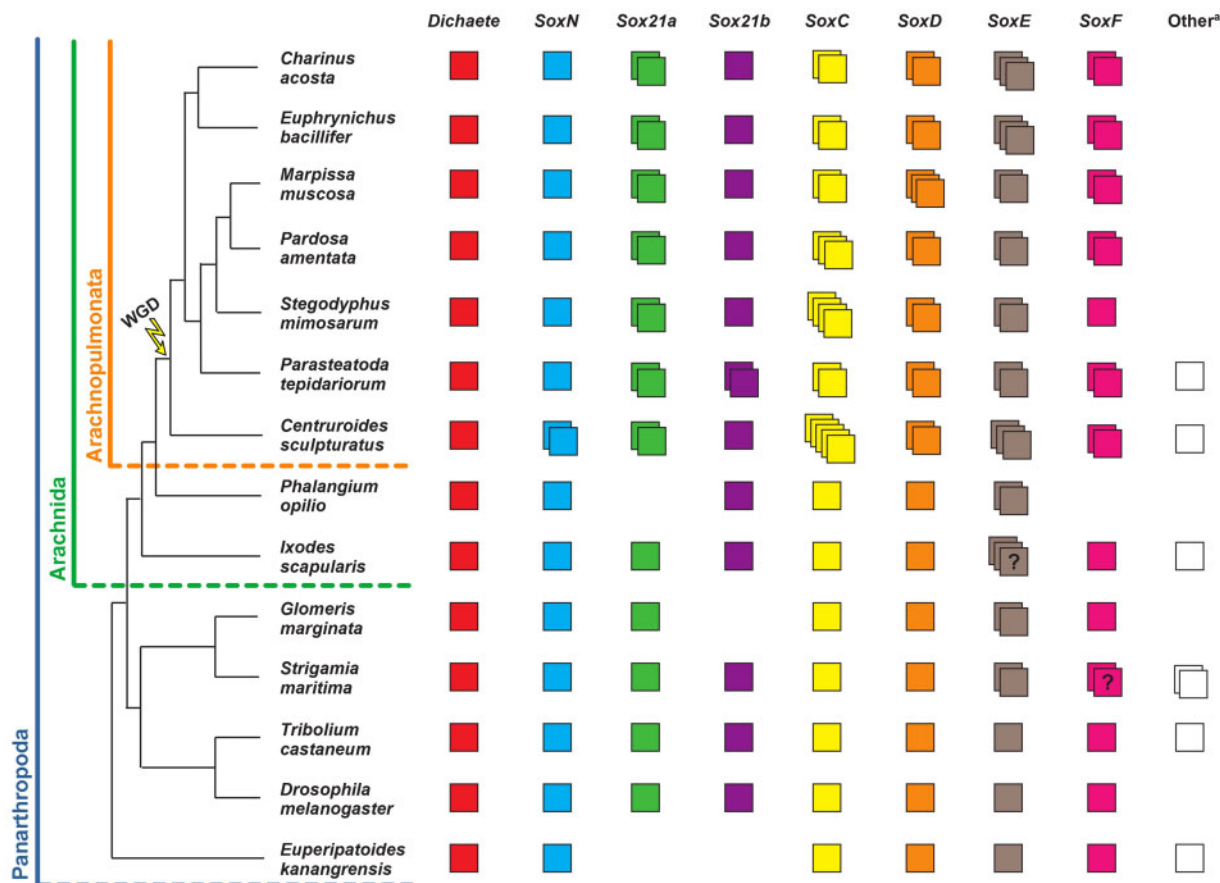


Fig. 1. Panarthropod Sox gene repertoires. Sox gene repertoires in the surveyed species (*Centruroides sculpturatus*, *Phalangium opilio*, *Ixodes scapularis*, *Strigamia maritima*, *Marpissa muscosa*, *Pardosa amentata*, *Charinus Acosta*, and *Euphrynichus bacillifer*) and other panarthropods for which the Sox gene repertoires were previously identified (*Parasteatoda tepidariorum*, *Stegodyphus mimosarum*, *Glomeris marginata*, *Tribolium castaneum*, *Drosophila melanogaster*, and *Euperipatoides kanangrensis*). Each colored box represents a different gene² = unclear number of copies due to incomplete sequences. Other^a = unresolved or highly divergent sequences.

S2, Supplementary Material online). 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) (supplementary fig. S2, Supplementary Material online). Bayesian analysis using the same HMG sequence alignment yielded similar results (supplementary fig. S3, Supplementary Material online), further supporting the identity of the sequences surveyed. 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. 1). Consistent with previous surveys (Janssen et al. 2018; Paese, Leite, et al. 2018), a single copy of *Dichaete* was found in all species analyzed (fig. 1). A single copy of SoxN was also found in most species, with the exception of *C. sculpturatus* where two sequences with identical HMG domains were found, although they differ outside this domain

(fig. 1). A single copy of Sox21a is present in non-WGD arthropods, whereas two copies were found in all seven arachnoplumonate species analyzed (fig. 1). Additional SoxB-like sequences were also found in *C. 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, Leite, et al. 2018), these sequences diverge significantly from other group B genes and therefore cannot be reliably classified (fig. 1 and supplementary fig. S2, Supplementary Material online).

All non-WGD arthropods (insects, myriapod, harvestman, and tick) have single copies of SoxC and SoxD, but in arachnoplumonates, at least two copies for each of these groups were found (fig. 1). At least two SoxE and SoxF genes were also found in all arachnoplumonates except for *S. mimosarum*, where SoxF is only represented by one gene, and in *Ph. opilio* SoxF is potentially missing (fig. 1). SoxE also appears to be duplicated in *Ph. opilio*, *G. marginata*, *S. maritima*, and *I. scapularis*, although *Smar-SoxE2* has an incomplete domain, and two of the three copies in *I. scapularis* contain nonoverlapping fragments of the HMG domain (indicating they might be the same gene) (fig. 1 and supplementary fig. S2, Supplementary Material online). An additional SoxE-like sequence was found in *S. maritima*, although this sequence is

highly divergent from other *SoxE* genes (fig. 1 and supplementary fig. S2, Supplementary Material online). We note that *SoxE* duplication may be relatively common in invertebrate lineages compared with 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. 1 and supplementary fig. S2, Supplementary Material online).

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. 1). This suggests that after the arachnopolunonate 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 *Ph. opilio*

To help better understand the evolution and roles of *Sox* genes in arachnids, we compared the expression patterns of the *P. tepidariorum* *Sox* genes with their orthologs in *Ph. opilio*. Although we previously characterized *Sox* gene expression patterns in *P. tepidariorum*, we only detected expression for 6 out of 15 *Sox* genes present in this species (*Pt-SoxN*, *Pt-Sox21b-1*, *Pt-SoxC1*, *Pt-SoxD1*, *Pt-SoxE1*, and *Pt-SoxF2*) (Paese, Leite, et al. 2018). 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 in situ hybridization (ISH) probes on 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, Leite, et al. 2018) (supplementary figs. S4–S10, Supplementary Material online). However, we found that *Pt-D* is expressed in a diffuse salt and pepper pattern in the developing cephalic lobe and prosoma at stages 6 and 7 as well as potentially in the SAZ (although this may be background signal) (supplementary fig. S4, Supplementary Material online). We also detected *Pt-D* expression in segmental stripes in L1–L3 at stages 8.1 and 8.2 as well as along the ventral midline and in the prosomal appendages at stage 10.1 (supplementary fig. S4, Supplementary Material online). In addition, we were now able to detect expression of *Pt-Sox21a-1* and *Pt-Sox21a-2* in nonoverlapping patterns mainly in the neuroectoderm, which is suggestive of sub- and/or neofunctionalization (supplementary fig. S6, Supplementary Material online). Furthermore, *Pt-Sox21a-1* appears to be expressed in segmental stripes in L1 and L2 at stage 8.1 and more ventrally restricted stripes in the opisthosoma at stage 9.2, which suggests that it may play a role in segmentation (supplementary fig. S6, Supplementary Material online). We also observed *Pt-SoxE-2* expression in a similar pattern to *Pt-SoxE-1* in the prosomal and opisthosomal limb buds, although

with some temporal and spatial differences, as well as unique expression of *Pt-SoxE-2* in opisthosomal cells that are possibly the germline progenitors (Schwager et al. 2015) (supplementary fig. S9, Supplementary Material online). We also found that although *Pt-SoxF-1* and *Pt-SoxF-2* are both expressed in the developing prosomal and opisthosomal appendages, each has also a specific domain of expression: *Pt-SoxF-1* is expressed in the secondary eye primordia (supplementary fig. S10, Supplementary Material online), and *Pt-SoxF-2* is expressed along the dorsal border of the prosomal segments and along the edge of the nonneurogenic ectoderm (supplementary fig. S10, Supplementary Material online).

We were able to detect expression for four of the seven identified *Ph. opilio* *Sox* genes (*SoxN*, *Sox21b*, *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 polymerase chain reaction (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, Leite, et al. 2018) as well as at the tips and base of the prosomal appendages (supplementary fig. S5, Supplementary Material online). *Pt-SoxC-1* and *Po-SoxC* are also expressed in the developing neuroectoderm as well as in the prosomal appendages at later stages of development, although *Pt-SoxC-1* is also expressed in a few cells, where *Po-SoxC* is not detected (supplementary fig. S7, Supplementary Material online). *Po-SoxD* and *Pt-SoxD-2* also exhibit similar expression in the precheliceral region, prosomal appendages, and as segmental stripes (supplementary fig. S8, Supplementary Material online), but an additional domain of *Pt-SoxD-2* expression was detected in the opisthosomal appendages (supplementary fig. S8, Supplementary Material online).

Finally, 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. 2). We were also able to detect expression for *Pt-Sox21b-2* in the last two opisthosomal segments (fig. 2) as well as a faint signal in the developing neuroectoderm that did not appear to overlap with *Pt-Sox21b-1* expression at stage 9.2 (fig. 2 and supplementary fig. S11, Supplementary Material online). These results suggest that *Sox21b* played a role in segmentation ancestrally in arachnids and that there may have been subfunctionalization of *Sox21b* function after duplication in *P. tepidariorum*.

Are Other *Sox* Genes in Addition to *Sox21b-1* Require for Segmentation in *P. tepidariorum*?

Our analysis of *Sox* gene expression suggested that in addition to *Pt-Sox21b-1*; *Pt-D*, *Pt-Sox21a-1*, and *Pt-SoxD-2* may also be involved in segmentation in *P. tepidariorum* (supplementary figs S4, S6, and S8, Supplementary Material online). Therefore, we carried out parental RNAi (pRNAi) knockdown of these three *Sox* genes as well as repeating *Pt-Sox21b-1* knockdown as a positive control and for further analyses of this treatment (supplementary fig. S12 and file 3, Supplementary Material

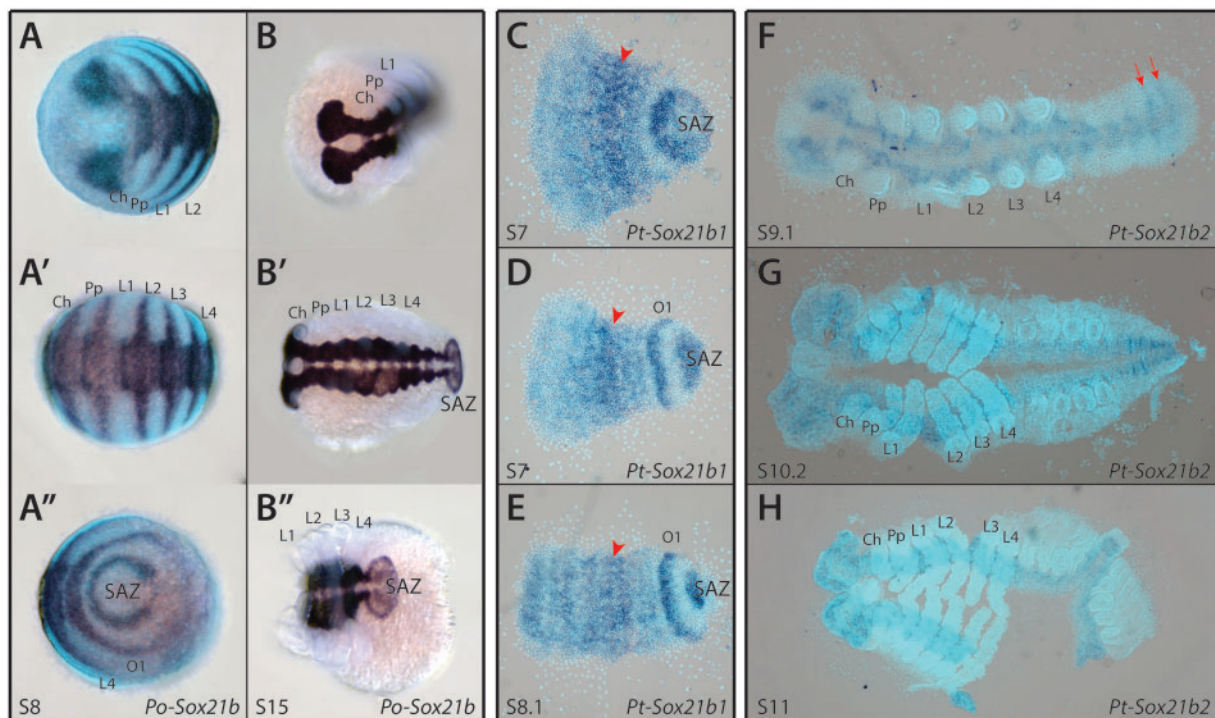


FIG. 2. Expression patterns of *Phalangium opilio* and *Parasteatoda tepidariorum* Sox21b genes. *Po-Sox21b* has a segmental expression pattern at stage 8 (A–A''), similar to that previously observed for *Pt-Sox21b-1* (C–E). Expression of *Pt-Sox21b-1* in the prosomal region appears to be stronger in 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 (C–H) are flat mounted. Ch, chelicerae; Pp, pedipalps; L1–L4, walking legs 1–4; O1, opisthosomal segment 1.

online). We did not observe any effect on embryonic development from knocking down *Pt-D*, *Pt-Sox21a-1*, and *Pt-Sox2D2*, although we were able to reproduce the same phenotypic effects at an approximately similar frequency as we previously obtained with *Pt-Sox21b-1* pRNAi (Paese, Schönauer, et al. 2018) (supplementary fig. S12 and file 3, Supplementary Material online). This suggests that although *Pt-Sox21b-1* is essential for segmentation, *Pt-D*, *Pt-Sox21a-1*, and *Pt-Sox2D2* may not be required.

Further Analysis of Sox21b-1 Function in Prosomal Segmentation in *P. tepidariorum*

Knockdown of *Sox21b-1* expression in *P. tepidariorum* inhibits the formation of all leg-bearing segments, although in some embryos L1 still forms (Paese, Schönauer, et al. 2018). To further evaluate the effect of *Pt-Sox21b-1* knockdown on the regulation of anterior segmentation, we characterized the expression patterns of *Pt-Dll*, *Pt-hb*, and *Pt-Msx1*, which have known roles in this process (Schwager et al. 2009; Pechmann et al. 2011; Leite et al. 2018; Akiyama-Oda and Oda 2020), in *Pt-Sox21b-1* pRNAi knockdown embryos.

Pt-Dll is required for the development of L1 and L2 segments and is also expressed in the SAZ (fig. 3A–D) and later in the prosomal limb buds as well as some developing opisthosomal appendages (Pechmann et al. 2011). The early ring-like expression domain in L1 is still present in stage 5 knockdown embryos ($n = 5$) and maintained as a stripe up to stage 7

($n = 3$) (fig. 3E–G). In *Pt-Sox21b-1* pRNAi embryos, at stage 8.1, the L1 stripe of *Pt-Dll* expression is perturbed compared with wild-type embryos and the L2 stripe is missing ($n = 5$) (fig. 3H–I'). This is consistent with the loss of L1 and more commonly L2 upon *Pt-Sox21b-1* knockdown and suggests that 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. 3G–J'). In some stage 8.1 *Pt-Sox21b-1* pRNAi embryos expression of *Pt-Dll* was observed in the presumptive L3–L4 region, adjacent to the disorganized cell mass of the presumptive SAZ, perhaps indicating there was incomplete knockdown of *Pt-Sox21b-1* in these embryos ($n = 2$) (fig. 3H).

Pt-hb is expressed in the prosoma (fig. 4A and B) and 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. 4), which is consistent with the loss of segments observed after this treatment (Paese, Schönauer, et al. 2018). 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. 4C and D).

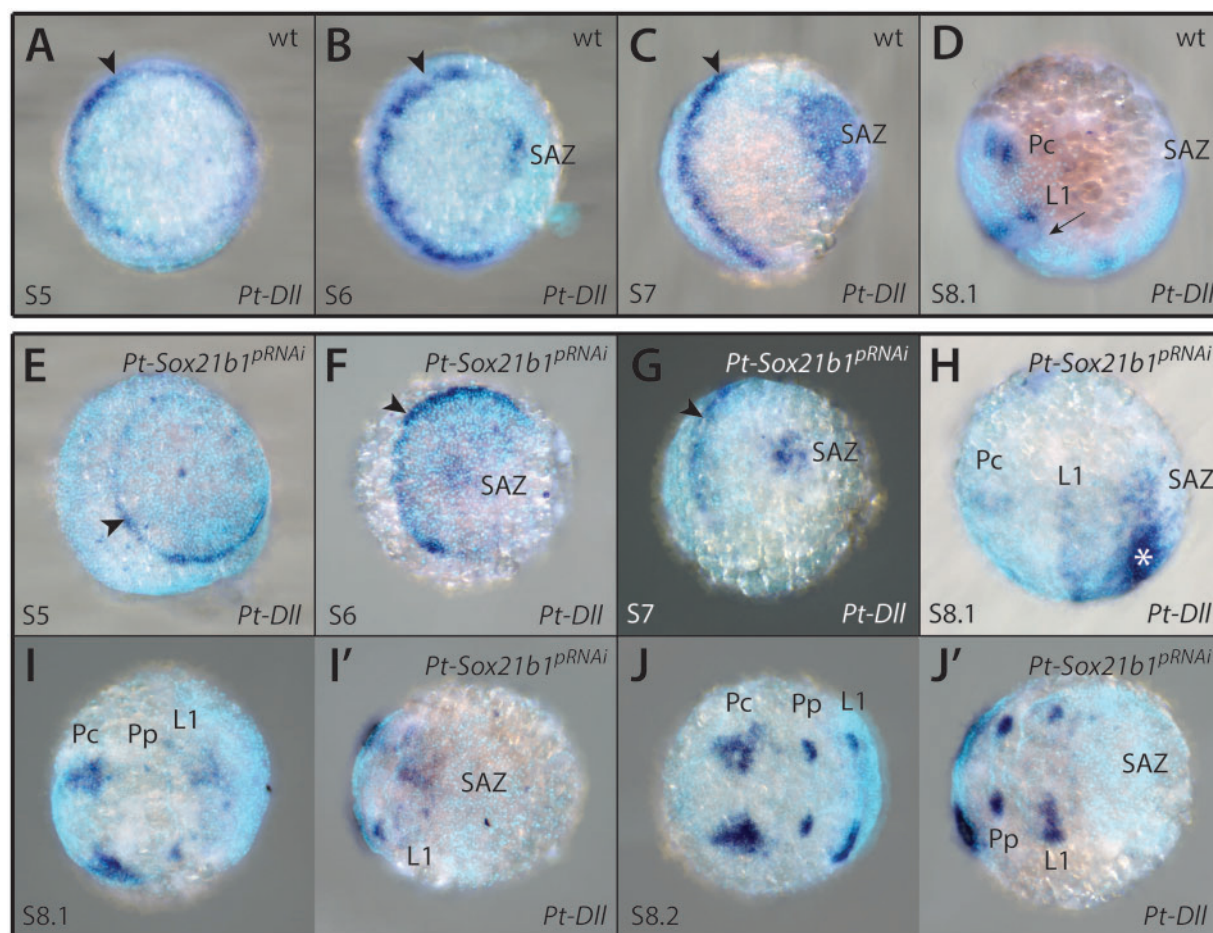


Fig. 3. Effect of *Pt-Sox21b-1* pRNAi knockdown on *Pt-Dll* expression. *Pt-Dll* is expressed in the presumptive region of the L1 segment from as early as stage 5 (A, black arrowhead), first as a ring-like pattern that transforms into a stripe at the germ disc to germ band transition (B, black arrowhead). Additional expression domains in the SAZ and precheliceral region arise at stages 6 (B) and 7 (C), respectively, as well as a faint stripe in the L2 segment at stage 8.1 (D, black arrow). At stage 8.2, expression is restricted to the precheliceral domains and developing prosomal appendages (Pechmann et al. 2011). Expression of *Pt-Dll* does not seem to be affected in stages 5 (E) and 6 (F) *Pt-Sox21b-1* pRNAi embryos. At stage 7 (G), *Pt-Sox21b-1* knockdown leads to the reduction of *Pt-Dll* expression in the SAZ, although the anterior stripe of expression is unaffected (black arrowhead). In stage 8.1 knockdown embryos (H and I'), expression is also reduced in the presumptive L1 segment and the faint stripe in the L2 segment is missing. In some embryos ($n = 2$), ectopic expression was observed near the posterior of the germ band, adjacent to the presumptive SAZ (H, asterisk). The precheliceral and developing prosomal appendage expression domains are seemingly unaffected (H–J'). Anterior is to the left in all images. Pc, precheliceral region; Pp, presumptive pedipalpal segment; L1, presumptive L1 segment.

Subsequently, although the precheliceral domain and pedipalpal stripes of *Pt-hb* expression develop normally, the L1/L2 domain does not split into segmental stripes, and the L3 stripe cannot be distinguished from L4 (fig. 4D), whereas in some embryos, all tissue and expression posterior of L1 is lost ($n = 4$) (fig. 4F). These results suggest that although *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–L4 and the formation of these segments.

Pt-Msx1 is expressed in a segmental pattern during early embryogenesis (fig. 5A–D) and regulates both prosomal and opisthosomal segmentation (Leite et al. 2018; Akiyama-Oda and Oda 2020). *Pt-Msx1* is still expressed in *Pt-Sox21b-1* pRNAi embryos at stage 5, although due to the highly deformed state of the embryos analyzed, it is difficult to tell

whether the pattern is restricted to the presumptive L2–L4 segments or if it extends anteriorly into the presumptive head and L1 regions ($n = 5$) (fig. 5E). In early stage 7, *Pt-Sox21b-1* pRNAi 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. 5F–G). However, at late stage 7, although 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. 5H) (and is consistent with the effects of *Pt-Sox21b-1* embryonic RNAi (eRNAi) – see below). 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. 5F–H). Taken together, these results are similar

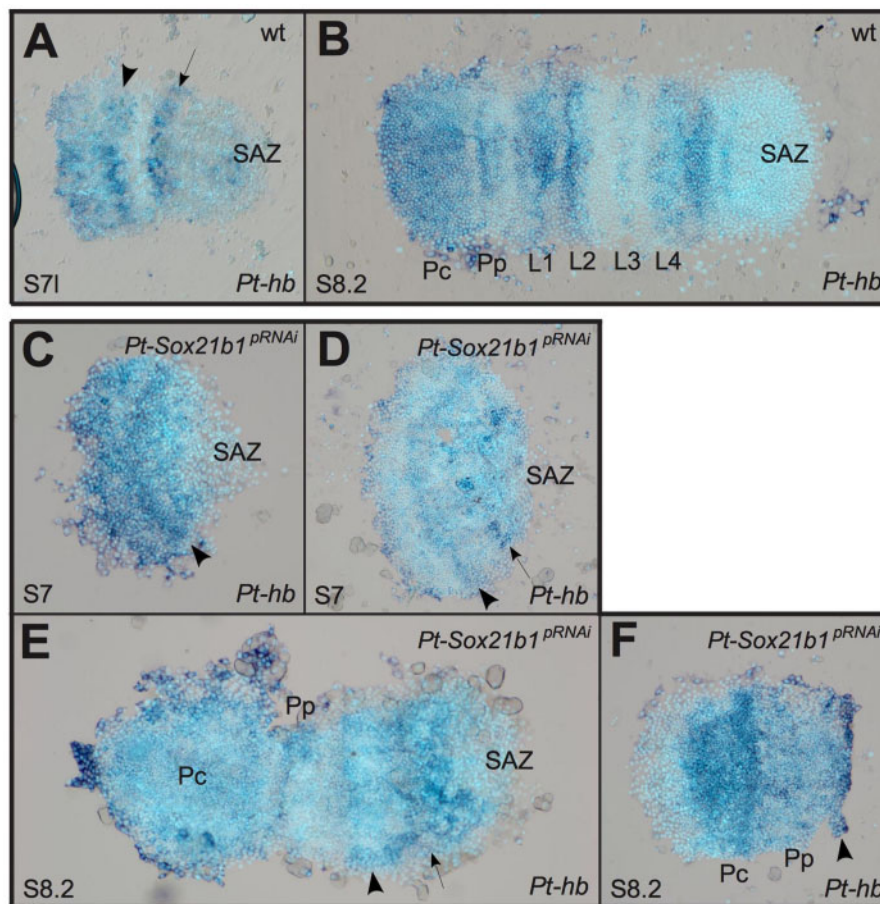


FIG. 4. Effect of *Pt-Sox21b-1* pRNAi knockdown on *Pt-hb* expression. At late stage 7 (A), two bands corresponding to the L1/L2 segments (black arrowhead) and presumptive L4 segment (black arrow) are observed. Subsequently, by stage 8.2 (B) *Pt-hb* expression becomes segmentally restricted, with a domain covering the precheliceral region and cheliceral segment, and as stripes in all other anterior segments, with the strongest expression in the L1, L2, and L4 segments. Expression of *Pt-hb* in the presumptive L1/L2 segments is unaffected in stage 7 *Pt-Sox21b-1* pRNAi embryos (C and D, black arrowhead), although expression in the presumptive L4 segment is absent in some embryos (black arrow). At stage 8.2 (E and F), expression in the precheliceral and pedipalpal segments is unaffected by *Pt-Sox21b-1* knockdown, however, separation of the L1/L2 band of expression does not seem to occur (black arrowhead), and a single band of expression is observed in the presumptive L3/L4 segments (black arrow). Anterior is to the left in all images. Pc, precheliceral region; Pp, presumptive pedipalpal segment; L1–L4, presumptive L1–L4 segments.

to the effect of *Pt-Sox21b-1* knockdown on *Pt-hb* expression and suggest that although *Pt-Sox21b-1* is not required for *Pt-Msx1* activation, it is necessary for splitting of the broad L2–L4 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 organizing the prosomal cells into segments.

Clonal Analysis of *Pt-Sox21b-1* Knockdown in Early Stages of Segmentation

Since pRNAi mediated knockdown of *Pt-Sox21b-1* can result in severe phenotypic effects during early embryogenesis, it is sometimes difficult to ascertain direct effects of *Pt-Sox21b-1* on segmentation gene expression (Paese, Schönauer, et al. 2018). To address this problem, we used eRNAi to induce *Pt-Sox21b-1* knockdown in small subsets of embryonic cells to analyze 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* on injected embryos (fig. 6). We observed that *Pt-Sox21b-1* expression was reduced in the cells of all clones obtained ($n = 12$) (fig. 6D–I). In most embryos, the effect of eRNAi on *Pt-Sox21b-1* expression appeared to extend to cells adjacent to the clone region ($n = 11$) (fig. 6D–I); a nonautonomous effect was also previously reported for eRNAi with other genes (Kanayama et al. 2011; Schönauer et al. 2016). Interestingly, at stages 7 ($n = 3$) (fig. 6E) and 8.1 ($n = 7$) (fig. 6F and G), clones in the presumptive leg segments appeared to cause a distortion of the germ band along the antero-posterior axis that extended posteriorly into the opisthosoma as a consequence of fewer cells in the affected area compared with regions adjacent to the clone. 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.

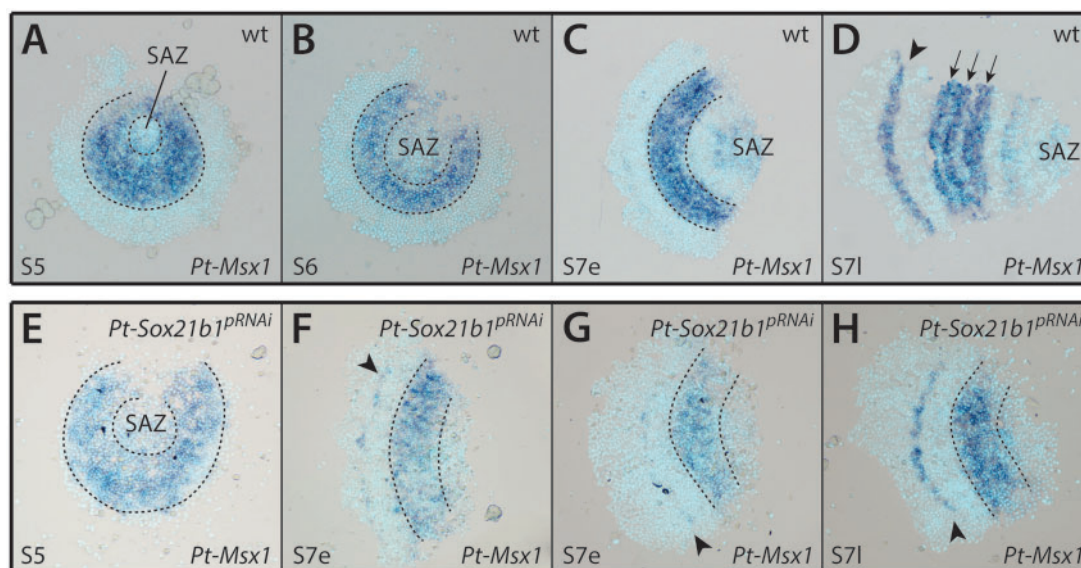


Fig. 5. Effect of *Pt-Sox21b-1* pRNAi knockdown on *Pt-Msx1* expression. *Pt-Msx1* is first expressed at stage 5 (A) in the presumptive region of the L2–L4 segments (dashed lines), forming a band of expression at stage 6 (B). At early stage 7 (C), two additional domains of expression arise in the SAZ and in a faint stripe corresponding to the presumptive head segments. This anterior stripe of expression becomes broader and more distinct by late stage 7 (D, black arrowhead) and the L2–L4 domain starts splitting into three stripes corresponding to each leg segment (black arrows). Faint expression is still present in the SAZ and a faint stripe can be detected in the first opisthosomal segment. Expression of *Pt-Msx1* in stage 5 *Pt-Sox21b-1* pRNAi embryos (E) is seemingly unaffected. At early stage 7 (F and G), *Pt-Sox21b-1* knockdown does not affect the band of expression in the L2–L4 presumptive region (dashed lines). However, this domain does not split into stripes in late stage 7 (H; dashed lines). The anterior stripe of expression is unaffected in stage 7 knockdown embryos (F–H, black arrowhead), although *Pt-Msx1* expression in the SAZ is lost. Anterior is to the left in all images except A and B. All embryos are flat mounted.

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 *Pt-Delta* or *Pt-caudal* via double fluorescent ISH. We chose these genes because *Pt-Dl*, like *Pt-Sox21b-1*, is expressed in the developing prosoma as well as dynamically in the opisthosoma in a similar pattern to *Pt-cad* (Oda et al. 2007; McGregor et al. 2008; Schönauer et al. 2016). Moreover, although expression of both *Pt-Dl* and *Pt-cad* is lost upon *Pt-Sox21b-1* pRNAi knockdown, it was unknown where and when the dynamic expression of these two key segmentation genes overlapped with that of *Pt-Sox21b-1* (Paese, Schönauer, et al. 2018).

Pt-Dl and *Pt-Sox21b-1* are coexpressed in the prosoma at stage 7 (fig. 7A and A''), but their expression dynamics are largely out of phase in the SAZ (fig. 7A' and A''). At stage 8.1, *Pt-Sox21b-1* is strongly expressed in the anterior SAZ (fig. 7B'), ubiquitously in the prosoma and in a stripe domain in the forming head lobes (fig. 7B and B'). At this stage, *Pt-Dl* expression in the anterior SAZ and the cheliceral/pedipalpal region does not overlap with *Pt-Sox21b-1* expression (fig. 7B). Overall, these double ISH experiments show that *Pt-Dl* and *Pt-Sox21b-1* may work together to specify prosomal segments but they are out of phase with each other in the SAZ.

In the case of *Pt-cad*, at late stage 6, the expression of this gene overlaps with *Pt-Sox21b-1* in the SAZ (fig. 8A' and 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. 8B and B''), the anterior SAZ expression of *Pt-cad* appears even stronger and, as was the case with *Pt-Dl*, *Pt-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. 8C and C''). *Pt-cad* is also expressed in lateral regions of the anterior SAZ (fig. 8C''), as well as in a lower cell layer of the fourth walking leg segment, but these domains do not overlap with *Pt-Sox21b-1* expression (fig. 8C). Taken together, the relative expression of *Pt-Sox21b-1* and *Pt-cad* suggest that 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 but the role of *Pt-cad* is restricted to sequential production of opisthosomal segments (Schönauer et al. 2016).

Clonal Analysis of the Effect of *Pt-Sox21b-1* Knockdown on *Pt-Dl* Expression

We found that *Pt-Sox21b-1* and *Pt-Dl* overlap in their expression in the prosoma but they are out of phase in the developing opisthosoma (fig. 7). The loss of *Pt-Dl* expression during the development of both tagmata when *Pt-Sox21b-1* is knocked down using pRNAi suggests that, although prosomal

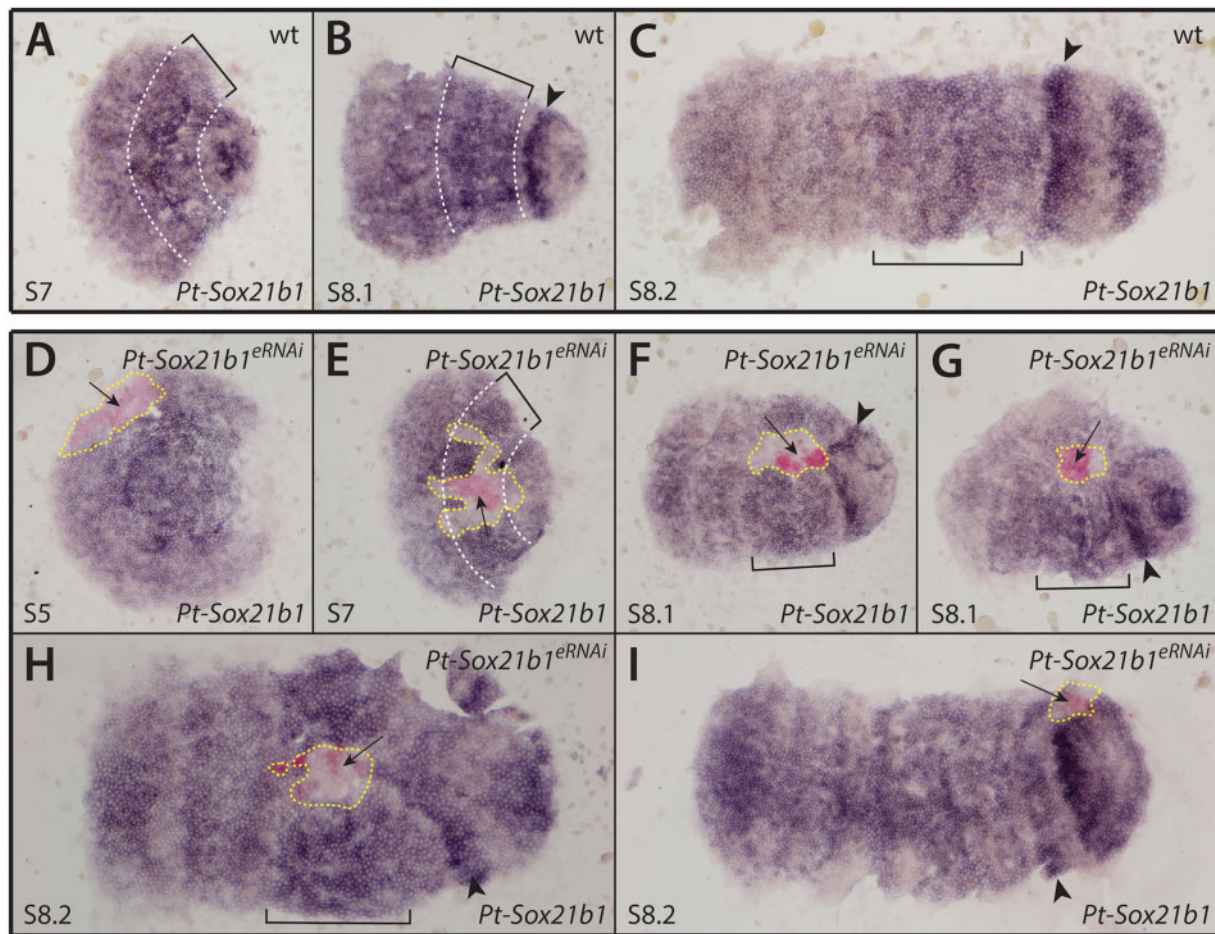


FIG. 6. Effect of *Pt-Sox21b-1* eRNAi knockdown in early segmentation. (A–C) Expression pattern of *Pt-Sox21b-1* in stages 7 (A), 8.1 (B), and 8.2 (C) wild-type embryos. (D–I) Expression pattern of *Pt-Sox21b-1* in stages 5 (D), 7 (E), 8.1 (F, G), and 8.2 (H, I) *Pt-Sox21b-1* eRNAi embryos. Anterior is to the left in all images. All embryos are flat mounted. Dashed white lines and brackets mark the presumptive L2–L4 segments. Arrowheads: presumptive O1 segment. Arrows indicate the knockdown clones, and yellow dashed lines indicate the area affected by eRNAi knockdown.

expression of *Pt-Dl* is dependent on *Pt-Sox21b-1*, perhaps even directly, opisthosomal expression of *Pt-Dl* is only lost indirectly in *Pt-Sox21b-1* pRNAi embryos because the SAZ does not develop properly (Paese, Schönauer, et al. 2018). To test this further, we carried out ISH for *Pt-Dl* in embryos with *Pt-Sox21b-1* eRNAi clones.

At stage 7, *Pt-Dl* is expressed in the presumptive cheliceral/pedipalpal segment, leg-bearing segments L2–L4, and in the SAZ (fig. 9A). At this stage, *Pt-Sox21b-1* knockdown leads to the loss of *Pt-Dl* expression in the developing L2–L4 segments ($n = 5$) (fig. 9B–F). Furthermore, we also observed that the *Pt-Sox21b-1* knockdown again leads to distortion of the prosomal tissue, which is apparent in constricted *Pt-Dl* expression surrounding the *Pt-Sox21b-1* knockdown clone area (fig. 9B, E, and F). During stage 8, *Pt-Dl* expression in *Pt-Sox21b-1* knockdown clones in the head region appeared normal ($n = 4$) (fig. 10B–D and supplementary fig. S13, Supplementary Material online). However, we again observed loss of *Pt-Dl* expression in the L2–L4 segments as well as distortion of the surrounding tissue at this stage ($n = 3$) (Fig. 10B and D and supplementary fig. S13, Supplementary Material online).

These results suggest that *Pt-Dl* expression in L2–L4 is dependent on *Pt-Sox21b-1*, and moreover, that the physical organization of these cells into separate prosomal segments by this Sox gene is regulated via *Pt-Dl*. Unfortunately, we were unable to obtain clones in the opisthosoma but it is likely that expression of *Pt-Dl* in this tissue only indirectly requires *Pt-Sox21b-1* to regulate the formation of the SAZ.

Discussion

The Expanded Sox Repertoire of Arachnoplumonates

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 (Paese, Leite, et al. 2018). We find that all arachnoplumonates we analyzed have at least two copies of each Sox gene family, except for the Sox 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*, and *SoxF* ohnologs after the WGD in the ancestor of arachnoplumonates that was not shared with harvestmen and ticks. Our data indicate that *SoxE* may have

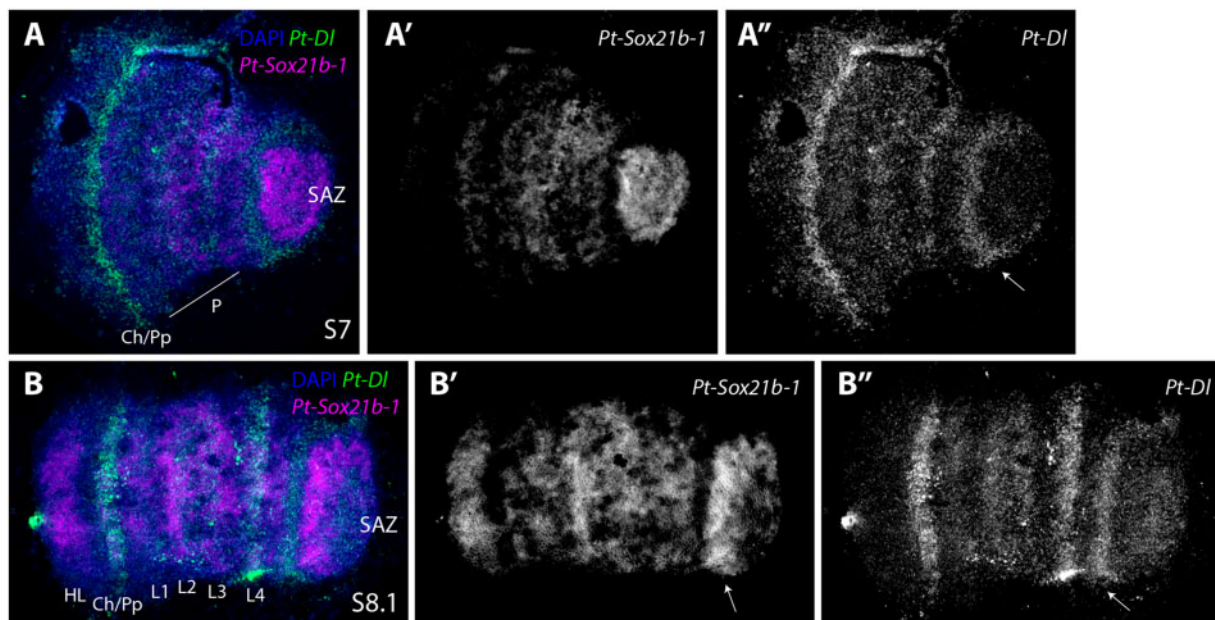


FIG. 7. Expression of *Patasteatoda tepidariorum* *Sox21b-1* and *Df*. Double-fluorescent in situ of *Pt-Sox21b-1* and *Pt-Df* in stages 7 (A–A'') and 8.1 (B–B'') embryos. At stage 7 *Pt-Sox21b-1* is expressed strongly in the SAZ and throughout the forming prosoma (A'). *Pt-Df* is also expressed in the anterior SAZ (white arrow in A''), but this expression does not overlap with *Pt-Sox21b-1* (white arrow in A''). *Pt-Df* and *Pt-Sox21b-1* are coexpressed in the forming prosoma at stage 7 (A). *Pt-Df* also exhibits a strong expression domain in the developing cheliceral–pedipalpal region (5 A''). At stage 8.1, *Pt-Sox21b-1* is strongly expressed in the anterior SAZ (arrow in B'), ubiquitously in the prosoma and in a stripe domain in the forming head lobes (B, B'). At this stage, *Pt-Df* is expressed in the anterior SAZ (white arrow in B''), 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. P, prosoma; Ch, chelicerae; Pp, pedipalps; HL, head lobe; L1–L4, walking legs 1–4.

been duplicated in arachnids prior to the arachnoplumate WGD because there are two copies in all arachnids surveyed including the tick and harvestman. The three copies of *SoxE* genes in *C. sculpturatus* and the two whip spider species could represent retention of a further copy after WGD or lineage-specific duplications. We suggest that there may have been subsequent lineage-specific duplication of *SoxN* in *C. sculpturatus*, and *Sox21b* in *P. tepidariorum*, although we cannot completely exclude that these are true ohnologs and only one was retained in other lineages. Analysis of additional lineages may help to resolve this. 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, Schöner, et al., 2018). Our survey also suggests that there have been lineage-specific duplications of *SoxC* and *SoxD* genes in the spiders *P. amentata* and *M. muscosa*, respectively, and perhaps other families in *C. sculpturatus* (fig. 1). 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 (Schwager et al. 2017; Leite et al. 2018). Interestingly, the main fate of retained duplicated *Sox* genes in vertebrates appears to be subfunctionalization, although there are likely cases of neofunctionalization, for example, in teleosts (De Martino et al. 2000; Cresko et al. 2003; Klüver et al. 2005; Voldoire et al. 2017).

To evaluate the fates of retained spider *Sox* ohnologs, we compared their expression during embryogenesis with 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 expression and roles of these genes in other arthropods (Janssen et al. 2018). 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. Comparison of *Sox21b* 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.

Pt-Sox21b-1 Is Used in Different GRNs Underlying Prosomal and Opisthosomal Segmentation

Pt-D, *Pt-Sox21a-1*, and *Pt-SoxD-2*, like *Pt-Sox21b-1*, are also expressed in patterns which could indicate that they play a role in segmentation. Indeed, it is intriguing that *Pt-D* and *Pt-Sox21a-1* are both expressed in L1 and this segment is sometimes retained after *Pt-Sox21b-1* pRNAi (Paese, Schöner, et al. 2018). However, only knockdown of *Pt-Sox21b-1* has a detectable phenotypic effect. This shows that although *Pt-Sox21b-1* is required for segmentation, it suggests that these other *Sox* genes may only act partially redundantly

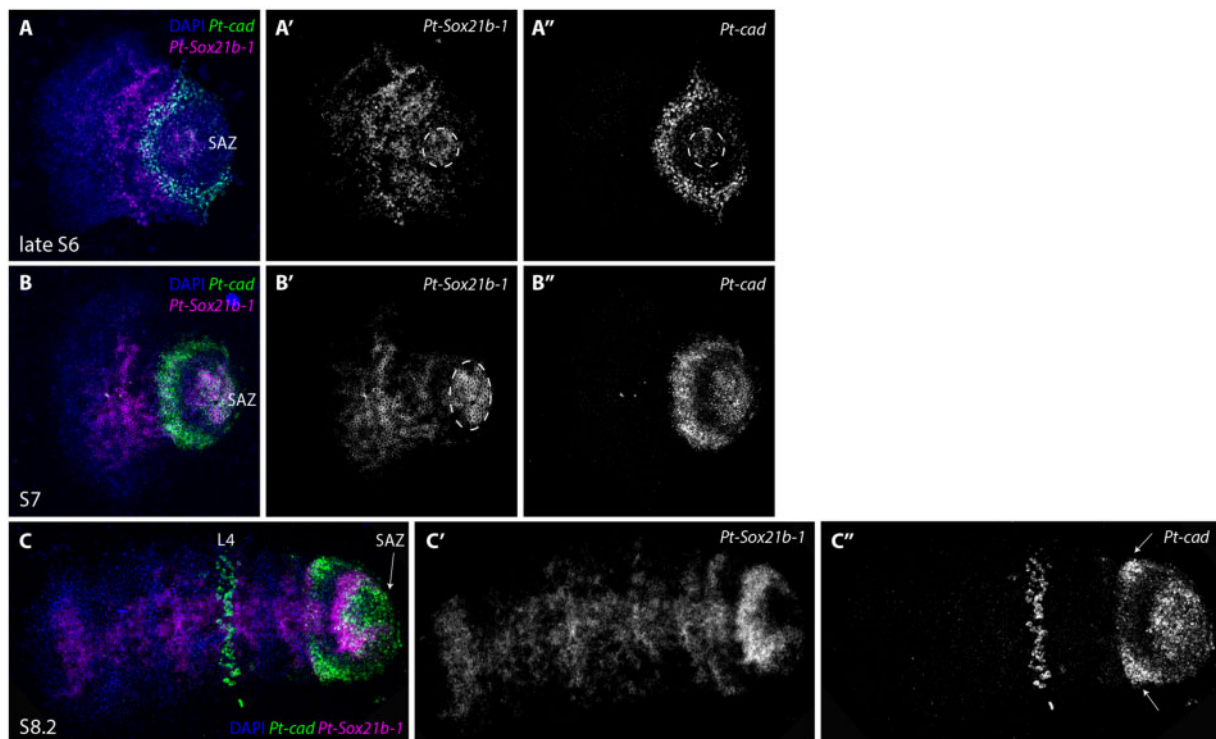


FIG. 8. Expression of *Parasteatoda tepidariorum* *Sox21b-1* and *cad*. Double-fluorescent in situ of *Pt-Sox21b-1* and *Pt-cad* in stages 6 (A–A''), 7 (B–B''), and 8.1 (C–C'') embryos. (D–F''). At late stage 6, *Pt-Sox21b-1* is expressed in a faint circular domain in the posterior SAZ overlapping with *Pt-cad* (dashed circle in A''), although expression of these two genes does not overlap in the anterior SAZ (A, A''). *Pt-Sox21b-1* 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 partially overlaps with *Pt-Sox21b-1* (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 fourth walking leg segment but these domains do not overlap with *Pt-Sox21b-1* expression (C). Anterior is to the left in all images. L4, fourth walking leg.

(or compensate for each other) as has been suggested for Sox genes in other animals (Wegner 1999; Phochanukul and Russell 2010; Heenan et al. 2016). It is also possible that the pRNAi knockdown of *Pt-D*, *Pt-Sox21a-1*, and *Pt-SoxD-2* may not have been fully penetrant. Note that the knockdown of expression of the focal gene in pRNAi experiments in *P. tepidariorum* in the absence of a phenotype is difficult to conclusively verify (see Oda et al. 2007) and further analysis of the function of these genes using eRNAi might be needed.

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, Leite, et al., 2018) (fig. 11). Here we show that *Pt-Sox21b-1* does not appear to be necessary for the activation of *Pt-Dll*, *Pt-hb*, and *Pt-Msx1*. This suggests that *Pt-Sox21b-1* could repress *Pt-Dll*, *Pt-hb*, and *Pt-Msx1* to generate stripes from broader expression domains during prosomal segmentation and/or acts to regulate the division and viability of cells and organize them into segments (fig. 11). We suggest that these aspects of *Pt-Sox21b-1* function in the formation of the limb-bearing segments likely involve direct regulation of parts of *Pt-Dl* expression in this tagma or at least indirectly via other factors. This is consistent with the loss of prosomal segments observed following *Pt-Sox21b-1* pRNAi and the disruption of these segments in *Pt-Dl* pRNAi embryos (Oda

et al., 2007). Intriguingly, it was recently shown that *Pt-hedgehog* represses *Pt-Msx1* during segmentation, but it is unclear whether this role of *Pt-hh* is dependent on *Pt-Sox21b-1*, and potentially *Pt-Dl*, or not (Akiyama-Oda and Oda 2020). This shows that further work is needed to decipher how *Pt-Dl*-mediated regulation of prosomal segments is integrated with *Pt-hh*, the gap-like functions of *Pt-hb*, *Pt-Dll*, and *Pt-Msx1* as well as other factors involved in cell organization in the germband such as Toll genes (Schwager et al. 2009; Pechmann et al. 2011; Benton et al. 2016; Akiyama-Oda and Oda 2020).

Pt-Sox21b-1 also regulates formation of the SAZ and the subsequent addition of opisthosomal segments (fig. 11). We further characterized 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* in some SAZ cells, but that its dynamic expression is out of phase with *Pt-Dl*. We therefore interpret the previously observed loss of *Pt-Dl* upon *Pt-Sox21b-1* pRNAi in the SAZ as a side effect of the perturbed development of this tissue rather than regulation of *Pt-Dl* by *Pt-Sox21b-1*. 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 clock-like mechanisms for sequential production of segments from the SAZ requires further investigation.

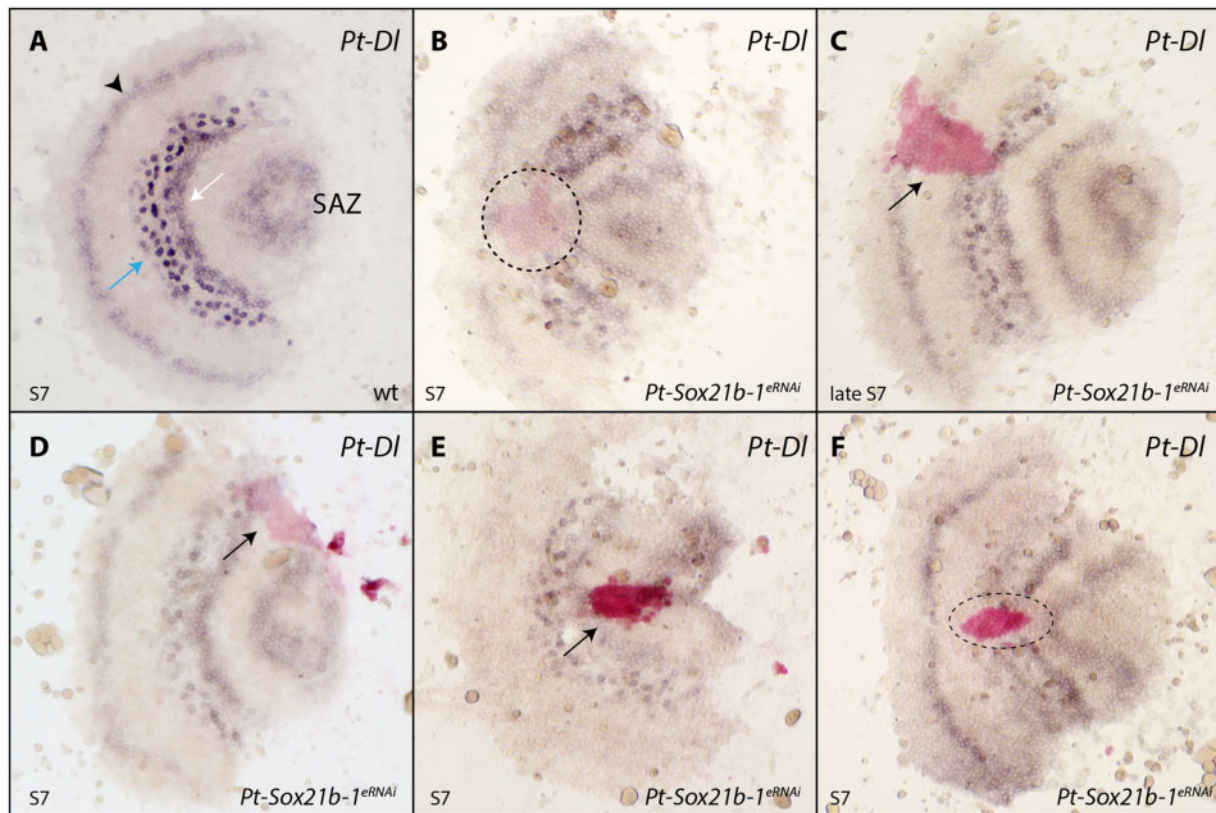


FIG. 9. Effect of *Pt-Sox21b-1* eRNAi knockdown on *Pt-Dl* expression during early prosomal development. At stage 7 (A), *Pt-Dl* expression clears from the center of the SAZ, is expressed in two adjacent domains in the forming prosoma, an anterior salt-and-pepper pattern (blue arrow), and a posterior solid band of expression (white arrow), and is also expressed in the presumptive pedipalpal segment (black arrowhead). *Pt-Sox21b-1* knockdown causes the loss of *Pt-Dl* expression in the prosoma (B–F) as well as the distortion of *Pt-Dl* expression domains in the tissue surrounding the *Pt-Sox21b-1* clones (B, E, F). Anterior is to the left. All embryos are flat mounted. Black arrows point at knockdown clones marked by pink staining (C, D, E) and dashed lines the estimated extent of the entire knockdown region surrounding the stained area, including potential nonautonomous affects (B, F). Pp, presumptive pedipalpal segment; L1–L4, presumptive L1–L4 segments; O1, presumptive O1 segment; O2, presumptive O2 segment.

The simultaneous prosomal segment production and sequential opisthosomal segment addition in spiders have similarities to the simultaneous and sequential mechanisms used in short-germ and long-germ insects, respectively (Clark et al. 2019). Indeed, many of the same genes are involved in segmentation in spiders and insects including SoxB genes, as well as Wnt signaling, *cad* and pair-rule gene orthologs (Clark and Peel 2018; Clark et al. 2019). However, further work is needed to better understand the gene regulatory networks underlying both prosomal and opisthosomal segmentation in spiders, including the role of *Sox21b-1* and potentially other Sox genes. It is intriguing that although both mechanisms in spiders are regulated by *Sox21b-1*, many of the other components differ between segment formation in these two tagmata. For example, *cad*, *even-skipped*, and *runt* are involved in sequential but not simultaneous segmentation in *P. tepidariorum* (Schönauer et al. 2016). Therefore, while better understanding the oscillatory mechanisms used in the SAZ of spiders might add to recent new insights into understanding of how simultaneous segmentation may have evolved from sequential segmentation in insects (Clark 2017), much remains to be discovered regarding how the gap gene-like mechanism employed in the

spider prosoma relates to mechanisms found insects and other arthropods.

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-h light-dark cycle. Embryos were staged according to Mittmann and Wolff (2012) and fixed as described in Akiyama-Oda and Oda (2003). Embryos were collected and stored in RNAlater (Invitrogen) from captive-mated females of the amblypygids *C.*, at 1 day, 1 month, and 2 months after the appearance of the egg sacs, and *E. bacillifer*, at ~30% of embryonic development (Harper et al. 2020). Mixed-stage embryos were collected from a female *P. amentata* (collected in Oxford) and *M. muscosa* (kindly provided by Philip Steinhoff and Gabriele Uhl) and stored in RNAlater (Harper et al. 2020). *Ph. opilio* embryos were collected and prepared as previously described (Sharma et al. 2012).

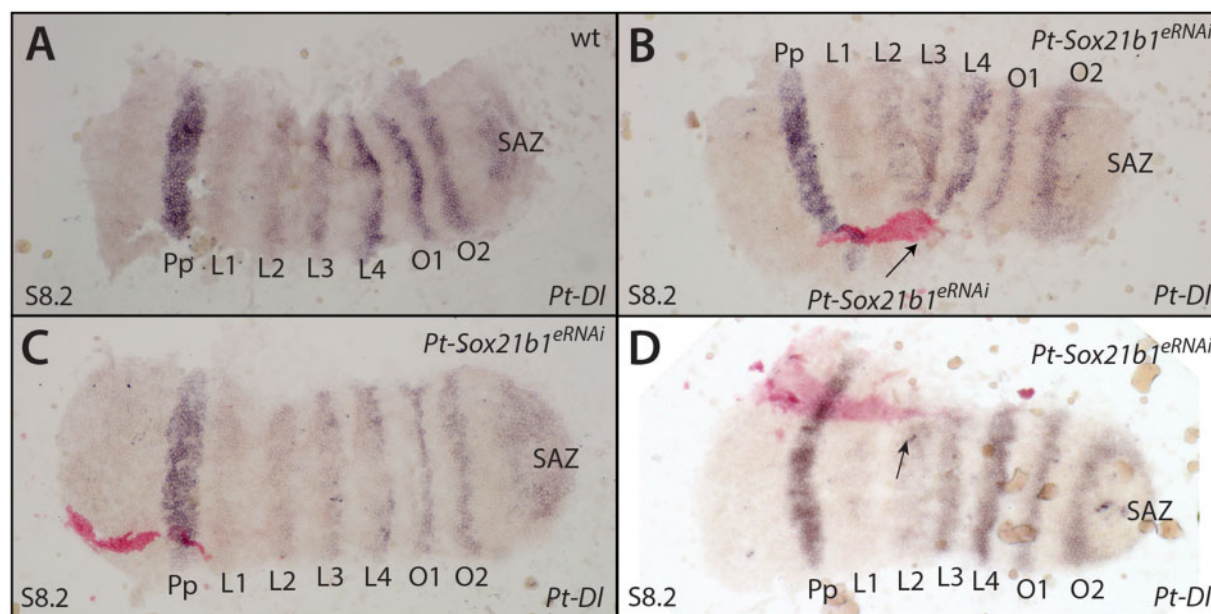


FIG. 10. Effect of *Pt-Sox21b-1* eRNAi knockdown on *Pt-Dl* expression during later prosomal development. At stage 8.2 (A), *Pt-Dl* expression is restricted to segmental stripes of the prosoma of varying strength, the opisthosoma, and the SAZ (A). *Pt-Sox21b-1* knockdown clones in the developing head do not appear to affect *Pt-Dl* expression (B–D). Anterior is to the left. All embryos are flat mounted. Knockdown clones are marked by pink staining and black arrows point at tissue constriction caused by *Pt-Sox21b-1* knockdown. Pp, presumptive pedipalpal segment; L1–L4, presumptive L1–L4 segments; O1, presumptive O1 segment; O2, presumptive O2 segment.

Transcriptomics

RNA was extracted from the embryos of *C. acosta*, *E. bacillifer*, *P. 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 bp 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/FilterUncorrectablePEfastq.py>, last accessed March 2021). Reads were trimmed for adapter contamination using TrimGalore! (available at <https://github.com/FelixKrueger/TrimGalore>, last accessed March 2021) 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 November 2, 2019; ten species, 2,934 BUSCOs; Harper et al. 2020). These RNA-Seq data sets are available in the BioSample database under accessions SAMN18203375, SAMN18203376, SAMN18203377, and SAMN18203378.

Identification of Sox Genes

To identify the Sox gene repertoires of *C. acosta*, *E. bacillifer*, *M. muscosa*, *P. amentata*, *C. sculpturatus*, *Ph. opilio*, *I. scapularis*, and *S. 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; *P. amentata*—this study; *C. sculpturatus*—PRJNA422877; *Ph. opilio*—PRJNA236471; *I. scapularis*—PRJNA357111; and *S. maritima*—PRJNA20501), using the HMG domain protein sequences previously identified in *P. tepidariorum*, *S. mimosarum*, and *D. melanogaster* (Paese, Leite, et al. 2018). Predicted protein sequences were obtained using the open reading frame (ORF) finder NCBI online tool (<https://www.ncbi.nlm.nih.gov/orf-finder/>; 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-vanncbi.nlm.nih.gov/blast/smart-blast/>; 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, Supplementary Material](#) online.

Sequence identity was further verified through the construction of maximum likelihood trees ([supplementary figs. S1 and S2, Supplementary Material](#) online). Protein sequence alignments of HMG domains from *C. acosta*, *E. bacillifer*, *M. muscosa*, *P. amentata*, *S. mimosarum* (Paese, Leite, et al. 2018), *P. tepidariorum* (Paese, Leite, et al. 2018), *C. sculpturatus*, *Ph. opilio*, *I. scapularis*, *S. maritima*, *G. marginata* (Janssen et al. 2018), *T. castaneum* (Janssen et al. 2018), *D. melanogaster* (Paese, Leite, et al. 2018), 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

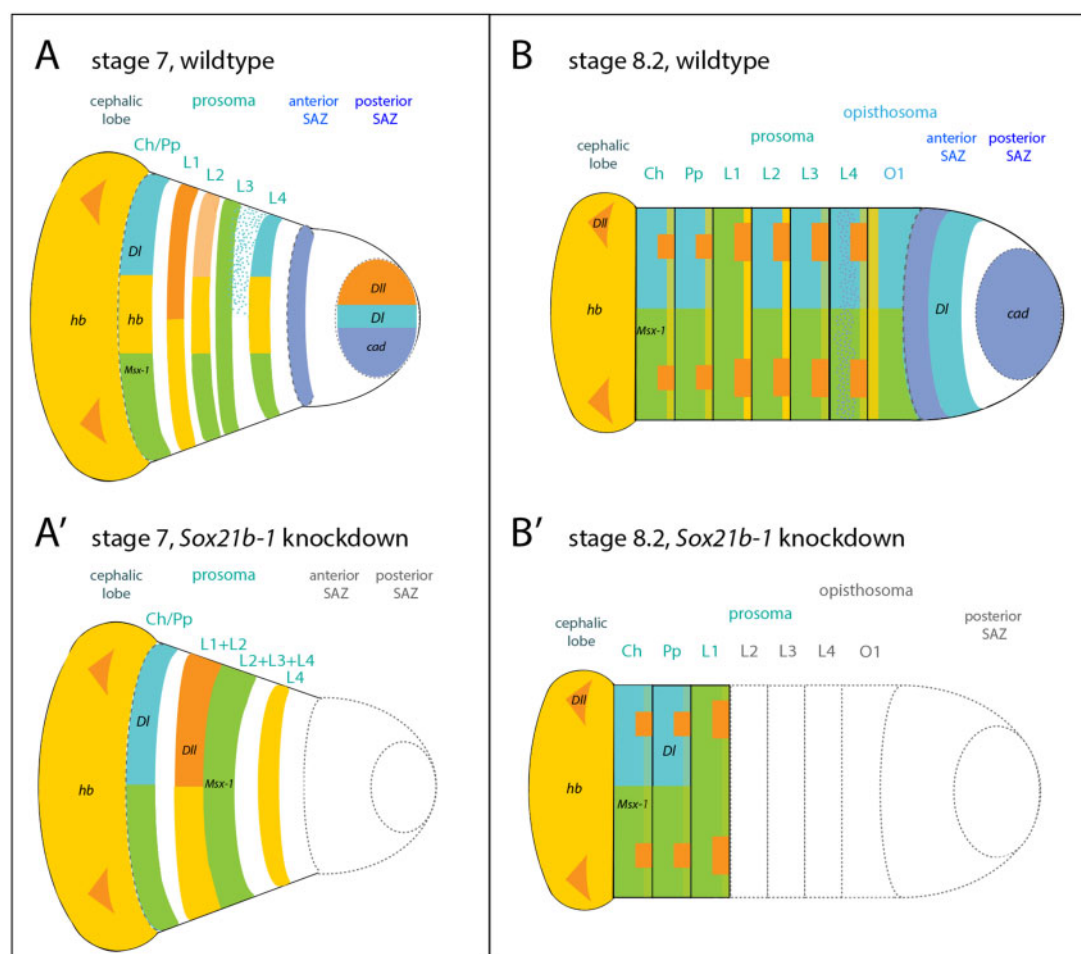


FIG. 11. Summary of the effect of *Sox21b-1* knockdown on spider segmentation. (A, B) wild-type embryos. (A', B') of *Sox21b-1* knockdown embryos. At stage 7, *cad* (blue) is expressed in a circular domain in the posterior SAZ as well as in the anterior SAZ in a stripe (A). *Df* (turquoise) is expressed in a circular domain in the posterior SAZ, in a solid domain in L4, adjacent to an anterior salt-and-pepper domain, as well as in the cheliceral/pedipalpal segment (A). At this stage, *hb* (yellow) is expressed in the prospective L1, L2, and L4, the cheliceral/pedipalpal segment and the cephalic lobe (A). *Dll* (orange) is also expressed in the posterior SAZ, L1, and in two distinct domains in the cephalic lobe (A). *Msx1* (green) is expressed in the cheliceral/pedipalpal segment, and the developing L2–L4 segments (A). In *Sox21b-1* knockdown embryos, the SAZ is missing (dashed gray lines) and expression of *Df*, as well as *hb* expression do not split into stripes, (L1 + L2) (A'). *Msx1* expression does not resolve into stripes but remains as a solid domain (L2 + L3 + L4). *Df* expression in L4 and the SAZ is lost and only remains in the cheliceral/pedipalpal segment (A'). At stage 8.2, *cad* is expressed in the posterior SAZ, in a stripe domain in the anterior SAZ and in a salt-and-pepper pattern in L4 (B). *Df* expression can be observed in the anterior SAZ, posterior of *cad* expression and segmentally in O1, L2–L4 and the cheliceral/pedipalpal segment (B). *Msx1* is expressed segmentally in all prosomal and opisthosomal segments at stage 8.2 (Leite et al. 2018) (B). *Df* is expressed in the developing prosomal limb buds, and two distinct domains in the cephalic lobe (Pechmann et al. 2011) (B). *hb* is expressed in the cephalic lobe, in all prosomal segments, whereby expression in the cheliceral, pedipalpal, L3, and L4 appear fainter, compared with expression in L1 and L2 (B). At stage 8.2, the SAZ as well as opisthosomal and prosomal segments up to L1 are lost in a *Sox21b-1* knockdown embryo (gray dashed lines) (B'). *Sox21b-1* knockdown results in downregulation of *hb* expression in the remaining prosomal segments and only faint *hb* (light yellow) stripes remain (B'). Anterior *Msx1*, *hb*, *Df*, and *Dll* expression is still observed in the *Sox21b-1* knockdown embryo (B').

algorithm (1,000 replicates; Stamatakis et al. 2008). Bayesian phylogenetic analysis was performed using MrBayes v.3.2.7a with an LG + Γ substitution model (10,000,000 generations, default settings) (Ronquist et al. 2012). Alignments (Phylip files) and gene trees (tre files) are provided in supplementary files 2 and 4–7, [Supplementary Material](#) online. The resulting trees were visualized and processed in FigTree v1.4.4 (<http://tree.bio.ed.ac.uk/software/figtree/>, last accessed March 2021).

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 *Ph. 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>, last accessed March 2021), 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 8, Supplementary Material](#) online. The template for probe synthesis was generated through two rounds of standard PCR method using OneTaq 2x Master Mix (New England Biolabs): 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 for the sense probe. 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 labeling mix (Roche) or Fluorescein RNA labeling mix (Roche), according to manufacturer's guidelines.

In Situ Hybridization

Colorimetric 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-min washes in PBS-Tween-20 (0.02%) (PBS-T), and at step 18, the embryos were incubated for 30 min. Postfixation was followed by ethanol treatment to decrease background: Embryos were incubated for 10 min in inactivation buffer (75 g glycine, 600 μ l 1 N HCl, 50 μ l 10% Tween-20, and dH₂O 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 min in 50% ethanol in PBS-T, and finally washed twice with PBS-T. Embryos were then counterstained with DAPI (1:2,000; Roche) for ~20 min and stored in 80% glycerol in 1 \times PBS at 4 °C. Imaging was performed using a Zeiss Axio Zoom V.16. DAPI overlays were generated in Adobe Photoshop CS6.

Double fluorescent ISH protocol was modified from [Clark and Akam \(2016\)](#): Fixed embryos were gradually moved from methanol to PBS-T and washed for 15 min. Embryos were then transferred to hybridization buffer, hybridized overnight at 65 °C, and washed posthybridization as detailed in [Prpic et al. \(2008\)](#). Embryos were incubated in blocking solution (Roche) for 30 min and AP-conjugated anti-DIG (1:2,000; Roche) and POD-conjugated anti-FITC (1:2,000; Roche) added and incubated for 2 h. Tyramide biotin amplification (TSA Plus Biotin Kit, Perkin–Elmer) was performed for 10 min, followed by incubation for 90 min in streptavidin Alexa Fluor 488 conjugate (1:500; ThermoFisher Scientific). AP signal was visualized by Fast Red staining (Kem En Tec Diagnostics). Counterstaining with DAPI (1:2,000; Roche) was carried out for 5–10 min. 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. Images were processed using Adobe Photoshop CS6 and FIJI software.

Double-Stranded RNA Preparation

Synthesis of double-stranded RNA (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 μ g/ μ l 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 μ l 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 5–8.2 for *Pt-Sox21b-1* and stages 7–9.2 for *Pt-Sox21a-1*, *Pt-SoxD-2*, and *Pt-D* 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, Schönauer, et al. 2018](#)). The same range and approximate frequencies of three phenotypic classes were observed for all cocoons from injected females ([Paese, Schönauer, et al. 2018](#)). In class I embryos, all segments posterior to the first leg (L1) segment do not form properly. In class II embryos, only the head, cheliceral, and pedipalpal segments are formed, and the L1 segment is missing in addition to the missing segments of class I embryos. Class III embryos do not form a germ band, forming instead a disorganized cellular mass in the center of the germ disc. 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, Schönauer, et al. 2018](#)). For *Sox21a-1*, *SoxD-2*, and *Dichaete* parental RNAi, dsRNAs of 441, 785, and 675 bp, respectively, were used ([supplementary file 8, Supplementary Material](#) online).

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 consisting of 5 μ l of FITC-dextran, 5 μ l of biotin-dextran, and 2.5 μ l of dsRNA. Embryos were subsequently fixed at stages 5–8.2 of development. Visualization of eRNAi clones was achieved by detecting the coinjected biotin-dextran with the Vectastain ABC-AP kit (Vector Laboratories) after ISH, according to the manufacturer's protocol.

Supplementary Material

[Supplementary data](#) are available at *Molecular Biology and Evolution* online.

Acknowledgments

This study was funded in part by a Leverhulme Trust (Grant No. RPG-2016-234 to A.P.M. and A.S.), an NERC (Grant No. NE/T006854/1 to A.P.M. and L.S.R.), a Nigel Groome Studentship from Oxford Brookes University to L.B.G., a BBSRC DTP Studentship to A.H., a John Fell Fund (Grant No. 0005632) from the University of Oxford to L.S.R., a BBSRC (Grant No. BB/N007069/1 to S.R.), and an NSF CAREER (Grant No. IOS-1552610 to P.P.S.). We thank Philip Steinhoff and Gabriele Uhl for kindly providing *Marpissa muscosa* embryos.

Author Contributions

L.B.G., A.S., S.R., L.S.R., P.P.S., and A.P.M. designed this project; L.B.G., A.S., A.H., G.B., M.S., S.A., and L.S.R. performed the experimental work; L.B.G., A.S., S.R., A.H., S.A., L.S.R., P.P.S., and A.P.M. analyzed data; and L.B.G., A.S., and A.P.M. wrote the paper with the help of all other authors.

Data Availability

The data underlying this article are available in the BioSample database under accessions SAMN18203375, SAMN18203376, SAMN18203377, and SAMN18203378 or in its [Supplementary Material](#) online.

References

- Akiyama-Oda Y, Oda H. 2003. Early patterning of the spider embryo: a cluster of mesenchymal cells at the cumulus produces DPP signals received by germ disc epithelial cells. *Development* 130(9):1735–1747.
- 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):2347–2357.
- Akiyama-Oda Y, Oda H. 2020. Hedgehog signaling controls segmentation dynamics and diversity via *mxs1* in a spider embryo. *Sci Adv*. 6:32917677.
- Andrews S. 2010. *FastQC: A Quality Control Tool for High Throughput Sequence Data* [Online]. Available from: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>. Accessed March 2021.
- Benton MA, Pechmann M, Frey N, Stappert D, Conrads KH, Chen YT, Stamatakis E, Pavlopoulos A, Roth S. 2016. Toll genes have an ancestral role in axis elongation. *Curr Biol*. 26(12):1609–1615.
- Bowles J, Schepers G, Koopman P. 2000. Phylogeny of the SOX family of developmental transcription factors based on sequence and structural indicators. *Dev Biol*. 227(2):239–255.
- Clark E. 2017. Dynamic patterning by the *Drosophila* pair-rule network reconciles long-germ and short-germ segmentation. *PLoS Biol*. 15(9):e2002439.
- Clark E, Akam M. 2016. Odd-paired controls frequency doubling in *Drosophila* segmentation by altering the pair-rule gene regulatory network. *Elife* 5:e18215.
- Clark E, Peel AD. 2018. Evidence for the temporal regulation of insect segmentation by a conserved sequence of transcription factors. *Development (Cambridge)* 145:1–15.
- Clark E, Peel AD, Akam M. 2019. Arthropod segmentation. *Development (Cambridge)* 146(18):dev170480.
- Cresko WA, Yan YL, Baltrus DA, Amores A, Singer A, Rodríguez-Marí A, Postlethwait JH. 2003. Genome duplication, subfunction partitioning, and lineage divergence: *sox9* in stickleback and zebrafish. *Dev Dyn*. 228(3):480–489.
- De Martino S, Yan YL, Jowett T, Postlethwait JH, Varga ZM, Ashworth A, Austin CA. 2000. Expression of *sox11* gene duplicates in zebrafish suggests the reciprocal loss of ancestral gene expression patterns in development. *Dev Dyn*. 217(3):279–292.
- Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, et al. 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc*. 8(8):1494–1512.
- Harper A, Baudouin-Gonzalez L, Schoenauer A, Seiter M, Holzem M, Arif S, McGregor AP, Sumner-Rooney L. 2020. Widespread retention of ohnologs in key developmental gene families following whole genome duplication in arachnophiles. *bioRxiv*. doi: 10.1101/2020.07.10.177725.
- Heenan P, Zondag L, Wilson MJ. 2016. Evolution of the Sox gene family within the chordate phylum. *Gene* 575(2 Pt 2):385–392.
- Hilbrant M, Damen WG, McGregor AP. 2012. Evolutionary crossroads in developmental biology: the spider *Parasteatoda tepidariorum*. *Development* 139(15):2655–2662.
- Iwasaki-Yokozawa S, Akiyama-Oda Y, Oda H. 2018. Genome-scale embryonic developmental profile of gene expression in the common house spider *Parasteatoda tepidariorum*. *Data Brief*. 19:865–867.
- Janssen R, 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(1):88.
- Kamachi Y, Kondoh H. 2013. Sox proteins: regulators of cell fate specification and differentiation. *Development* 140(20):4129–4144.
- Kanayama M, Akiyama-Oda Y, Nishimura O, Tarui H, Agata K, Oda H. 2011. Travelling and splitting of a wave of hedgehog expression involved in spider-head segmentation. *Nat Commun*. 2:500.
- Klüver N, Kondo M, Herpin A, Mitani H, Schartl M. 2005. Divergent expression patterns of *Sox9* duplicates in teleosts indicate a lineage specific subfunctionalization. *Dev Genes Evol*. 215(6):297–305.
- Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol*. 35(6):1547–1549.
- Leite DJ, Baudouin-Gonzalez L, Iwasaki-Yokozawa S, Lozano-Fernandez J, Turetzek N, Akiyama-Oda Y, Prpic NM, Pisani D, Oda H, Sharma PP, et al. 2018. Homeobox gene duplication and divergence in arachnids. *Mol Biol Evol*. 35(9):2240–2253.
- McGregor AP, Pechmann M, Schwager EE, Feitosa NM, Kruck S, Aranda M, Damen WG. 2008. Wnt8 is required for growth-zone establishment and development of opisthosomal segments in a spider. *Curr Biol*. 18(20):1619–1623.
- McKimmie C, Woerfel G, Russell S. 2005. Conserved genomic organisation of Group B Sox genes in insects. *BMC Genet*. 6:26.
- 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.
- Oda H, Akiyama-Oda Y. 2020. The common house spider *Parasteatoda tepidariorum*. *EvoDevo*. 11:6.
- Oda H, Nishimura O, Hirao Y, Tarui H, Agata K, 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(12):2195–2205.
- Paese CLB, Leite DJ, Schöner A, McGregor AP, Russell S. 2018. Duplication and expression of Sox genes in spiders. *BMC Evol Biol*. 18(1):205.
- Paese CLB, Schoenauer A, Leite DJ, Russell S, McGregor AP. 2018. A SoxB gene acts as an anterior gap gene and regulates posterior segment addition in a spider. *Elife* 7:e37567.
- Pechmann M, Khadjeh S, Turetzek N, McGregor AP, Damen WG, Prpic NM. 2011. Novel function of Distal-less as a gap gene during spider segmentation. *PLoS Genet*. 7(10):e1002342.
- Phochanukul N, Russell S. 2010. No backbone but lots of Sox: invertebrate Sox genes. *Int J Biochem Cell Biol*. 42(3):453–464.
- Prpic N-M, Schoppmeier M, Damen WG. 2008. Whole-mount in situ hybridization of spider embryos. *CSH Protoc*. 2008:pdb.prot5068.
- Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. 2012. MrBayes 3.2:

- efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol.* 61(3):539–542.
- Russell SRH, Sanchez-Soriano N, Wright CR, Ashburner M. 1996. The Dichaete gene of *Drosophila melanogaster* encodes a SOX-domain protein required for embryonic segmentation. *Development* 122(11):3669–3676.
- Schepers GE, Teasdale RD, Koopman P. 2002. Twenty pairs of Sox. *Dev Cell.* 3(2):167–170.
- Schönauer A, Paese CLB, Hilbrant M, Leite DJ, Schwager EE, Feitosa NM, Eibner C, Damen WGM, McGregor AP. 2016. The Wnt and Delta-Notch signalling pathways interact to direct even-skipped expression via caudal during segment addition in the spider *Parasteatoda tepidariorum*. *Development* 143(13):2455–2463.
- Schwager EE, Meng Y, Extavour CG. 2015. Vasa and piwi are required for mitotic integrity in early embryogenesis in the spider *Parasteatoda tepidariorum*. *Dev Biol.* 402(2):276–290.
- Schwager EE, Pechmann M, Feitosa NM, McGregor AP, Damen WG. 2009. Hunchback functions as a segmentation gene in the spider *Achaearanea tepidariorum*. *Curr Biol.* 19(16):1333–1340.
- Schwager EE, Sharma PP, Clarke T, Leite DJ, Wierschin T, Pechmann M, Akiyama-Oda Y, Esposito L, Bechsgaard J, Bilde T, et al. 2017. The house spider genome reveals an ancient whole-genome duplication during arachnid evolution. *BMC Biol.* 15(1):62.
- Seppely M, Manni M, Zdobnov EM. 2019. BUSCO: assessing genome assembly and annotation completeness. *Methods Mol Biol.* 1962:227–245.
- Setton EVW, Sharma PP. 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 PP, Schwager EE, Extavour CG, Giribet G. 2012. Hox gene expression in the harvestman *Phalangium opilio* reveals divergent patterning of the chelicerate opisthosoma. *Evol Dev.* 14(5):450–463.
- Song L, 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(9):1312–1313.
- Stamatakis A, Hoover P, Rougemont J. 2008. A rapid bootstrap algorithm for the RAxML Web servers. *Syst Biol.* 57(5):758–771.
- Voltaire E, Brunet F, Naville M, Volff JN, Galiana D. 2017. Expansion by whole genome duplication and evolution of the sox gene family in teleost fish. *PLoS One* 12(7):e0180936.
- Wegner M. 1999. From head to toes: the multiple facets of Sox proteins. *Nucleic Acids Res.* 27(6):1409–1420.
- Wilson MJ, Dearden PK. 2008. Evolution of the insect Sox genes. *BMC Evol Biol.* 8:120.