# Investigating the roles of Hes and Sox genes during embryogenesis of the spider *P. tepidariorum*

---

**Christian Louis Bonatto Paese**

This thesis was submitted in partial fulfillment of the requirements of the award of Doctor of Philosophy

June 2018

**“Our greatest glory is not in never falling, but in rising every time we fall”**

*Confucius*

Aquele uísque

“Lembro-me daqueles dias,
Em que assistíamos ao pica-pau
E rachavamos o bico dando risada de algumas passadas.
Lembro-me de que possuíamos aspirações e sonhos selvagens. Cada
qual, à sua maneira, pôs-se a caçar seu horizonte.
Sabíamos que nenhuma estrada se emenda no passado,
Todas se amarram apenas no porvir,
E, só com muita sorte, nos achamos novamente.
Mas a memória, ah!
Essa é uma cruel concubina,
Pois nunca nos esquecemos daqueles dias.
Eles voltam em espasmos mistos de tristeza e alegria.
No saldo da vida,
Parece que atingimos muito do que poderíamos,
Mas sempre ficamos na dúvida  De uma dose
daquele uísque:
Velho olho vermelho.”

Pablo Jamilk

Essa tese é o resultado de 3 anos que me trouxeram um incrivel auto-aprendizado.
Todo esse esforço é dedicado a minha família – em especial aos meus sobrinhos
Leonardo e Leticia.

## Table of Contents

## Abstract

One goal of evolutionary developmental biology is to understand the mechanisms that underlie the regulation and evolution of embryogenesis among different species. One approach to answer these questions is to study and compare the expression and function of “toolbox” genes in different lineages. Members of two conserved gene families, Hes and Sox, are involved in both segmentation and neurogenesis in different metazoans. Genes such as *hairy* and *Dichaete* are involved in the segmentation cascade in *Drosophila melanogaster*, whereas *deadpan* and *Sox21B* acts in the regulation of the neurogenic precursors in the same species. To broaden our understanding of the evolution and function of these genes, for this project, I studied them in the commonhouse spider *P. tepidariorum*: an emergent organism to studies in comparative embryology and genetics, and which has experienced a whole genome duplication in its evolutionary history. I characterized the expression of four Hes-like genes and fifteen Sox genes in this species. Briefly, *hairy* and *Sox21B-1* are expressed in prosoma as well as in the posterior segment addition zone and forming segments. *deadpan*, *hey*, *side*, *SoxNeuro*, *Sox C-1* and *Sox D-2* are expressed in the central nervous system, and *Sox E-2* and *Sox F-2* in the mechanoceptors of the walking limbs and peripheral nervous system. I also carried out functional analysis of *Sox21b-1* and demonstrated its involvement in the gene regulatory network that is responsible for the posterior segment addition in the spider, as well as a striking relation with secondary layers formation. Finally, I generated new resources to help develop new functional tools in *P. tepidariorum* in order to make genomic manipulations in this species, as well as to improve techniques such as live imaging and in situ hybridization. Taken together my research embraced different aspects

of the spider embryogenesis and prove that the chelicerates are emerging research systems to the study of evolutionary developmental biology.

## List of Abbreviations

| | |
|---|---|
| **°C** | degrees Celsius |
| **µg** | microgram |
| **µl** | microliter |
| **aa** | amino acid |
| **bHLH** | Basic Helix-Loop-Helix |
| **BLAST** | Basic Local Alignment Search Tool |
| **BMP** | Bone Morphogenetic Protein |
| **bp** | base pairs |
| **Cas9** | CRISPR Associated Protein 9 |
| **cDNA** | Complementary DNA |
| **Ch** | cheliceres |
| **CNS** | Central Nervous System |
| **CRISPR** | Clustered regularly interspaced palindromic repeats |
| **C-terminal** | carboxy-terminal |
| **DAPI** | 4-6-diamidino-2-phenylindol |
| **ddH2O** | double-distilled water |
| **DIG** | Digoxigenin |
| **DNA** | deoxyribonucleic acid |
| **dsRNA** | double stranded deoxyribonucleic acid |
| **EMT** | Epithelial-to-mesenchymal transition |
| **eRNAi** | embryonic RNAi |
| **FGF** | Fibroblastic Growth Factor |
| **h** | hour |
| **HMG** | High Mobility Group |
| **HR** | homologous recombination |
| **Im** | inner yolk mass |
| **INT/BCIP** | 2-[4-iodophenyl]-3-[4-nitrophenyl]-5-phenyltetrazolium chloride)/ 5-bromo4-chloro-3-indolyl-phosphate |
| **L** | walking leg |

| | |
|---|---|
| **Lb** | labrum |
| **ml** | millilitre |
| **NBT/BCIP** | nitro blue tetrazolium/5-bromo-4-chloro-3-indolylphosphate |
| **NHEJ** | non-homologous end joining |
| **nt** | nucleotide |
| **N-terminal** | amino-terminal |
| **O** | opisthosomal |
| **Oc** | yolk columns |
| **PcL** | precheliceral lobe |
| **PCR** | polymerase chain reaction |
| **PNS** | Peripheral Nervous System |
| **Pp** | pedipalp |
| **pRNAi** | parental RNAi |
| **PSM** | Pre Somitic Mesoderm |
| **RNA** | ribonucleic acid |
| **RNAi** | RNA interference |
| **rpm** | rotations per minute |
| **RT** | reverse transcription |
| **SAZ** | segment addition zone |
| **sgRNA** | short guide RNA |
| **snRNA** | small nuclear RNA |
| **SOX** | Sry-related HMG box |
| **st** | stage of development |
| **TF** | Transcription Factor |
| **WGD** | Whole Genome Duplication |
| **wt** | wild-type |

## List of Figures

## List of Tables

# General Introduction

## Introduction

The field of evolutionary developmental biology is focused on understanding how changes in the ontogenetic mechanisms of different organisms give rise to phenotypic divergence on one hand and reconstructing ancestral features on the other (Sommer, 2009). This relies on comparative studies between several organisms to characterize similarities and differences. For example, studies of segmentation help to explain how chordates, annelids, and arthropods sub compartmentalize their bodies into spatiotemporally controlled segments. The process has similarities in a broader view, but intrinsic mechanisms are different in many aspects indicating that segmentation has likely evolved several times (Patel, 1994; Scholtz, 2002; Peel, 2005). Also, studies on neurogenesis have provided insights into how animals with a functional nervous system pattern their cells in order to form the neural precursors that will give rise to the neurons (Harteinstein, 2015).

This doctoral research analyzed the expression of the toolkit genes of the Hes (Hairy/Enhancer-of-Split) and the Sox (SRY-box) gene families specifically during segmentation and neurogenesis in the common house spider *P. tepidariorum* (Arthropoda, Chelicerata, Arachnida, Araneae), providing new insights into how these two processes are regulated during chelicerate development, compared to other animals. In addition, my work continued the effort to further develop state-of-the-art tools for the study of developmental biology in this spider.

### 1 Spider studies in embryology and developmental biology

Arachnid embryology studies date back to 1862, when the Swiss anatomist ReneEdouarde Claparede started analysing embryos of spiders and mites (Claparede, 1862). This initial research was then followed by a period of around 30 years of scarce studies until histological techniques started being used in different species of chelicerate embryos, including horseshoe crabs (Kingsley, 1892; Patten & Hazen, 1900).

The first embryological description for *P. tepidariorum* (formerly *Theridium* and *Acharaneae*) was carried out by Montgomery (1909), who mostly focused on their embryology, rather than on comparative studies. In his work, the meroblastic intralecithal type of cleavage, in which the egg is mostly composed by yolk, was first described (Montgomery, 1909). After fertilization occurs at the periphery of the yolk, the first nuclei migrate towards the center of the egg, hence the name intralecithal (inside of the yolk) where the initial cleavages take place (Montgomery, 1909; Holm, 1952; Mittmann & Wolff, 2012). The divisions occur in a syncytial environment until the 16-cell stage, when the cells migrate towards the periphery of the yolk-mass with every consecutive division (Figure 1.1) (Suzuki & Kondo, 1995; Kanayama, 2010; Pechmann, 2016).

**COPYRIGHTED IMAGE REMOVED FROM ELECTRONIC VERSION**

**Figure 1. 1. Early cleavages in *P. japonica.*** Paraffin sections of early cleavages in the spider P. japonica, which belongs to the same genus as P. tepidariorum. A) 2-nucleus stage, with the yolk columns (oc) externally and the inner mass (im) in the centre. B) 8-nucleus stage. C) 16nucleus stage, when the cells starts the migration towards the periphery, and the blastocoel appears. D) 32-nucleus stage (Suzuki & Kondo, 1995).

In *P. tepidariorum*, the germ disc is formed by the blastoderm cells between 16 to

27 hours after egg laying (Figure 1.2). This is the primordial embryonic tissue in the spider. In the center of the disc an aggregation of cells that will form the cumulus (the morphological homolog of the dorsal organizer), which is the region where the gastrulation will occur (Figure 1.2) (Anderson, 1973; Mittmann & Wolff, 2012). What is striking about the gastrulation in the spiders is that the cumulus has a mesenchymal characteristic, which means that after the invagination of the ectodermal cells, the cells that went through a epithelial-to-mesenchymal process (EMT) migrate towards the rim of the germ disc (Akiyama-Oda & Oda, 2003; Mittmann & Wolff, 2012). Ake Holm, in his classical studies of spiders (Holm, 1952; Holm, 1954), discussed the potential that the dorsal organizer of several different species had to induce an entire axis, following what was known about the organizers in frogs, started experiments with transplantation of this region called the cumulus (Spemann & Mangold, 2001). His explants of the cumulus to different regions of the germ disc gave rise to spiders with a duplicated axis (e.g. two heads in one axis), and these were amazingly described in his studies. Sadly, at the moment there was no further understanding of the molecular mechanisms behind this interesting experiment (Holm, 1952).

50 years later Akiyama-Oda & Oda (2003) discovered that the mechanism controlling cumulus migration is BMP signalling. They found that the knockdown of *decapentaplegic* (the arthropod BMP-4 orthologue) causes several deffects in the cumulus migration, mesoderm formation and axis specification (Akiyama-Oda & Oda, 2003). Interestingly, BMP signalling is also involved with the dorsoventral patterning in the organizer of frogs, which was the experiment on which Holm based his transplantations in the first place (Holm, 1952; Spemann & Mangold, 2001).

After the transition of radial to axial symmetry mediated by the cumulus prosomal segments are formed and opisthosomal segments are added sequentially to the germband of spider embryos (Figure 1.2). One of the focus of this thesis is in the mechanisms regulating the addition of segments. In spiders this was initially studied in the late '90s with studies of the spider *Cupiennius salei* and was followed by *P. tepidariorum* (Damen, 1999; Abzhanov, 2001; Damen, 2002; Akiyama-Oda & Oda, 2003; McGregor, 2008b). *P. tepidariorum* have a well-described embryonic development in stages, which facilitates the comparative studies with other species (Figure 1.2) (Mittmann & Wolff, 2012). This species also produces a high number of embryos every 2 or 3 days, ranging from 250 embryos per cocoon, and they are easily accessed from the females (Mittmann & Wolff, 2012), and is easy to culture. In addition, techniques such as in situ hybridization for the detection of mRNA transcripts in the embryos (Abzhanov, 1999; Akiyama-Oda & Oda, 2003, McGregor *et al*., 2008), immunohistochemistry for the analysis of proteins (Schwager, 2015; Paese *et al*., 2018), gene knockdown with RNA interference, injected both in to females for a systemic analysis (Akiyama-Oda & Oda, 2003), or in the embryos for clonal analysis (Kanayama *et al*., 2010) exemplify the powerful tools available for studying the genetic regulation of the development of this organism and build on classic studies of arachnid development.

COPYRIGHTED IMAGE REMOVED FROM ELECTRONIC VERSION

**Figure 1. 2. *P. tepidariorum* development.** (A) Cellularisation is complete at stage 2 and the germ disc including the primary thickening (pt) in the center forms at stage 4. The germ band with the segment addition zone (SAZ) has developed at stage 8 and elongated with the limbs becoming morphologically visible. Inversion occurs between stage 10 – 13, marked by the internalisation of yolk. At stage 14 the embryo is fully developed with a clear constriction (arrowhead) between prosoma (Pro) and opisthosoma (Op). (Adapted from Hilbrant *et al*., 2012).

Indeed, the sequencing of the *P. tepidariorum* recently revealed that there has been a whole genome duplication (WGD) in the ancestor of spiders and scorpions (Schwager *et al*., 2017). Interestingly, duplicated copies of many of important developmental genes, including Hox genes as well as other regulatory factors such as microRNAs, have been retained in spiders and other arachnids since this event (Leite *et al.,* 2016; Schwager *et al.,* 2017). This obviously has interesting implication for the evolution of the regulation of developmental processes in these animals including segmentation and neurogenesis, which I studied for this thesis. Below I introduce the processes of segmentation and neurogenesis in different clades and compare with what is already known for *P. tepidariorum*, in order to provide a background for the results that will be discussed in the following chapters.

## 2 Segmentation

One characteristic of some metazoan groups is their ability to build a body from sequentially repeated (morphologically and/or functionally) units, commonly called segments (Hannibal, 2003). However, the way in which that these segments form and are patterned can differ amongst species.

Primary segmentation is a process that occurs in the mesodermal pouches (hereby coelomic cavities), which are divided by epithelial junctions, as found in molluscs and echinoderms (Tautz, 2004). These animals are classified as “pseudosegmented” by Ballavoine (2014), who considers that there is a certain degree of segmentation in many extant phyla, in addition to arthropod, annelids and vertebrates. This view is not universally accepted in the field, however, where the consensus appears to be that the biological definition of segmentation doesn't not applies to these species (Hannibal, 2013).

Secondary segmentation is a process in which clusters of cells along the axis are in an undifferentiated state (Peel, 2003). These cells are then patterned into segments after the activation of several molecular signals, usually in an anterior to posterior direction (Tautz; 2004; Peel, 2005; McGregor, 2008). Three extant, phylogenetically distant, clades exhibit this body plan feature: vertebrates, annelids, and arthropods. Vertebrate somitogenesis is a process that involves the division of the presomitic mesoderm into somites, which will give rise to the vertebrae and patterned neurons (Pourquie, 2003). Similarly, annelids divide their entire body into morphofunctionally similar metameres, that are ectodermal, mesodermal and endodermal (Balavoine, 2014), whereas arthropods subdivide their ectoderm in repeated segments (Hannibal, 2003).

While there is no consensus on the evolutionary origin of how these processes arose in these three different clades (Davis & Patel, 1999; Tautz, 2004), there are three different scenarios that are debated in the field. The first view suggests that segmentation evolved independently in annelids, arthropods and vertebrates, which would be a clear example of an evolutive convergence (Figure 1.3 – green balls). The second view explains that segmentation arose independently in chordates and the protostome ancestor. This would require many losses of segmentation in the lophotrochozoans and ecdysozoans, which is less parsimonious than the other views (Figure 1.3 – orange). Finally, it was also proposed that segmentation arose in the common ancestor of all bilaterials, the termed

Urbilateria, and that this process was retained in arthropods, annelids and vertebrates, but lost in all the rest of the phyla (Figure 1.3 – Red).

The evolution of the process of segmentation is therefore an important question in the evo-devo field. Intriguingly it has been found that some molecular mechanisms are common to segmentation in these three phyla and while this may indicate ancestral regulatory mechanisms for this process, this pattern could also be explained by their independent co-option (Chipman, 2008). However, it is important to characterize the regulation of segmentation to be able to understand this process and compare different animals. Thus, this part of the introduction will first focus on the differences between segmentation among vertebrates, annelids, and arthropods, and then expand in more detail on what is currently known regarding the control of segmentation in spiders.

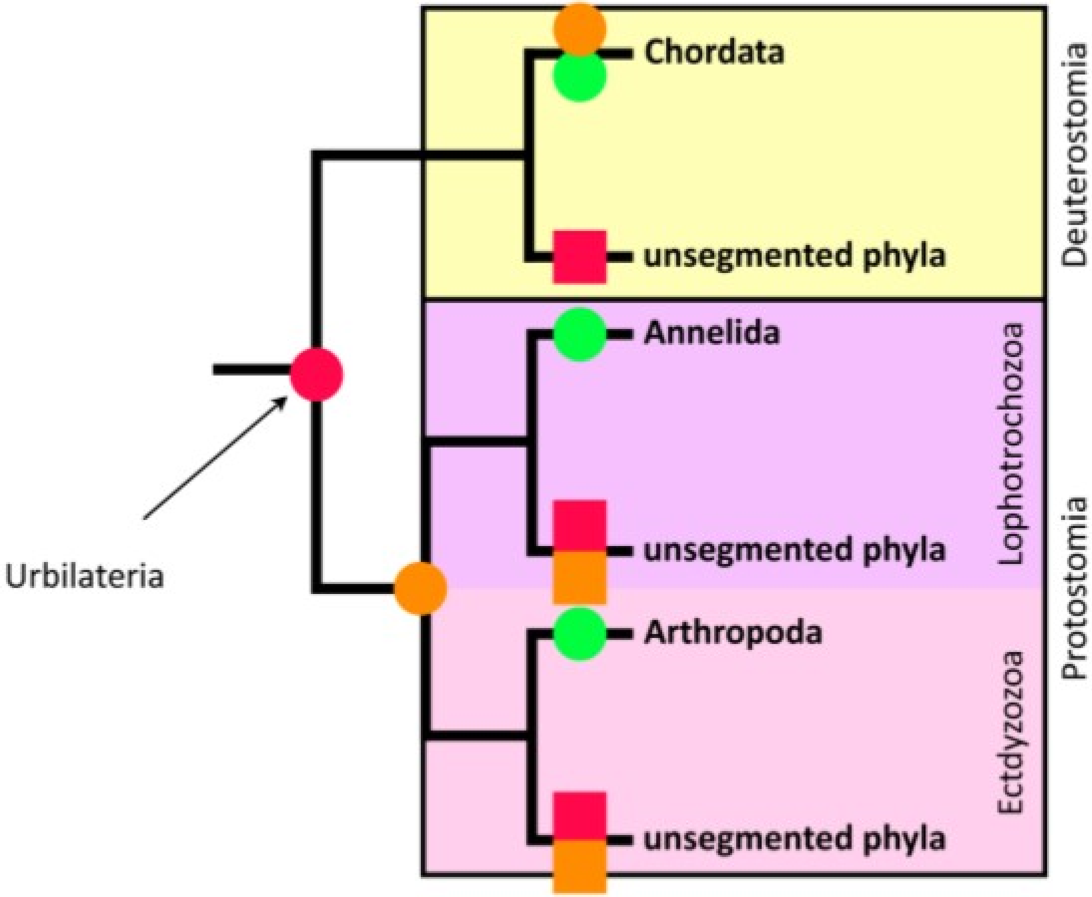


**Figure 1. 3. Possible origins of segmentation in the metazoans.** Summary of the three views on the origins of segmentation in bilaterians (after Davis & Patel, 1999). The colored circles represent the gain of segmentation, whereas the colored squares the loss of the mechanism. The differences in the colors are explained in the text above.

## 2.1 Segmentation in Vertebrates

The formation of segments in vertebrates occurs in a zone of high cell proliferation in the presomitic paraxial mesoderm (PSM). The “blocks” of segments that are patterned in the PSM are called somites, hence the name somitogenesis (Pourquie, 2000). The somites are then the visible expression of segmentation in vertebrates, producing derivative structures such as the vertebrae, the intervertebral disks, and the ribs (Tam and Trainor, 1994).

At the cellular level, the cells in the primitive streak have an ectodermal potential, and they acquire a mesenchymal fate after gastrulation, which allows their migration to differentiate into mesoderm (Tam & Trainor, 1994). These mesodermal progenitors in the PSM are called somitomeres (Pourquie, 2000). This is the first important difference between vertebrate, annelids and arthropods segmentation: in vertebrates, the segments are from mesodermal original, whereas in annelids they are ectodermal, mesodermal and endodermal, and only of ectodermal origin in arthropods (Pourquie, 2003; Peel, 2005).

To explain somitogenesis at the molecular level, Cooke & Zeeman (1976) developed a model called the “clock and wavefront”. In summary, this mechanism is controlled by opposing gradients: FGF and Wnt emanating from the posterior and retinoic acid pathway from the anterior region of the embryo (Cinquin, 2007). Also, the undifferentiated cells in the posterior part of the PSM are highly dependent on *hes7* in mammals and *her* in zebrafish, and these genes are downstream targets of the Notch signaling (Kageyama, 2007). These genes have a cyclic expression in the PSM (clock), and act together with the FGF and Wnt signalling pathways in the PSM to give rise to the aforementioned somites in an anterior-to-posterior manner (Figure 1.4) (Pourquie, 2007). n the anterior part of the PSM at the determination front, which is the point where Wnt, FGF and retinoic acid gradients meet to subdivide the PSM into the new somites (Figure 1. 4) (Pourquie, 2003; Gibb, 2010).

COPYRIGHTED IMAGE REMOVED FROM ELECTRONIC VERSION

**Figure 1. 4. Genetic control of somitogenesis in chick.** A) Dorsal view schematic of a chick embryo with the position of somites and the PSM that flank the axial neural tube. B) Periodic waves of expression of the *cHairy1* gene (successive waves shown in different colours) across the PSM share the same periodicity as somite formation. C) A schematic diagram integrating the domains of various signalling activities in the PSM – the wavefront of determination on the lefthand side, and the clock on the right. (Adapted from Gibb, 2010).

## 2.2 Segmentation in Annelids

The phylum Annelida is comprised of a diverse group of animals with vastly different morphological and ontogenetic features, most of which undergo indirect development with several larval stages (Balavoine, 2014). Much is now known about the embryonic development of these animals (Shimizu, 2001), and one of the best studied annelids in regard to developmental biology is the nereid worm *Platynereis dumerilii*, due to the transparency of the embryos enabling easy observation of embryogenesis (Balavoine, 2014).

In this species, the most anterior part of the body is non-segmented (protostomium), and the posterior segments are added through a budding process that emanates from a segment addition zone (SAZ) (Wanninger *et al.,* 2009). In this region, stem-cells, called teloblasts, divide into successively smaller cells with ectodermal (ectoteloblasts) or mesodermal (mesoteloblasts) potential. Strikingly, during annelid segmentation the three cell layers develop at a similar rate, in contrast to what is seen in vertebrates and arthropods, in which where the endoderm is patterned only later in development (Weeden, 1997).

At a molecular level, it is known that annelid development relies much on the Wnt/β-Catenin and Hedgehog signaling pathways, but there is no evidence for the involvement of Notch like in vertebrates. (Shimizu, 2001; Dray *et al*., 2010). However, orthologues of *hes/hey* (annelid orthologues of the *Drosophila* gene *hairy*) are expressed in the SAZ in an oscillatory pattern, similar to what is observed in vertebrates (Kageyama, 2012; de Rosa *et al.,* 2015), suggesting that they play an important role in the SAZ.

## 2.3 Segmentation in Arthropods

Arthropods present a wide variety of segmentation modes (Hannibal, 2003) but the first difference in the mechanism of segment patterning relates to how the invaginating cells that undergo gastrulation form a germ band (Figure 1.5). In some insects, such as *Drosophila melanogaster* and *Musca domestica*, the blastomeres that undergo gastrulation forms an extended germ band that takes up the entire intervitelline space – and the germ band will later retract and in the space of minutes the intersegmental borders become visible (Sommer & Tautz, 1991). Due to this special characteristic where the entire embryo forms the germ band, it was given the name “long germ band” mode of embryogenesis (Figure 1.5) (Liu, 2005).

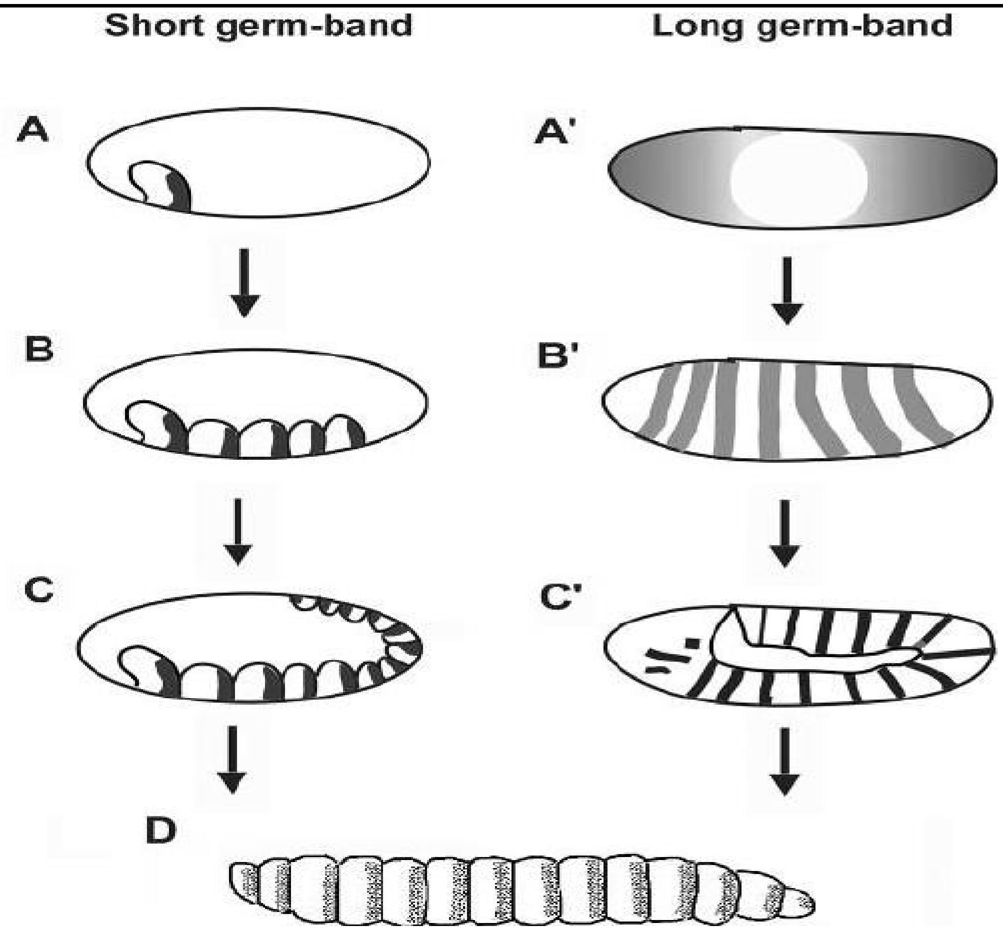


**Figure 1. 5. Diagram of germ bands types in arthropods.** Schematic representation which illustrates the difference between the modes of germ band development. A-C) The short germ band embryogenesis presents the cephalic region in formation, with a segment addition zone in the posterior region, being later differentiated. A’ – C’) The long germ band mode of embryogenesis shows that all the segments are patterned simultaneously soon after gastrulation (Based on Newman, 2005).

An important question that was asked in order to understand the control of segmentation in arthropods was which molecular mechanisms are involved (Peel, 2005).

In order to understand this, genetic screening of the fruit-fly *Drosophila melanogaster* was carried out in the late ‘70s and this and subsequent work identified the genes and pathways involved in the classic segmentation cascade of this insect (Akam, 1985; McDonald, 1986; Akam, 1989).

In summary, segmentation in *D. melanogaster* is regulated by a hierarchical cascade of genetic activation and repression, which can be subdivided into four classes of genes (Figure 1.6):

- Maternal effect genes: The mRNA or proteins from these genes are deposited into the egg by the female via the nurse cells of the ovarioles before fertilization. Upon fertilization, the gene products provide the coordinates to pattern the anterior-posterior axis. *hunchback* and *bicoid* are good examples of maternal effectors that acts in the anterior and *caudal* in the posterior (Figure 1.6) (Frohnhofer, 1986; Mlodzik, 1987; Hulskamp & Tautz, 1991);

- Gap determinants: After fertilization maternally supplied proteins activate the expression of gap genes, which were named for the phenotypes that resembles a gap in the segmental pattern of the embryo (e.g. lack of trunk, but formation of head and abdoman). Gap genes are expressed in patterns marking future groups of contiguous segments and subdivide the embryo into the presumptive head, thorax, abdomen and posterior region of the embryo. Examples of this class are the genes *Kruppel* and *knirps*, which contribute to determining the thorax and abdomen (Figure 1.6) (Knipple, 1985; Rothe, 1989);

- Pair-Rule genes: Gap genes and maternal effectors that are involved in the trunk region of the embryo activate the expression of the pair-rule genes. These genes are expressed in a double segment periodicity in alternate parasegments, and the phenotypes caused by their knockdown show a deletion of every second segment. These genes, which include *even-skipped*, *hairy* and *runt*, have an important role in segment specification (Figure 1.6) (Ingham, 1985; Frasch, 1987; Ingham, 1988).

- Segment Polarity genes: The polarity of each parasegment is defined by the expression of the segment polarity genes. These genes are are regulated by gap and pair rule genes, and are expressed in either the anterior or posterior compartments of the parasegments. For example, engrailed set the anterior boundary of the parasegment and marks the posterior comparment of the segment, whereas the expression of *wingless* marks the posterior of the parasegment, which corresponds to approximately the middle of the segment (Figure 1.6). In *D. melanogaster*, *hedgehog* also acts as a posterior segment polarity gene in the segmentation mechanism (Kornberg, 1985; Levine, 2008; Peel, 2005).

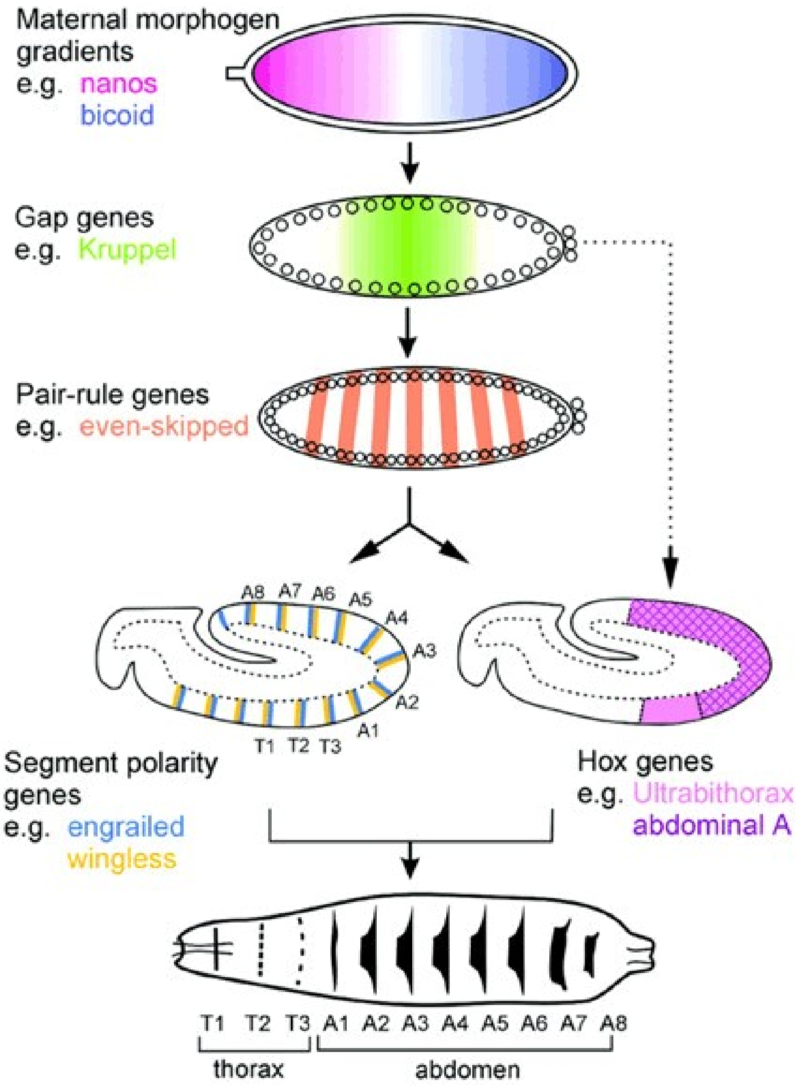


**Figure 1. 6. Molecular cascade controlling segmentation in D. melanogaster.** Schematic representation of the molecular cascade that is initiated by maternal proteins, which activate the downstream expression of gap, pair-rule and segment polarity genes in dipteran species (Based on Sanson, 2001).

However, in contrast to *D. melanogaster*, the majority of arthropod species, including most other insects, myriapods, crustaceans and chelicerates exhibit the “short germ band” mode (Figure 1.5). It received this name because after the invagination of the blastomeres, the ectodermal layer will pattern just a small part of the presumptive head ectoderm and form a few of the anterior segments (Liu, 2005; Shaw *et al*., 2007).

Subsequently, posterior segments are continuously added from a region of undifferentiated cells with both ectodermal and mesodermal potential, called the segment addition zone (SAZ) or growth zone (Peel, 2005; McGregor *et al*, 2008).

The cells in the SAZ are thought to be maintained in an undifferentiated state until molecular signals are expressed and activates the rearrangement of these cells into segments (Scholtz, 2002; Peel, 2005; McGregor *et al*., 2008). However, many open questions remain concerning how short germ arthropods pattern their segments including how this is regulated and at the genetic and cellular levels. To address these questions, many studies have compared the *Drosophila* segmentation cascade to the regulation of segmentation in short germ arthropods.

Interestingly, it was discovered that many of the same genes regulate embryogenesis in long germ band and short germ band arthropods and aspects of the *Drosophila* segmentation cascade are also involved in segmentation in short germ arthropods (Peel, 2005; Clark & Akam, 2017). However, short germ arthropods also use mechanisms that have been probably been lost in long germ insects like *Drosophila* (Patel, 1997; Levine, 2008; Clark & Peel, 2017).

In short germ arthropods, orthologues of the *D. melanogaster* gap genes such as *hunchback*, *kruppel* and *giant* were studied and it has been found that their expression is similar in *Gryllus*, *Oncopeltus* and *Schistocerca* (Patel *et al*., 2001; Liu & Kaufman, 2004a, 2004b; Mito *et al*., 2005). RNAi experiments confirmed that the knockdown of these genes results in loss of segments and/or homeotic transformations (e.g Mito *et al*., 2005).

The pair-rule gene orthologues are also involved in the segmentation in short germ arthropods, but their expression and function can vary from what is seen in the fruit-fly. For example in *T. castaneum*, the pair rule genes are also expressed in a double-segment periodicity (Patel, 1994; Choe *et al*., 2006) and knockdown of these genes results in phenotypes that resembles the fruitfly ones (Choe *et al*., 2007). However, in *Oncopeltus*, the orthologue of *even-skipped* is expressed with single-segment periodicity (Liu & Kaufman, 2005), which is also the case in spiders (Schönauer et al., 2016).

Segment polarity genes are the most conserved in regards of their expression patterns and functions in short germ arthropods in comparison with *D. melanogaster*. Functional studies have only been carried out in insects, but the expression patterns of *engrailed* and *wingless* orthologues in other arthropods including X and Y for, example, demonstrates their conserved expression throughout the arthropod clade (Hughes, 2002; Simmonnet, 2004).

It was found that orthologues of several genes that act during the canonical *Drosophila* segmentation cascade were also shown to play a role segmentation in chelicerates (Damen, 2000; Abhzanov, 2001; Damen, 2002; McGregor *et al.,* 2009). Similarities were found in the expression of Hox genes as well as pair-rule orthologues like *hairy*, *even-skipped* and *runt,* being involved in the patterning of the segments in the spiders *P. tepidariorum* and *C. salei* (Damen, 2000; Abzhanov; 2001). Also, in chelicerates like the spiders *P. tepidariorum* and *C. salei*, the expression of segment polarity genes is also conserved with respect to other arthropods and as mentioned above while the pair rule orthologues are involved in segmentation, they appear to be expressed with single segment periodicity (Damen, 2007; Kanayama, 2011). Moreover, there is evidence that *hb* and *distal-less* have gap gene like functions in *P. tepidariorum* (Pechmann *et al*., 2009; Schwager *et al*., 2009).

However, it was discovered in spiders and later in a few insects that like vertebrates but in contrast to Drosophila that the Delta-Notch signalling pathway is involved in segmentation (Palmeirim; 1997; Pourquie, 2003; Pueyo, 2008; Mito *et* al., 2011). (+ all the other relevant citations).

### 2.3.1 The GRN for formation of the SAZ and segment addition in *P.*

Functional analysis of *Delta* and *Notch* in *P. tepidariorum* demonstrated that their expression in the germ disc is essential for the formation of the SAZ from the caudal lobe, (Oda *et al*., 2007). It addition, was shown the dynamic expression of Delta and Notch in the SAZ of *C. salei* is required for segment formation (Stollewerk, 2003; Schoppmeier, 2005).

Furthermore, it appears that this dymanic expression of *Delta* in *P. tepidariorum* is regulated by the Wnt signalling pathway ligand, (McGregor *et al*., 2008). When this gene was knocked down using RNA interference (parental RNAi), the embryos were unable to differentiate the most posterior segments and displayed a truncated opisthosoma (McGregor *et al*., 2008). This finding was further investigated by Schönauer *et al* (2016), whi found that interplay between Delta-Notch and Wnt signaling is required to specify the SAZ and segment addition. This work also showed that other segmentation genes in the spider, such as *caudal* and *even-skipped*, act downstream of *Wnt8* and *Delta* (Schönauer *et al*., 2016; McGregor). The characterization of this gene regulatory network (GRN) segment addition in spider's segmentation is a great step towards an understanding of how segmentation is regulated in chelicerates compared to other arthropods but is still incomplete and many details remain to be added (Figure 1.7).

COPYRIGHTED IMAGE REMOVED FROM ELECTRONIC VERSION

**Figure 1. 7. Gene Regulatory Network of the posterior segment addition in *P. tepidariorum*.** Dl-N activates Wnt8 expression to maintain a pool of undifferentiated cells in the posterior SAZ. Wnt8 is then required for dynamic expression of Dl, which results in the formation of a stripe of Dl expression in anterior SAZ cells. Wnt8 and Dl-N are also required to activate cad expression, and together, these factors activate eve and runt expression. In anterior SAZ cells, Dl then suppresses Wnt8 expression and in combination with cad, eve and runt expression leads to segment formation. Arrowheads and flat arrows indicate activation and repression, respectively, although it is not known if these interactions are direct or whether additional factors are required. (Adapted from Schönauer et al., 2016).

### 2.3.2 The *hairy/Hes* genes in spiders

The gene *hairy* is a member of a conserved gene family that is involved in embryogenesis in species throughout metazoan phylogeny: The Hairy and Enhancer of Split (HES) complex (Jennings, 1999; Gazave, 2014). These genes encode transcription factor proteins that have a basic helix-hoop-helix DNA binding domain (bHLH) and an Orange domain, which is binding a domain for the Notch signalling pathway (Baker, 2018). In addition, Hes genes contain the YRPW domain (or WRPW) that is present in all the bHLH TF's, which binds Groucho, a downstream target of the Notch signalling pathway (Jennings, 2008).

As mentioned previously, *cHairy1* is a gene that is expressed in an oscillatory pattern during somitogenesis (in mice) (Kageyama, 2007). In annelids the expression of the *hey/hes* genes regulates the formation of segments from the SAZ, and *hairy* acts as a pair rule gene in *D. melanogaster.* The spider orthologue of *hairy* is also expressed in the SAZ of both *C. salei* and *P. tepidariorum* (as well as during the patterning of prosoma) (Damen, 2000; Pechmann, 2009). In *P. tepidariorum*, like Dl, the cyclic expression of *hairy* is lost in *Wnt8* pRNAi embryos (McGregor *et al.*, 2008; Schönauer *et al.*, 2016). However, despite these studies, very little is known about the role of *hairy* and Hes genes more generally during spider embryogenesis.

The study by Dearden (2015) identified several orthologues of the Hes-like genes in the *P. tepidariorum* genome. These included the orthologues of the *Drosophila* genes Deadpan and Hey (*Hairy and E(spl)* related with a *Y*RPW domain), which are involved in the differentiation of neural precursors (neuroblasts) in insects, Clockwork Orange (cwo) which is involved in the control of circadian rhythm, and Side (Similar to Deadpan) which control the migration of motor neurons (Ingham, 1985; Siebert, 2009; Dearden, 2015). Unfortunately, Dearden (2015) just did bioinformatic analysis to compare the Hes gene family in metazoans, thus no gene expression was further analysed for the spiders (Figure 1.8). For this reason, one of the aims of my thesis is to characterize properly the orthologues for the Hes-like genes found in the *P. tepidariorum* genome.

COPYRIGHTED IMAGE REMOVED FROM ELECTRONIC VERSION

**Figure 1. 8.** ***Schematic representation of the phylogenetic relationships of bHLH Orange.*** *Bayesian phylogram of representative bHLH-orange domain proteins reconstructed using the WAG model of protein evolution. bHLH Orange domain proteins from Chelicerates (PTEP,* P. tepidariorum*; LHE,* Latrodectus hesperus*; ISCW,* Ixodes scapularis*; CSCU,* Centruroides exilicauda*; XP,* Metaseiulus occidentalis*) Insects (honeybee, GB;* Drosophila*; gene names) a myriapod,* Strigamia maritima *(smar_), and an Annelid,* Platynereis *(AGS identifiers) with well characterised HES genes (Source Dearden, 2015).*

Interestingly, in other arthropods the Hes genes act in a similar manner to their counterparts in vertebrates. It has been shown by Chipman (2008) that *hes-4* and *hes-7* likely have roles in segmentation in the centipede *Strigamia maritima*, where they are expressed in a posterior-to-anterior manner in the SAZ and subsequently in nascent segments. In the cockroach *P. americana*, the expression of *hairy* orthologs are clearly involved in a similar mechanism (Pueyo, 2008) because an oscillatory pattern is apparent visible in the posterior of the germ band (Figure 1.9). A similar pattern was also found in another short germ band insect, the flour beetle *Tribolium castaneum* (see Aranda, 2008) (Figure 1.9). However, no functional study of *hairy* was carried out in either of these species to fully understand the role of this gene in segment addition (Figure 1.9).

COPYRIGHTED IMAGE REMOVED FROM ELECTRONIC VERSION

**Figure 1. 9. *hairy* expression in the cockroach and in the flour beetle.** Expression of the *hairy* orthologues in P. americana and T. castaneum. It is noteworthy that the expression in the growth zone cycles in a posterior-to-anterior manner. Adapted from Aranda (2008) and Pueyo (2008).

Therefore, one of the aims of my thesis was to better understand how *hairy* and other bHLH Hes-like genes are expressed in spider embryos to compare to their possible roles in segmentation in other arthropods. In addition, this provided an opportunity to understand if the genes involved in *D. melanogaster* neurogenesis (such as *deadpan*, *hey* and *side*) play a similar role during spider embryogenesis.

## 3 Neurogenesis

Metazoans have the ability to sense signals from the external environment and transduce these stimuli into a cellular and molecular signal, which allows the organisms to respond in many different ways. This of course relies on the specialization of the cells in a complex nervous system. The comparative studies in different taxa allow us to understand how the ontogenetic similarities arose, especially in how these cells differentiate into ganglia and neurons, as well as which neurogenic genes are conserved and act in a similar way in different species. In the following I will focus on the two well studied groups – vertebrates and arthropods – and specifically highlight what is currently known in the common house spider.

### 3.1 Neurogenesis in Vertebrates

Neuronal differentiation in vertebrates occurs in three distinct processes (Lewis, 1996). The first occurs in the central dorsal neuroepithelium that contains a high number of postmitotic cells and undifferentiated proliferative neural cells, which will subdivide and differentiate into the neurons of the neural plate (Haubensak *et al*., 2004). The neural plate then folds and cells with neuroectodermal potential will form the neural tube. This process is dependent on Notch signaling (Lewis, 1996) and Sox proteins (Kamachi, 2013), especially Sox2, to help maintain the pool of proliferative cells.

The second mechanism entails the migration of neural crest cells, a subset of mesenchymal cells that give rise to the peripheral nervous system and annexes, both neuroectodermal and epithelial ectodermal (Simoes-Costa, 2015). This process is highly dependent on Wnt, FGF and Notch signaling, but also relies on Sox proteins (SimoesCosta, 2015).

Finally, the selection of cells from the epithelium to become neurogenic is made by the delamination of these cells from the epithelium and acquisition of migratory ability (epithelial to mesenchymal transition – EMT). This process is called cranial placode neurogenesis (Lassiter *et al*., 2014), and is also dependent on the Wnt, FGF and Notch signaling pathways. The first and the latter processes most resemble arthropod neurogenesis, for the specificity of cell-types (e.g. neural cells in the neural plate cells and neuroblasts in the ventral neuroectoderm) and migratory mechanism (e.g. mesenchymal cells with neurogenic potential) (Lewis, 1996).

### 3.2 Neurogenesis in Arthropods

Arthropods display a great variety of neurogenic patterning, which depends on the mode by which cells rearrange themselves to delineate the CNS. In onychophorans, insects, and crustaceans, a single cell with stem-cell potential delaminates from the neuroectoderm and can divide both symmetrically or in a stem cell manner (Figure 1.10) (Stollewerk, 2016). This mode of neurogenesis resembles the vertebrate placodal development, in which single cells acquire the potential to induce differentiation into neurogenic fate (Scotting, 1996). On the other hand, chelicerates and myriapods have clusters of postmitotic precursors in both the procephalic and ventral neuroectoderm (Stollewerk, 2016). This mode of neurogenesis can be compared to that of dorsal neuroectoderm differentiation in vertebrates, in which a group of cells acquire the ability for neurogenic potential, instead of specific cells being specified to pattern the neurons (e.g. neuroblasts in *D. melanogaster*) (Stollewerk, 2001; Stollewerk, 2016).

COPYRIGHTED IMAGE REMOVED FROM ELECTRONIC VERSION

**Figure 1. 10. Pattern of neural precursor/progenitor specification and division.** Onychophorans: single neural progenitors delaminate and divide symmetrically to produce intermediate neural precursors, which divide again. Chelicerates: most neural precursors are postmitotic. Insects: neuroblasts divide asymmetrically to produce GMCs, which divide again to produce neurons and glia. Crustaceans show the same division pattern. (Adapted from Stollewerk, 2016).

Like in *D. melanogaster,* differentiation of the neuronal progenitors in other arthropods is regulated by the *achaete-scute* homologues (ASH) and Notch signalling (Kageyama, 2007; Stollewerk, 2016). Briefly, proneural genes of the ASH group are expressed in clusters of cells in the ventral neuroectoderm. By a process called lateral inhibition, regulated by Notch, only one cell continues to express these genes, which differentiates into a neuroblast, whereas all of the other cells in the clusters will have an epidermal fate (Stollewerk, 2003).

### 3.3 Sox genes and its involvement in neurogenesis

Another important group of genes involved in both vertebrate and insect neurogenesis is the Sox group. This family of genes encompasses a set of conserved metazoan transcriptional regulators that play critical roles in a range of important developmental processes (Kamachi, 2013), including stem cell maintenance and nervous system development (Sarkar, 2013). The Sox family is defined by a set of genes containing an

HMG class DNA-binding domain sharing greater than 50% sequence identity with that of SRY, the Y-linked sex determining factor in eutherian mammals (Sinclair, 1990). In chordates, the family contains approximately 20 genes, which have been subdivided into eight groups (A-H) based in the DNA-binding domain sequences (Bowles, 2000; Heenan, 2016).

Representatives of the Sox family have been identified in all metazoans examined to date - mainly Groups B to F, with other groups specific to particular lineages. Note that Sox-like sequences have been reported in the genome of the choanoflagellate *Monosiga brevicollis* (King, 2008), these are more closely related to the non-sequence specific HMG1/2 class of DNA binding domain and thus true Sox genes are restricted to metazoans (Figure 1.11) (Zhong, 2011; Schnitzler, 2014).

COPYRIGHTED IMAGE REMOVED FROM ELECTRONIC VERSION

**Figure 1. 11. Sox factors throughout the metazoan phylogeny.** Sox-like HMG-box genes have been described in choanoflagellates. However, it is thought that true Sox genes arose in the common ancestor of the eumetazoans and SoxF genes were lost in poriferans and placozoans. SoxA group is a vertebrate novelty (Adapted from Schnitzler, 2014).

Although vertebrate Sox genes have been intensively studied due to their critical roles in development (Kamachi, 2013), with the exception of the fruit fly *Drosophila melanogaster*, they have been less well characterised in other invertebrates. *D. melanogaster* contains eight Sox genes (four group B genes and one each in groups C to F), which is generally consistent across the insect genomes examined to date (Prochanukul, 2011; Wilson, 2008; Wei, 2010). Of particular interest are the Group B genes of insects, which share a common genomic organisation that has been conserved across all insects examined to date, with three genes closely linked in a cluster (Prochanukul, 2011; McKimmie, 2005). Critical roles in early segmentation and nervous system development have been shown for *Dichaete* (*D*) (Russell, 1996; Nambu, 1996), and in CNS development for *SoxNeuro* (*SoxN*), where both these group B genes exhibit partial redundancy (Buescher, 2002; Overton, 2002).

The paucity of studies of Sox factors in other arthropods is one of the reasons why it was important to explore the expression and function of these genes further, especially in *P. tepidariorum* with the added interest that there has been a WGD in the lineage leading to this spider.

## 4 Aims of the thesis

The Hes-like and the Sox genes represent important toolkit genes in metazoans that are involved in many aspects of developmental biology, from stem-cell maintenance to cell death control. For this reason, the proposed aims for my thesis were:

- Description of the expression pattern of these genes in embryonic stages of the common house spider *P. tepidariorum*;
- Analysis of the phenotype caused by knockdown of one of the Sox factors and detailed study on its involvement in posterior segment addition in this short germ arthropod;
- Help to elucidate the phylogenetic relationships of these important genes in comparison to other metazoans;
- Investigation of the expression of selected orthologues in the developing nervous system of the spider;
- Development of new techniques for genomic modifications in *P. tepidariorum*.

# Chapter 1

# Expression of bHLH-TF Encoding Genes in *P. tepidariorum*

## Introduction

The bHLH family of transcription factors, which includes the Hes subfamily is the largest group of transcription factors with the ability to dimerize, and they regulate many different aspects of embryogenesis, such as segmentation, neurogenesis and apoptosis (Massari, 2000).

To gain a chelicerate perspective and therefore a broader view of the function and evolution of these genes, I characterized representatives of the Hes-like genes in *P. tepidariorum*. I started with the orthologue of *hairy*, *Pt-h*, which was previously described in spiders to be involved in posterior segmentation, as well as *deadpan*, *hey* and *side* - genes known for their involvement in neuronal differentiation in other animals.

Hes genes exhibit bHLH and Orange motifs in their coding region, as well as the motif WRPW four amino acids upstream of the stop-codon, which is recognized by the Groucho protein as a binding site for dimerization of these factors. In this chapter I will detail the identification, characterization and expression pattern of some of the bHLHcontaining genes present in *P. tepidariorum* genome.

## Identification and characterization of Hairy-Related orthologues

By conducting a tBLASTx search for conserved bHLH and Orange motifs in the genome, I identified several bHLH related-genes in *P. tepidariorum* (n> 80) (http://bioinf.unigreifswald.de/blast/P./blast.php). I focused on a small subset of bHLH-containing genes, that are related to *hairy* and *hes-like* genes because they have never been analysed in arthropods apart from *D. melanogaster*. I identified four genes that all contain the bHLH, Orange and Groucho binding domains (WRPW), that classify them as mostly related to the Hes genes in vertebrates (see the first four genes in Figure 2.1).

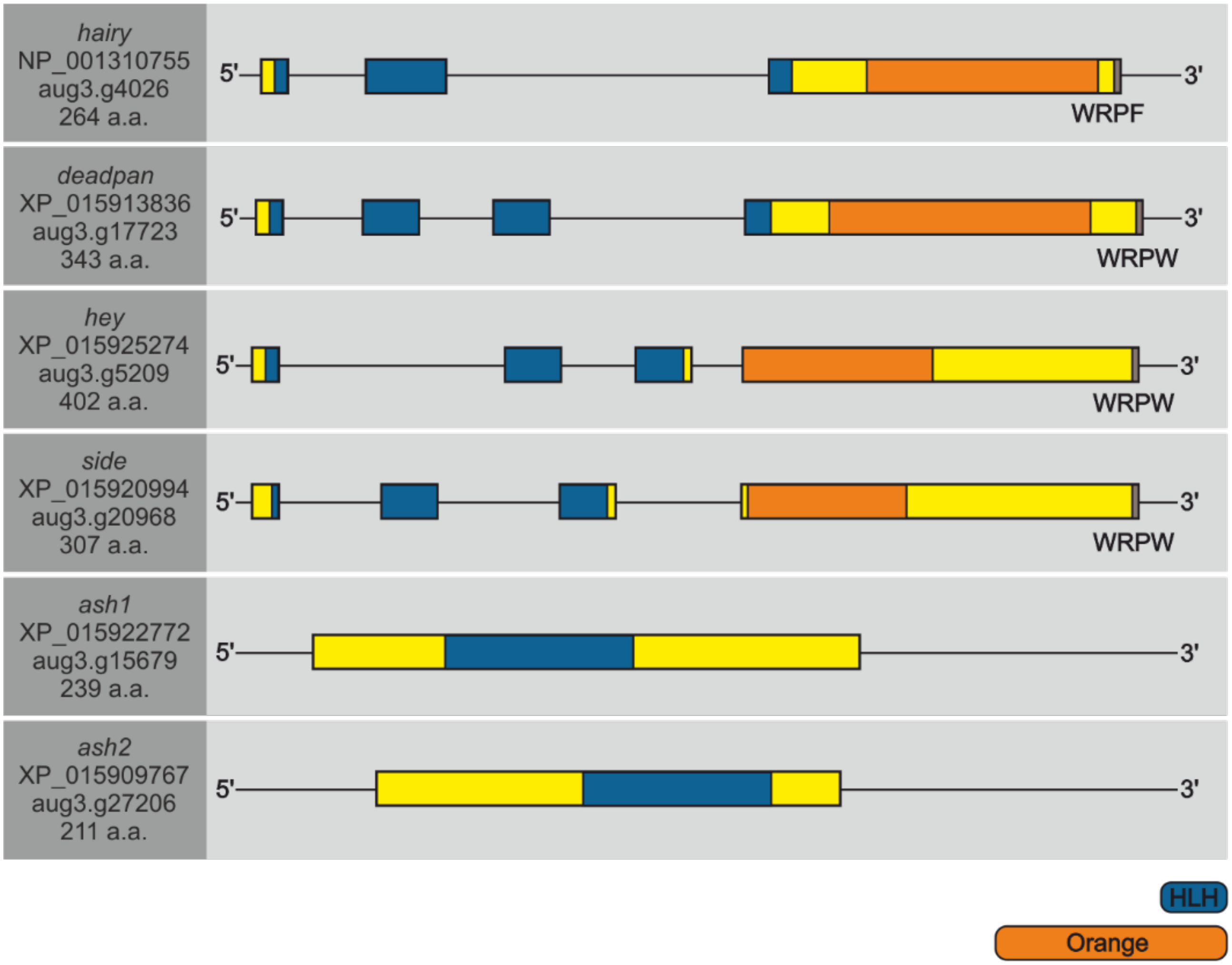


**Figure 2. 1. bHLH-containing genes found in *P. tepidariorum*.** Schematic representation of the six genes that the thesis focused on. On the left the gene name and GenBank ID, followed by the Augustus3 identifier and the amino acids length for these predicted proteins are given. The boxes represent the exons; blue colors for the bHLH domain, orange for the Orange domain, and light brown for the WRPW motif.

The first gene found was the previously identified orthologue of *Pt-h* (aug3.g4026) (McGregor *et* al., 2008; Pechmann *et* al., 2009). It contains three exons, with the predicted protein containing a bHLH domain at position 11 to 74, the Orange domain from 85 – 123, and the WRPW motif containing a change from a tryptophan to a phenylalanine (W to F) (Figure 2.1).

The second gene identified was the *deadpan* (aug3.g17723) orthologue, which exhibits four exons. The bHLH domain is in the same position as predicted for H (11 – 74), the Orange domain from position 89 – 127 and the WRPW at the end of the coding sequence (Figure 2.1). This gene was also previously annotated as a Deadpan group gene (Dearden, 2015).

*hey* (aug3.g5209) also has four exons, with the bHLH domain amino acids 20 – 82, the Orange domain from the 96 – 135, and again a WRPW motif immediately before the stop-codon (Figure 2.1). *side* (aug3.g20968) has four exons, the bHLH domain from amino acids 14 – 75, the Orange domain from 89 – 124 and again a WRPW motif is also present (Figure 2.1). As the name suggest this gene belongs to the Side group like *Pt-h*.

Note that in addition, I also identified two orthologues of the *achaete-scute* complex genes (namely *ash1* and *ash2*), known for being involved in neurogenesis in the spider *Cupiennius salei* (Stollewerk, 2001), which also encode a bHLH domain (final last genes in Figure 2.1). These genes don't contain either the Orange or the WRPW motif, only have the bHLH conserved domain (Figure 2.1). *Pt-ash1* (aug3.15679) and *Pt-ash2* (aug3.27206) just have one exon each.

To verify the annotation and identification of these genes predicted by BLAST, for these six *P. tepidariorum* (Pt) genes I made a maximum-likelihood tree using the software Mega, for 1000-bootstraps value and aminoacids comparison between the hes-like genes for the wandering spider *C. salei* (Cs), *Mus musculus* (Mm), *D. melanogaster* (Dm), the orb weaver spider *Nephila clavipes* (Nc), and the myriapod *S. maritima* (Sm) (Figure 2.2).

As predicted, there was strong support for placing *Pt-dpn* in the Deadpan group; *Pt-h* and *Pt-hey* in the Hairy group, *Pt-side* was grouped with the *D. melanogaster side* gene, and *Pt-ash1* and *Pt-ash2* had high bootstrap support for being placed in the Achaete-Scute complex with the *C. salei* orthologues.

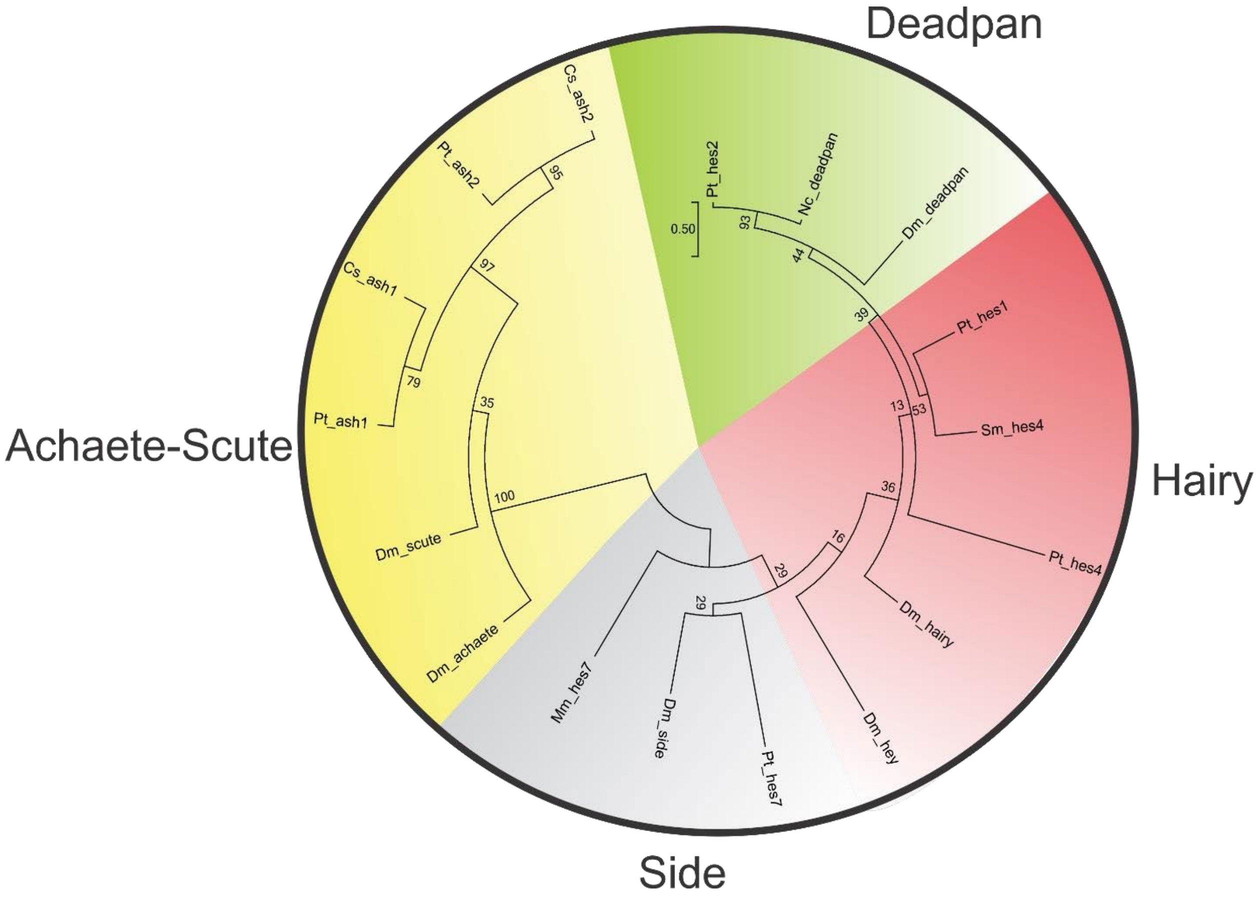


**Figure 2. 2. Phylogeny of bHLH domains in selected metazoans.** Phylogenetic tree showing the relationship between *C. salei* (Cs), *Mus musculus* (Mm), *D. melanogaster* (Dm), *Nephila clavipes* (Nc), *P. tepidariorum* (Pt) and *S. maritima* (Sm) based on bHLH domain sequences. The grouped genes are divided into different colours as highlighted outside the circle.

## Expression of *P. tepidariorum* bHLH-containing genes

To further characterise and compare the *hes-like,* and *ash1* and *ash2* genes in *P. tepidariorum*, I carried out in situ hybridizations to assay the expression of all six genes during embryogenesis in this spider.

*Pt-h* (aug3.g4026) expression matches the findings in Pechmann (2009). This gene is expressed strongly at stage 5, in both the presumptive anterior region of the germ disc, and broad expression is also observed in the centre of the germ disc (Figure 2.3 - A). When the germ band is formed, the expression is strong in the SAZ, but is also observed in the presumptive L2-L4 segments (Figure 2.3 - B-D). Dynamic expression is still seen in the posterior at stage 8.1 when strong expression also appears in the presumptive head lobes (Figure 2.3 - E-G). Strikingly, the expression in the anterior is not detectable at stage 8.2, when it is observed only in a segmental pattern in the opisthosomal segments (Figure 2.3 - H). Finally, at stage 9.1, proneural clusters in the entire ventral nerve cord show strong expression of *Pt-h*, and expression in the SAZ is still visible (Figure 2.3 - I).

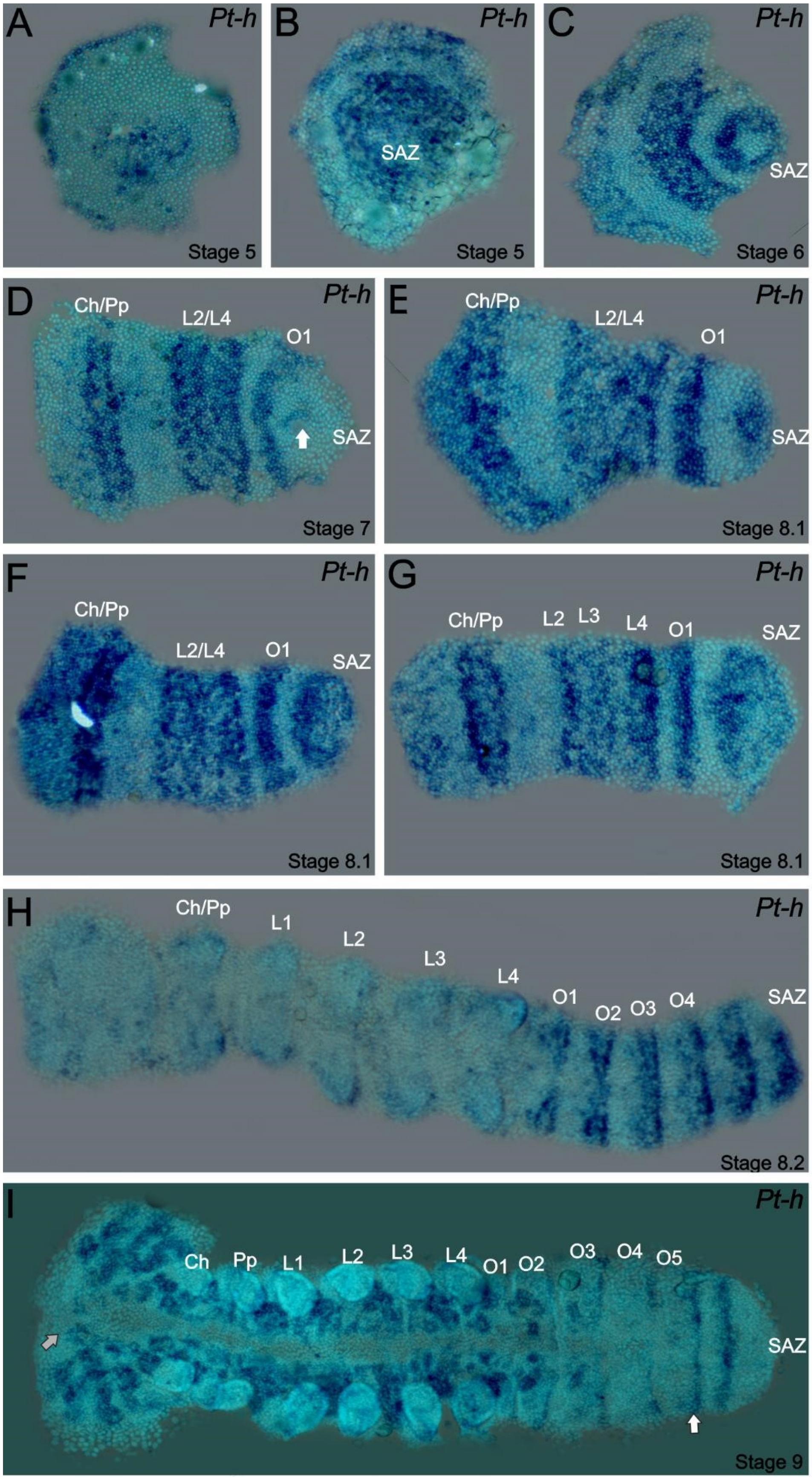

A
Pt-h
Stage 5
B
Pt-h
SAZ
Stage 5
C
Pt-h
SAZ
Stage 6
D
Ch/Pp
L2/L4
Pt-h
O1
SAZ
Stage 7
E
Ch/Pp
L2/L4
Pt-h
O1
SAZ
Stage 8.1
F
Pt-h
Ch/Pp
L2/L4
O1
SAZ
Stage 8.1
G
Pt-h
Ch/Pp
L2
L3
L4
O1
SAZ
Stage 8.1
H
Ch/Pp
L1
L2
L3
L4
O1
O2
O3
O4
SAZ
Pt-h
Stage 8.2
I
Pt-h
Ch
Pp
L1
L2
L3
L4
O1
O2
O3
O4
O5
SAZ
Stage 9

**Figure 2. 3. Expression of *Pt-h* in *P. tepidariorum*.** A - B) Stage 5 showing the expression at the rim of the germ disc and in the SAZ. C) hairy is expressed in the SAZ, forming segments and the most anterior region of the germ band at stage 6. D – E). At stage 7 and early stage 8, the expression is dynamic in the SAZ, and static in the Ch/Pp segment, and in the L2/L4 forming segments. F – G) The cyclic expression continues until mid-stage 8, from the SAZ and in the differentiated segments of the prosoma. H) At stage 8.2, the anterior expression is lost, and hairy is only expressed in the opisthosomal segments and the SAZ. I) The expression of this orthologue is now strong in the proneural clusters, and also faint expression remains in the SAZ. Ch: chelicerae, L1 – L4: prosomal segments 1 to 4, O1 – O4: opisthosomal segments 1 to 4, Pp: pedipalps; SAZ: segment addition zone.

*Pt-dpn* expression is first observed at mid-stage 9 in the head neurectoderm including the mushroom bodies, and clusters of neurogenic cells in the prosomal segments (Figure 2.4 – A). This gene is also expressed in stripes in the opisthosomal segments and SAZ (Figure 2.4 - A). This pattern is also visible at later stages when there is strong expression in the proneural clusters, although expression disappears from the SAZ at stage 10.1 (Figure 2.4 – B, C).

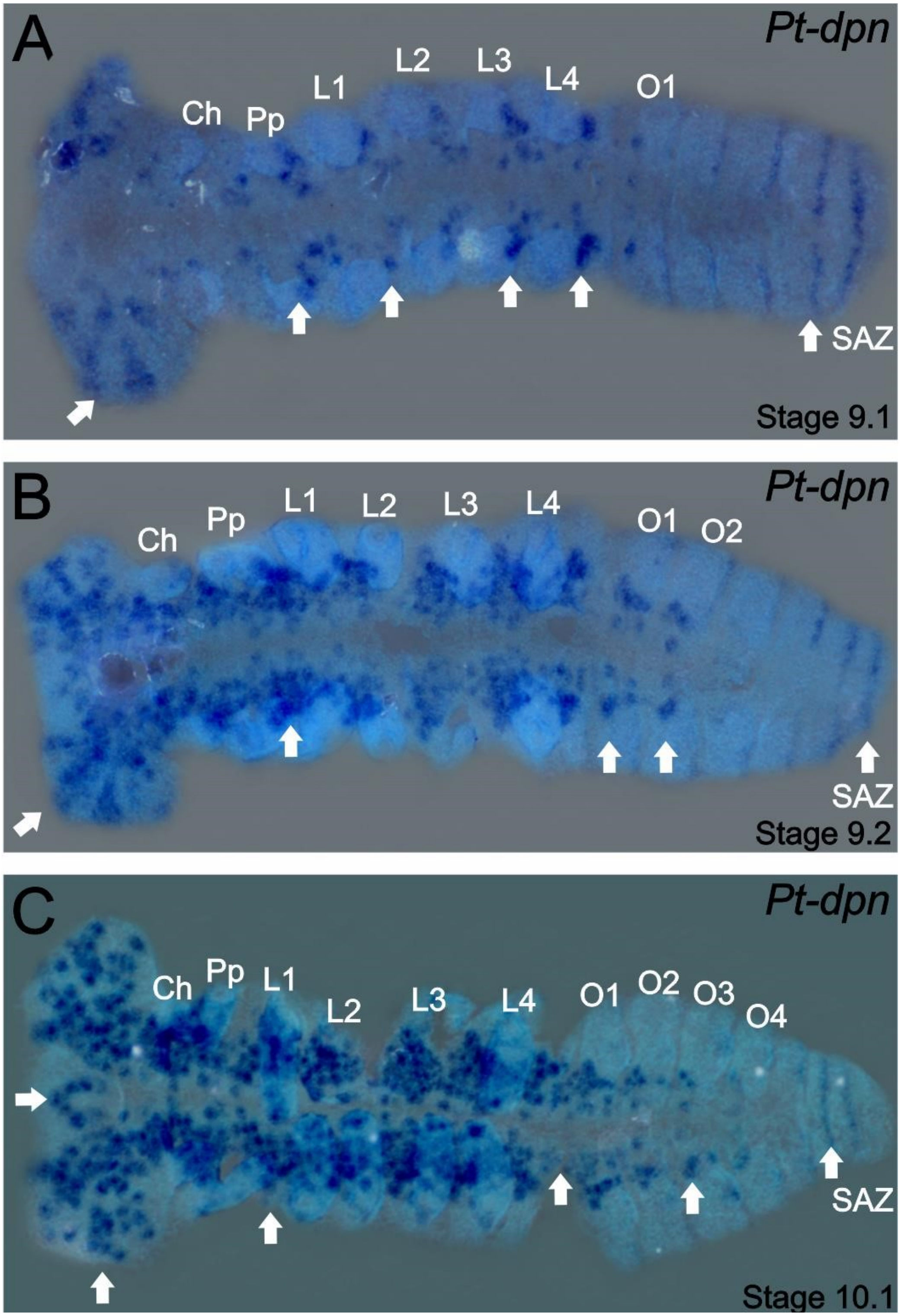


**Figure 2. 4. Expression of a *deadpan* orthologue in *P. tepidariorum*.** A) In situ hybridization showing the expression of this orthologue in the spider CNS at stage 9.1, in clusters of differentiating neurons (white arrows). Expression is also observed in the segments of the opisthosoma and SAZ. B and C) *Pt-dpn* is highly expressed in the CNS in stages 9.2 and 10.1,

but weaker in the opisthosomal segments. Ch: chelicerae, L1 – L4: prosomal segments 1 to 4, O1 – O4: opisthosomal segments 1 to 4, Pp: pedipalps; SAZ: segment addition zone.

For *Pt-hey* (aug3.g5209) strong expression was observed in the CNS at stage 9.2, specifically in the mushroom bodies and in internal clusters of the PNS throughout the entire ventral nerve cord (Figure 2.5 – A). *Pt-side* (aug3.g20968) exhibits expression in both the anterior neurectoderm, near the future mouthparts, and is weakly expressed in all the segments from the prosoma to the opisthosoma in a stripped pattern (Figure 2.5 – B).

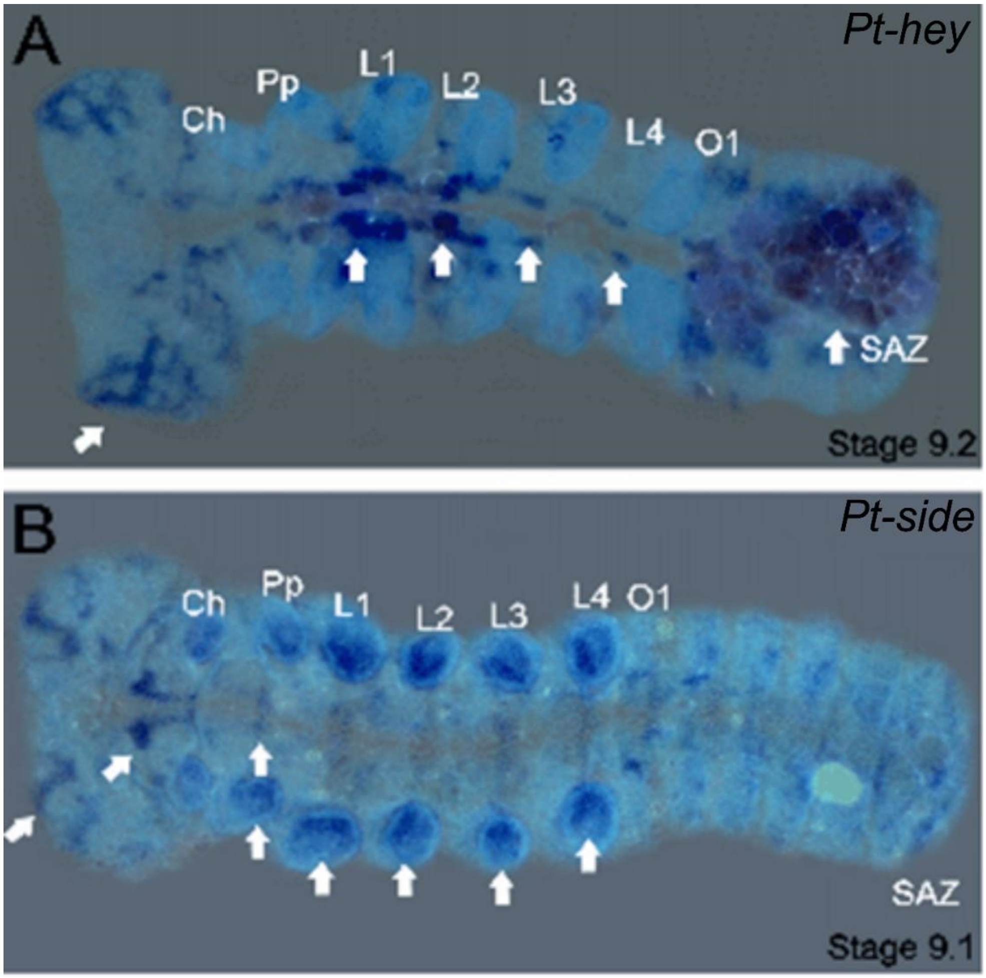


***Figure 2. 5. Expression of Pt-hey and Pt-side orthologues in P. tepidariorum.*** A) At stage 9.2, *hey* shows strong expression in the mushroom bodies and in some clusters of the ventral nerve cord, as indicated by white arrows. The arrow in the SAZ highlights some fragments of yolk that

couldn't be removed prior to the flat mounting. B) *side* is expressed in the brain neuroectoderm at stage 9.1, as well as in mouthparts and all the segments in the ventral nerve cord. The apparent expression in the limbs is background. Ch: chelicerae, L1 – L4: prosomal segments 1 to 4, O1 – O4: opisthosomal segments 1 to 4, Pp: pedipalps; SAZ: segment addition zone.

For the orthologues of *Pt-ash1* and *Pt-ash2,* no transcripts were detected with the set of RNA probes I designed. This might have been due to a probe synthesis that didn't work correctly, but I did not have the time to repeat the in situs for these two orthologues before finishing the thesis.

### Functional analysis of the bHLH genes

As mentioned in the introduction, the Hes-like genes are direct downstream targets of the Notch signaling in the segmentation process in both *D. melanogaster* and vertebrates. Thus, to better understand the role of this major pathway in the spider embryogenesis, I conducted functional analysis for the *Pt-Dl* knockdown embryos and evaluated the *Pt-h* expression. At stage 6 (n= 30) it was possible to observe a complete decrease of *Pt-h* in *Pt-Dl* RNAi embryos (Figure 2.6 – A and B). This suggest that in spiders, *Pt-h* is also dependent of the Delta-Notch pathway to activate its expression.

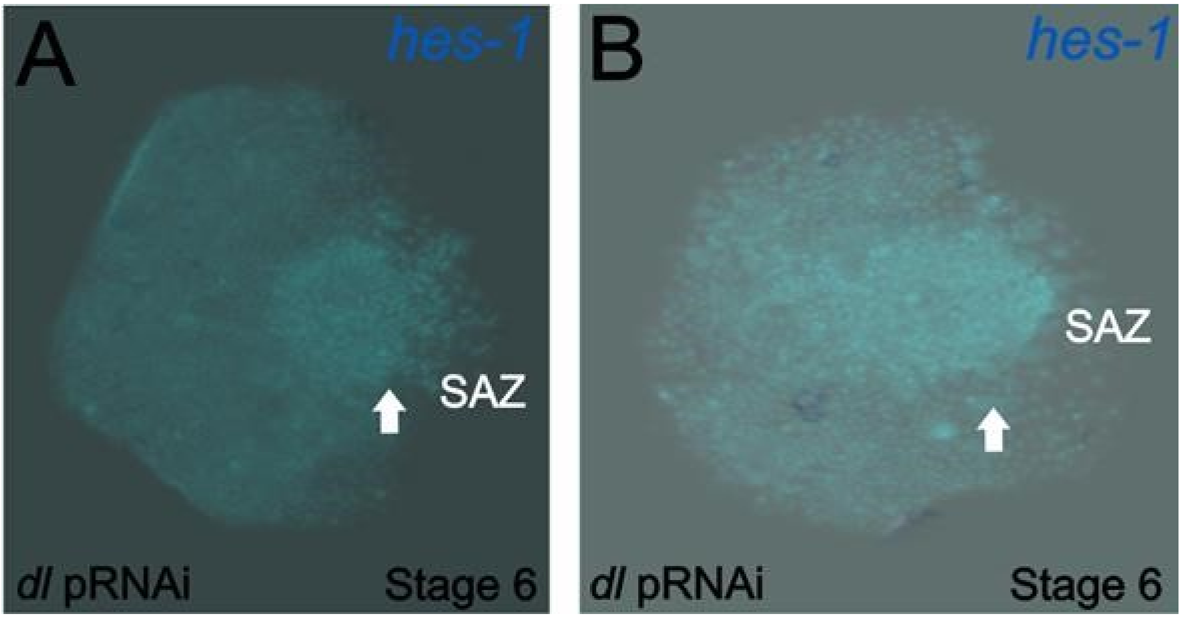


**Figure 2. 6. Expression of *Pt-h* in *Pt-Dl* knockdowns.** A and B) *Pt-h* in situ hybridization in embryos of stage 6, demonstrates the loss of this transcripts in the entire embryo. White arrows show an accumulation of cells in the SAZ, characteristic of *Pt-Dl* knockdowns. SAZ: segment addition zone.

## Discussion

### Conservation of bHLH-containing Hes-like genes in the common house spider

bHLH transcription factors (TFs) are DNA-binding proteins involved with several aspects of development (Murre *et al*., 1994). One of the most studied genes belonging to this complex is the *D. melanogaster* gene *hairy* (Carroll, 1988). This gene, that contains a bHLH domain in its coding sequence, is a member of a conserved gene family that is involved in various processes of embryogenesis in vertebrates and insects: The Hairy and Enhancer of Split (HES) family (Jennings, 1999). These genes encode proteins that have, apart from the basic helix-loop-helix binding site (bHLH), an Orange domain and are involved in the activation of the intracellular domain of the Notch signalling pathway.

As previously stated in Dearden (2015), chelicerates don't appear to have an enhancer-of-split complex, which has been shown to be a pancrustacean novelty amongst arthropods. In my thesis I corroborate the lack of the EoS complex, but my findings are not completely consistent with Dearden's analysis (2015). First, the author placed Hairy and Deadpan in the same group. In my analysis, I found that Deadpan is a single group containing the *Drosophila* and spiders' *deadpan* orthologues, whereas Hairy is grouped with Hey (Figure 2.2). Second, Dearden (2015) grouped *side* with *hey*, while I found the former to fall into a specific Side group with the orthologues from various other species (Dearden, 2015).

The evolution of the metazoans depended much in the plasticity seen for sets of genes arose by duplications. This was shown for the bHLH group in crustaceans in a genomic-wide study (Chang & Lai, 2018), different species of vertebrates (Ledent, 2001) and lophotrocozoans (Bao, 2017). For this reason, I focused my thesis in the identification of Hes-like genes of the common house spider, and further analyse if these genes show any aspect of neo or subfunctionalization, that would be an indicative of genes acquiring different functions as a result of the WGD.

### ***hairy*** **is expressed in a segmental pattern**

I first analysed the expression of *Pt-h*, the orthologue of the pair-rule gene involved in segmentation in *Drosophila* (Ingham, 1985). As mentioned in the introduction, in spiders *hairy* has the same pattern of expression in both *C. salei* (Damen, 2000) and *P. tepidariorum* (Pechmann, 2009; this study), which indicates that it is likely to be involved in head patterning and segment addition. What was shown for *P. tepidariorum* is the strong early expression of this gene in center of the germ disc, and then in a dynamic mode in the SAZ and the differentiating segments. This expression is similar to the *h* orthologues in other short-germ arthropods such as the cockroach *Periplaneta americana* (Pueyo, 2008) and the flour beetle *Tribolium castaneum* (see Aranda, 2008). In these two species, *hairy* is expressed in the differentiating growth zone, in a cyclic mode towards the anterior segments.

In the centipede *Strigamia maritima* it was seen a similar pattern as the one seen for *hairy* in Tribolium and Periplaneta (Chipman, 2008). Two *hes-like* genes, namely *StmHES1* and *Stm-HES4*, are expressed in a posterior-to-anterior manner in the SAZ and nascent segments (Chipman, 2008). In the maximum likelihood tree in Figure 2.2, *StmHES4* is grouped in the Hairy clade, exactly as in the cited work. My analysis was not conclusive for the *Stm-HES1* orthologue, because this gene was normally grouped as an outgroup of all of the hes-like genes. This might be because of the incomplete sequence of this gene in the GenBank.

The regulation of *Pt-h* by Wnt signaling was shown by McGregor (2008): the dynamic expression of *hairy* is lost in *Wnt8* pRNAi embryos. Schönauer *et al* (2016) demonstrated that *Wnt8* and the Delta-Notch pathway acts in a feedback loop in the posterior segment addition zone and controls the differentiation of new segments via *caudal* and *even-skipped*. Therefore, I knocked down *Dl* via parental RNAi to analyse what happen to the *Pt-h* orthologue. Strikingly, there was no detectable expression of this gene in embryos of stage 6 (Figure 2.6 – A and B). This was somehow expected if one considers that *Pt-h* is a direct downstream target of the Notch pathway in different metazoans (Cui, 2005; Kageyama, 2007).

The involvement of Delta-Notch pathway was also shown to play an important role in segmentation in the spider *C. salei* (Stollewerk, 2003). *Cs-delta* have two copies (*Csdelta1* and *Cs-delta2*), and the knockdown of the two copies in separated via embryonic RNAi leads to the perturbance of the dynamic expression of *Cs-hairy* in the posterior, but not to the depletion of its transcripts. Thus, there is a certain degree of functional redundancy in the Delta duplicates in *C. salei*, that acts in the activation of *Cs-hairy* in the absence of the other copy.

In vertebrates, the closest orthologues of the *hairy* gene, namely Hes genes in mouse and humans and Her genes in zebrafish, have in segmentation (*hes* and *her* 1 and 7) and neurogenesis, and are also dependent on the Notch (Kageyama, 2007). They play a role with the Wnt and FGF signaling in the somitogenesis mechanism in these vertebrates (Pourquie, 2007). Indeed, it is interesting that *Pt-h* also exhibits dynamic expression from the posterior, when compared to the orthologues of zebrafish and mice and is regulated by the same signaling pathways (Muller *et* al., 1996; Patel, 2003; Peel, 2003; Pourquie; 2007), and is also involved in the GRN that involves the *Dl* and *Wnt8* genes in the spider (McGregor *et al.,* 2008; Schonauer *et al*., 2016).

### ***deadpan*, *hey* and *side* are expressed in clusters of neurogenic cells**

*deadpan* and *hey* (H*airy and* E*(spl) related with a* Y*RPW domain)* are involved in the differentiation of neural precursors (neuroblasts) during insect neurogenesis (Bier, 1992; Leimeister, 1999), whereas *side* (Similar to Deadpan) function is to control the migration of motor neurons (Siebert, 2009; Dearden, 2015). Only studies in *D. melanogaster* are available for these three genes, and just a few of these have functional data available (Leimeister, 1999).

deadpan has a role in neurogenic differentiation but is also shown to have a gap expression pattern in earlier stages in *Drosophila*, that clears and become restricted to the neuroblasts (Bier, 1992). The closest orthologue of this gene in the annelid *Capitella teleta*, *Cte-hes2* is expressed in the brain, in the chaetal sacs and in the elongating part of the trunk, suggesting a role for this orthologue in neurogenesis and segmentation (Thamm, 2008).

In vertebrates, the closest orthologue of *deadpan* is *hes2* and for example in mice, this gene is expressed in the CNS and PNS, as well in some stripes in the vertebrae, but no functional studies have been carried out (Nishimura *et al*., 1998).

I showed that the spider orthologue of *deadpan* (*Pt-dpn*) has a similar later stages expression to the one seen for *Pt-h* in the spider as well as a stripped pattern in the opisthosoma and SAZ that becomes weaker at stage 10.1. The expression of *Pt-dpn* in the proneural clusters becomes progressively higher from stage 9.1 to 10.1. In the absence of functional data, it is difficult to be certain of its role, but I suggest from the expression pattern that this gene is involved with both the segmentation and neurogenesis during embryogenesis in this spider.

As mentioned in the introduction chapter, *hey* and *side* both have functions in differentiation of neural clusters in *Drosophila* (Leimeister, 1999; Siebert, 2009; Dearden, 2015), and also appear to be involved with neuronal differentiation and axonal migration, acting as CNS and PNS regulators, which likely corresponds to the ancestral role for the Hes genes in vertebrates (Sagner *et al*., 2018).

The expression pattern of the spider hey and side genes seems to be conserved compared to what is seen in *Drosophila* with expressed observed in the CNS and PNS. These two genes spider genes grouped with the *hes7* gene from mice, which has a role specifically in the posterior part of the pre-somitic mesoderm and is involved in segmentation (Kageyama, 2007). However, neither of these genes appears to be involved in the spider segmentation, instead, I suggest that their role is restricted to neuronal differentiation.

It is clear that to better understand the function of the bHLH-containing genes in spider embryogenesis functional experiments are needed. However, these experiments should aim to knock-out of the specific genes together and in combination because it is known that the bHLH-containing genes have a compensatory mechanism in retinal development of vertebrates (Cho, 2007), and that they also have a high level of functional redundancy (Kageyama, 2007).

# Chapter 2

# Expression of Sox Genes in *P. tepidariorum*

## Introduction

The Sox family of transcription factors is an important part of the genetic ‘toolbox’ of all metazoans examined to date and has been shown to play a range of developmental roles in vertebrates and insects. In *Drosophila*, these genes are involved with many different aspects of embryogenesis, from stem cell maintenance, cell proliferation and neural tissue specification (Kamachi, 2013; Sarkar, 2013; Reiprich, 2015). However, outside

*Drosophila* little is known about the extent or conservation of the Sox family in other arthropod species.

Chelicerates (e.g. spiders, horseshoe crabs and scorpions) are the sister group to mandibulates among arthropods and can therefore be informative about ancestral features of these animals. Furthermore, it is known that some of these animals, including spiders, have undergone a round of genome duplication in their evolutionary history, which offers an interesting perspective on the evolution of their genetic toolkit (Schwager *et al*., 2017). Therefore, to expand the current knowledge of the evolution and expression of the Sox genes in arthropods, Professor Steve Russell (University of Cambridge) and I isolated, and I characterised the expression of these genes in the common house spider *P. tepidariorum*. The results presented in this chapter are included in a manuscript entitled "Duplication and divergence of Sox genes in spiders", submitted in bioRxiv (Paese *et al*., 2017) in November of 2017 and for review to the journal BMC Evolutionary Biology.

## Isolation of Sox genes in spiders

In order to characterise the Sox gene complement of spiders, Professor Russell conducted TBLASTN searches of the genomes of *P. tepidariorum* (Paese *et* al., 2017; Schwager *et al*., 2017) and *S. mimosarum* (Sanggaard *et al*., 2014) using the HMG domain of the mouse Sox2 protein, recovering 16 and 15 sequences respectively.

Phylogenetic reconstruction of the spider Sox genes was generated with MUSCLE sequence alignments and PhyML maximum likelihood phylogenies using the HMG domains recovered from the BLAST searches, along with those from the eight *D. melanogaster* Sox genes and representatives of each subgroup from mouse. These analyses resulted in a clear classification into groups B-F as found in other invertebrate genomes (Figure 3.1). Note that Group A only contains the *SRY* gene specific to eutherian mammals and there are no Group G, H or I Sox genes found outside the vertebrates. Supporting this classification, phylogenetic trees constructed with the full-length sequences of the predicted spider Sox proteins and those from *Drosophila* yielded virtually identical results (Figure 3.1).

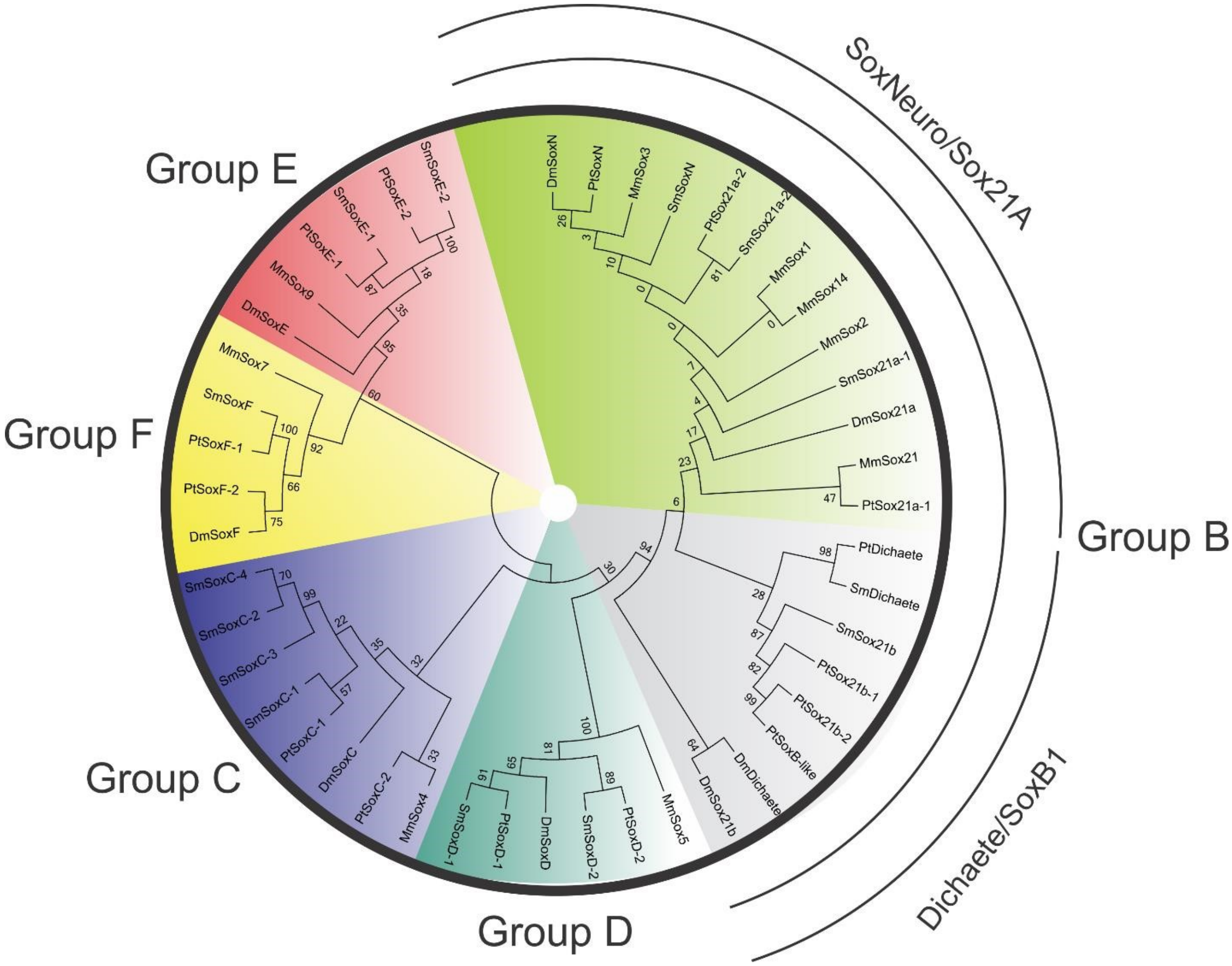


**Figure 3. 1. Phylogeny of Sox HMG domains in selected metazoans.** Phylogenetic tree showing the relationship between *M. musculus* (Mm), *D. melanogaster* (Dm), *P. tepidariorum* (Pt) and *S. mimosarum* (Sm) Sox genes based on HMG domain sequences. The grouped genes are divided into different colours as highlighted outside the circle.

All but three of these contained the highly conserved RPMNAFMVW motif that is characteristic of Sox proteins and three (*ptSoxC-2*, *ptSoxB-like* and *ptSox21b-2*) only show minor conservative substitutions. 14 of the *P. tepidariorum* sequences corresponded to annotated gene models (Paese *et al*., 2017). Two sequences were identical (*ptSox21b-1*, aug3.24914.t1 and aug3.g24896.t1), since the latter maps to a genomic scaffold of only ~7 kb it was presumed that this represents an assembly error and is thus considered  as a single gene. One genomic scaffold encoding a Sox domain (*ptSoxB*-like, Scaffold3643:28071..28299) is in a region of poor sequence quality and one cannot be sure it represents a *bona fide* gene but we have nevertheless included it in the subsequent analysis (Paese *et al*., 2017).

In common with many other gene families in spiders (Schwager *et al*., 2017), the Sox genes are mostly represented by two or more copies in each group (Figure 3.2). In other arthropods examined to date, as well as the onychophoran *Euperipatoides kanangrensi*, there is usually only a single copy of each gene, although there is a recent report of two Group E genes in the millipede *G. marginata* (Janssen *et al*., 2018). In the case of spider Groups D and E, the duplications likely predate the divergence of spiders as Professor Russell also show that for *Stegodyphus mimosarum* (Figure 3.1) (Paese *et* al., 2017). With Group F, there is only one gene identified in *S. mimosarum* but two in *P. tepidariorum*. In the case of group C, there appears to have been additional duplication events in *S. mimosarum* (Paese *et* al., 2017).

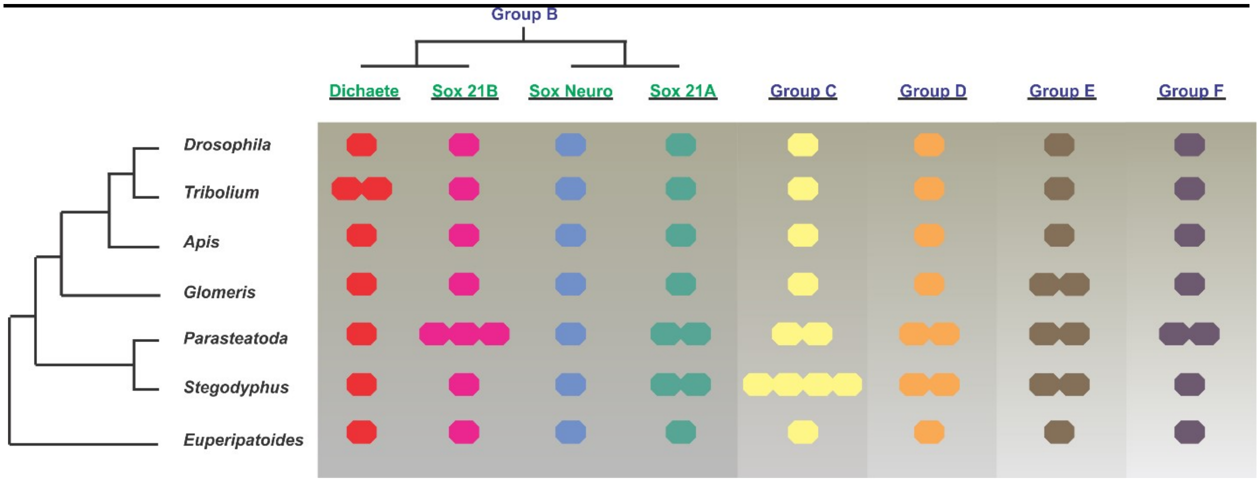


**Figure 3. 2. Repertoire of Sox genes in selected arthropods.** Diagrammatic representation of the complement of Sox genes in insects (*D. melanogaster*, *T. castaneum* and *A. mellifera*), the spiders (*P. tepidariorum* and *S. mimosarum*), the myriapod (*G. marginata*) and an onychophoran (*E. kanangrensis*). Each coloured circle represents a gene.

I also aligned the amino acid sequences of the HMG domain of the duplicated Sox genes found *in P. tepidariorum.* (Figure 3.3). This showed a high similarity of amino acid sequences for group D, E and F duplicates, whereas in contrast the duplicated SoxB and SoxC proteins show a high number of amino acid changes in their sequences. This suggest that these aminoacid changes could have lead to binding differences in the orthologues, thus a difference in expression patterns in genes that belongs to the same Sox group.

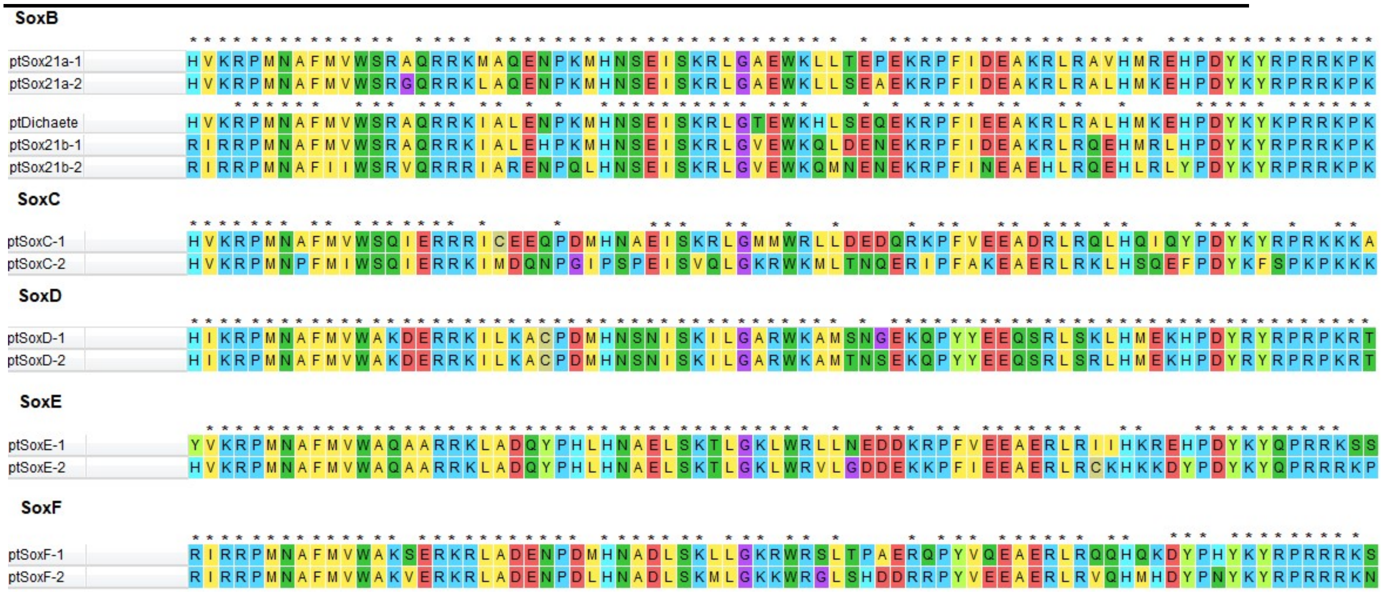


**Figure 3. 3. Alignment of the Sox orthologues aminoacids by ClustalW.** SoxB group genes were divided in the duplicated orthologues for *PtSox21a*, while *Dichaete* was aligned together with duplicated versions of *PtSox21b* genes. SoxC group is the one that shows the higher amino acid changes in their sequences, whereas the SoxD group is the most conserved in their protein sequence.

## Sox Gene Expression during *P. tepidariorum* embryogenesis

I next studied the expression of Sox genes during embryogenesis in *P. tepidariorum* and will present these results in alphabetical order of the groups below.

### SoxB group

For the SoxB family genes *ptSox21a-1*, *ptSox21a-2*, *ptSox21b-2* and *Dichaete*, no expression was detected during embryogenesis. This indicates that they might only be expressed at very low levels or in a few cells or that these genes are used during postembryonic development.

### *SoxN*

*pt-SoxN* was first isolated by Akiyama-Oda & Oda (2016). This gene was annotated in their paper as *sox2*, and only the expression at stage 9 in the nervous system was mentioned. Therefore, I aimed to characterize the expression of SoxN more fully. I found that expression of *pt-SoxN* commences at mid-stage 7 in the most anterior part of the germ band, a region corresponding to the presumptive head (Figure 3.4 - A). *ptSoxN* is subsequently expressed broadly in the developing head and follows neurogenesis in a progressive anterior-to-posterior pattern as new segments are added (Figure 3.4 - B). By mid-stage 9, *ptSoxN* is strongly expressed in the head lobes and in the ventral nerve cord (Figure 3.4 - C), however, after this stage no further expression was detected.

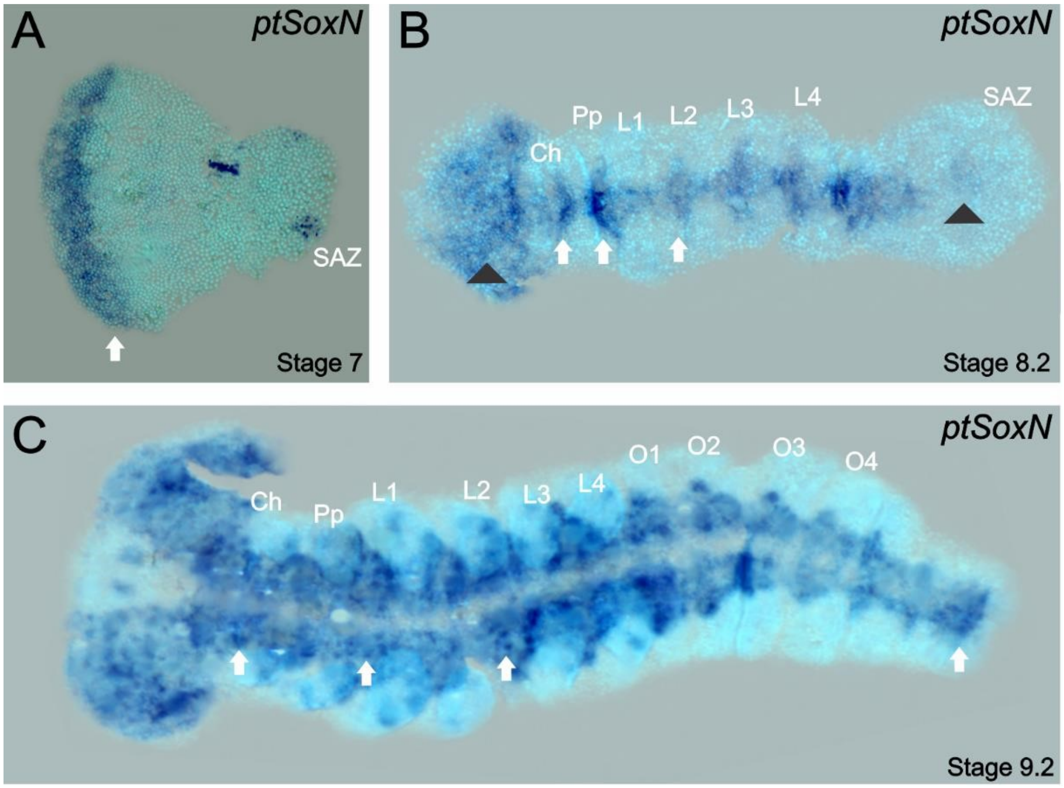


**Figure 3. 4. Expression of *ptSoxN*.** Flat-mounted embryos at different stages of development after RNA in situ hybridization. A) *ptSoxN* expression is restricted to the presumptive neuroectoderm in the most anterior region of the germ band at stage 7 (arrow). B) At stage 8.2, broad expression can be observed in the most anterior part of the embryo (black arrowhead) and in the ventral nerve cord appearing sequentially from anterior to posterior: white arrows indicate expression in clusters that will subsequently broaden, while expression in the posterior region adjacent to the SAZ is also observed (black arrowhead). C) At stage 9.2 expression is observed throughout the ventral nerve cord, with differentiating cell clusters indicated by arrows. Ch: chelicerae, L1 – L4: prosomal segments 1 to 4, O1 – O4: opisthosomal segments 1 to 4, Pp: pedipalps; SAZ: segment addition zone. Ventral views are shown for all embryos with the anterior to the left.

### *Sox21b-1*

Another member of the SoxB group, *Sox21b-1*, is expressed maternally and localized to in the Balbiani's body of pre-vitellogenic *P. tepidariorum* oocytes (Jędrzejowska, 2005) (Figure 3.5 - A). However, after fertilization *Sox21b-1* is not zygotically expressed until early stage 5, when weak expression is detected throughout the germ disc, with stronger expression in more central cells (Figure 3.5 - B). At late stage 5, expression becomes more restricted to the centre of the germ disc (Figure 3.5 - C). During stages 5 and 6, the cumulus migrates to the rim of the germ disc, opening the dorsal field and giving rise to an axially symmetric germ band (Figure 3.5 - D) (Mittmann & Wolff, 2012). In early stage 6 embryos, *Sox21b-1* can be observed in the middle of the presumptive prosoma in a broad stripe (Figure 3.5 - D), which develops further during stage 7 in the region where the leg bearing segments will form (Figure 3.5 - E). During these and subsequent stages, dynamic expression of *Sox21b-1* is observed in the SAZ and the most anterior region of the germ band that will give rise to the head segments (Figure 3.5 - F-H).

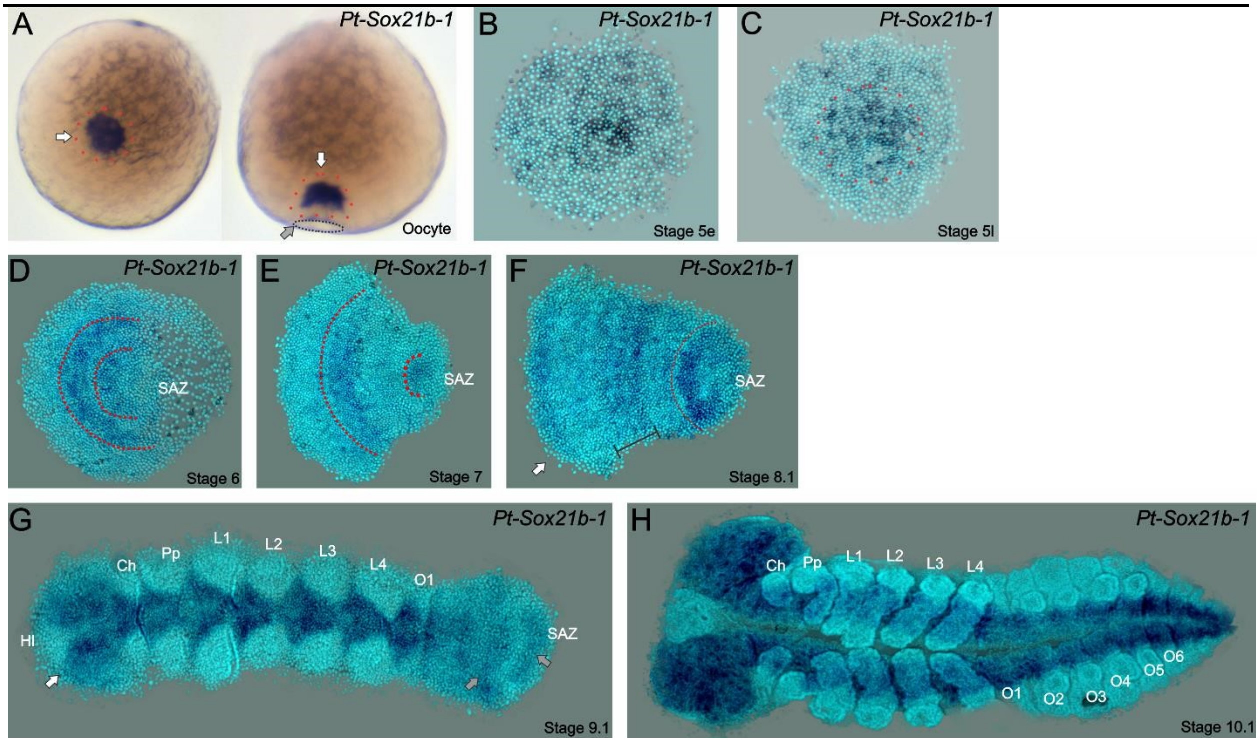


**Figure 3. 5. Expression of *Sox21b-1* in *P. tepidariorum* oocytes and embryos. A)** Dorsal (left) and lateral (right) views of pre-vitellogenic oocytes showing *Sox21b-1* mRNA in the Balbiani's body (red dashed circle and white arrows). The sperm implantation groove is indicated by a black dashed circle and grey arrow. **B**) At early stage 5, the expression of *Sox21b-1* appears in a salt and pepper pattern in the germ disc. **C**) Expression in the cumulus becomes stronger at late stage 5, with lower expression at the periphery of the germ disc (dashed red circle). **D**) At stage 6, Sox21b-1 is expressed in a broad stripe in the anterior (between the red dashed lines). **E**) At stage 7, there is expression in the region of the presumptive leg bearing segments and in the SAZ (both indicated by red dashed lines). **F**) At stage 8.1, *Sox21b-1* is still expressed in the SAZ and the presumptive leg bearing segments, but nascent expression is observed at the anterior of the germ band (indicated by the white arrows and black brackets). **G**) At stage 9.1, when the limb buds are visible expression of *Sox21b-1* becomes restricted to the ventral nerve cord (anterior white arrow) and can be observed in the SAZ (posterior grey arrows). **H**) At stage 10.1, *Sox21b1* expression is restricted to the ventral nerve cord and the head lobes. Ch: Chelicerae; HL: Head lobes; L1 to L4: Prosomal leg bearing segments 1 to 4; O1 to O6: Opisthosomal segments 1 to 6; SAZ: Segment addition zone. Ventral views are shown with anterior to the left, except as described for oocytes.

The widespread expression of both *SoxN* and *Sox21b-1* throughout the neuroectoderm strongly suggest that, as has been shown in vertebrates and flies, many cells in the developing CNS co-express these two related Sox genes. To confirm if their expression overlaps in the CNS, but not in the SAZ, I performed dual colour fluorescent *in situ* hybridisations with *SoxN* and *Sox21b-1* probes (Figure 3.6 - A-C). While expression of both genes clearly shows extensive overlap throughout the developing CNS, I was interested to note that at the very lateral regions of the neuroectoderm, *Sox21b-1* is expressed exclusively. As shown by the single gene in situs, only Sox-21b-1 is expressed in the SAZ (Figure 3.6 – A and B).

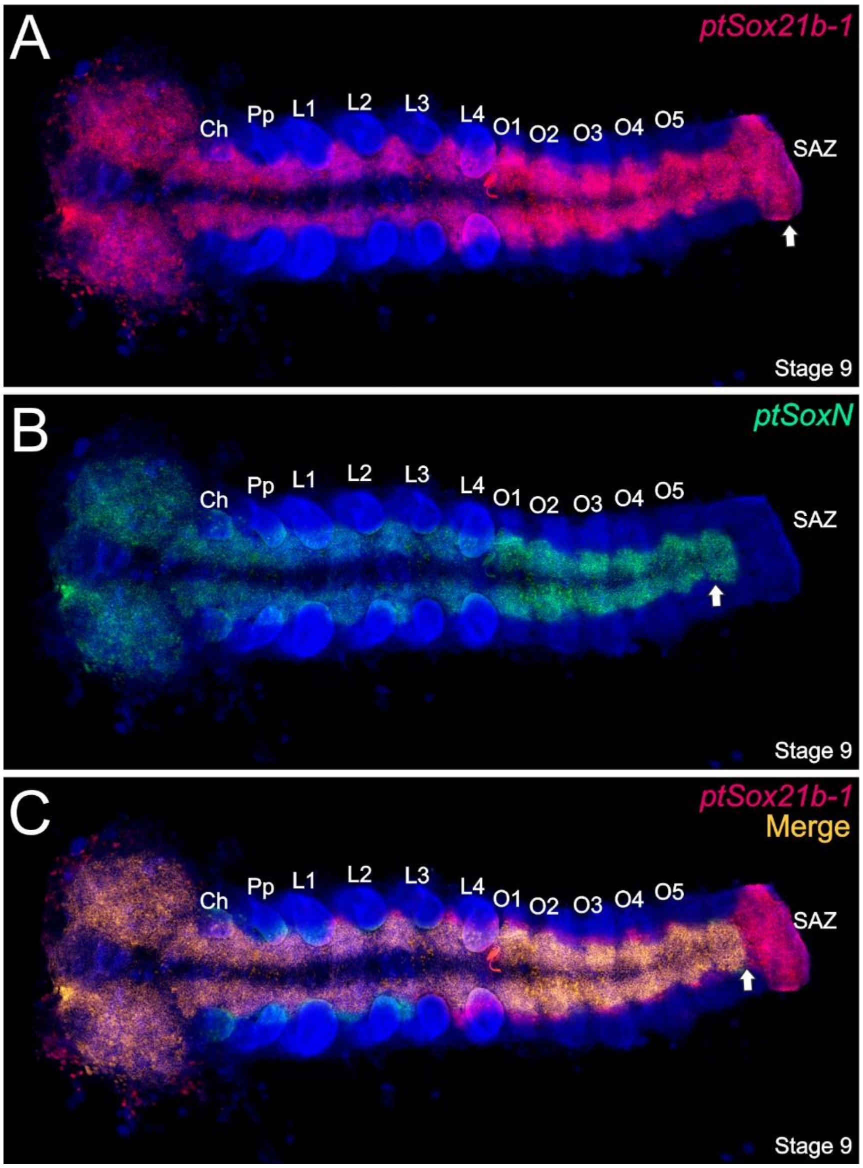


**Figure 3. 6. Double fluorescent in situ hybridization for *PtSox21b-1* and *PtSoxN*.** Dual colour in situ hybridization with (A) digoxigenin-labelled *PtSox21b-1* in red and (B) fluorescein *Pt-SoxN* in green. C) Merged figures A and B shows the overlap of these orthologues in the CNS, but lateral expression of only *PtSox21b-1*, and no expression of *PtSoxN* in the SAZ.

## SoxC group

For the Sox C group genes, I did not detect any expression for *ptSoxC-2*. However, *ptSoxC-1* expression was detected at mid-stage 6, in a pattern similar to that of *ptSoxN* in the presumptive head and anterior segments (Figure 3.7 - A). By stage 8.2, expression is apparent in neuroectodermal progenitors along the germ band and at the anterior region of the SAZ (Figure 3.7 - B), however by stage 9.1 (Figure 3.7 - C) expression is lost from the SAZ. Interestingly, from stage 9.1 onward, *ptSoxC-1* is expressed in the ventral nerve cord, from the head to the SAZ, however unlike the uniform expression of *ptSoxN*, *ptSoxC-1* is observed in clusters of cells, presumably undergoing neurogenic differentiation, progressively from the head through to opisthosomal segments as they differentiate in an anterior to posterior manner (Figure 3.7 - C).

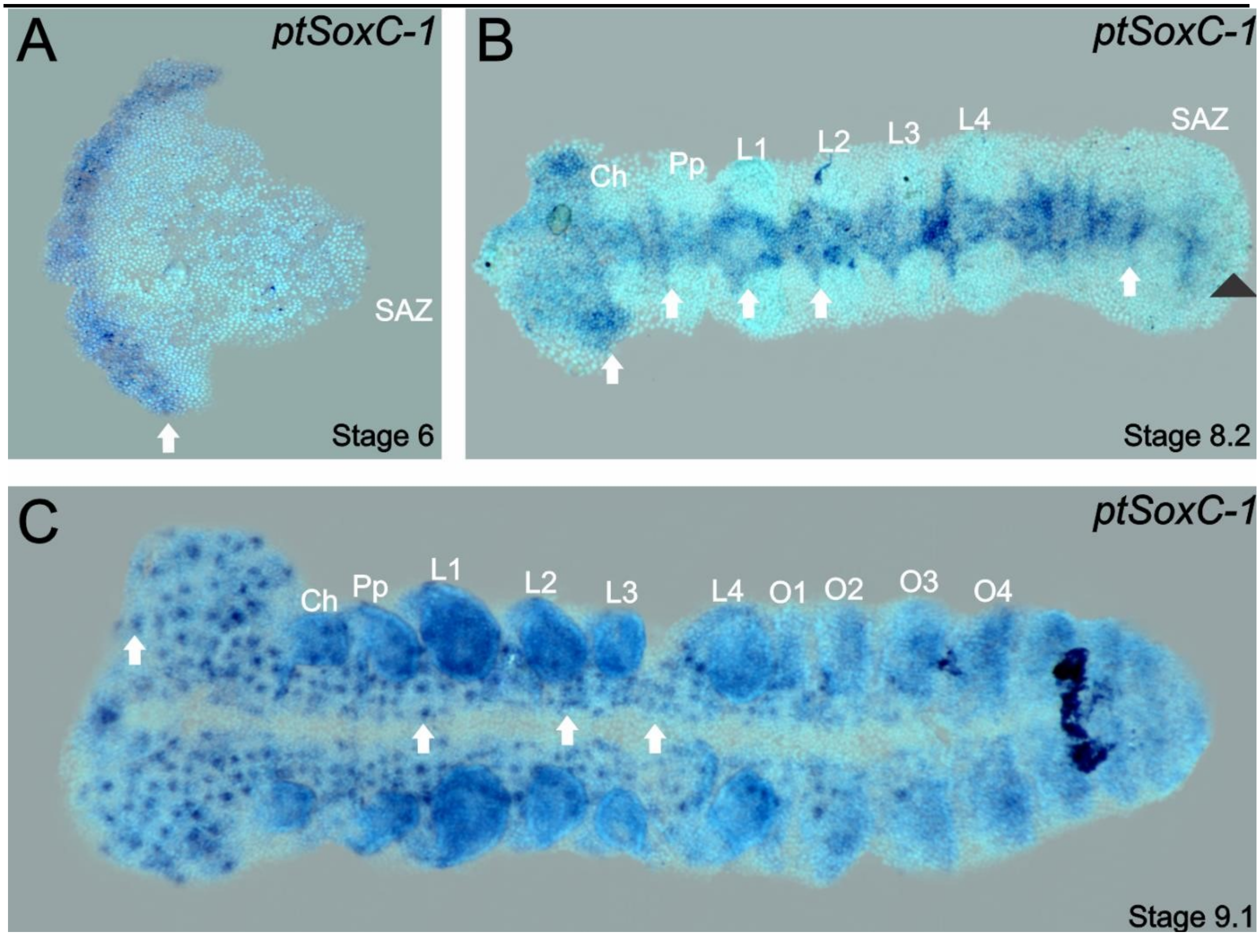


**Figure 3. 7. Expression of *ptSoxC-1*.** Flat-mounted embryos at different stages of development after RNA in situ hybridization. A) *ptSoxC-1* is strongly expressed in the presumptive neuroectoderm at stage 6 as indicated by the white arrow. B) At stage 8.2, strong expression is observed in the ventral nerve cord (white arrows) with the exception of the most posterior part of the SAZ (black arrowhead) C) At stage 9.1, expression is apparent in clusters of cells in the head and each anterior segment up to the third opisthosomal segment (O3): white arrows indicate localized expression. The signal in the limb buds is background and staining at the most posterior part of the O5 segment is an artefact of incomplete chorion removal. Ch: chelicerae, L1 – L4: prosomal segments 1 to 4, O1 – O4: opisthosomal segments 1 to 4, Pp: pedipalps; SAZ: segment addition zone. Ventral views are shown for all embryos with the anterior to the left.

## SoxD, E and F groups

Two genes in each of the *SoxD*, *E* and *F* families were identified. However, I did not detect any expression of Sox*D-2, SoxE-2* or *SoxF-1* during the *P. tepidariorum* embryonic stages examined. For *ptSoxD-2* I found no expression prior to stage 10, when I observed expression in the ventral nerve cord from the head to the most posterior part of the opisthosoma (Figure 3.8 - A).

*ptSoxE-1* is expressed in the developing limbs from stage 9 in small dots in the chelicerae, pedipalps and L1 buds, broader expression in L2 and L3, and in two dots in the L4 limb pairs, presumably corresponding to the differentiating peripheral nervous system (PNS) (Figure 3.8 - B).

Finally, the expression of *ptSoxF-2* is only detected at stage 9, in small domains at the tips of the L1 segment limb buds, but not in any other limb buds (Figure 3.8 - C).

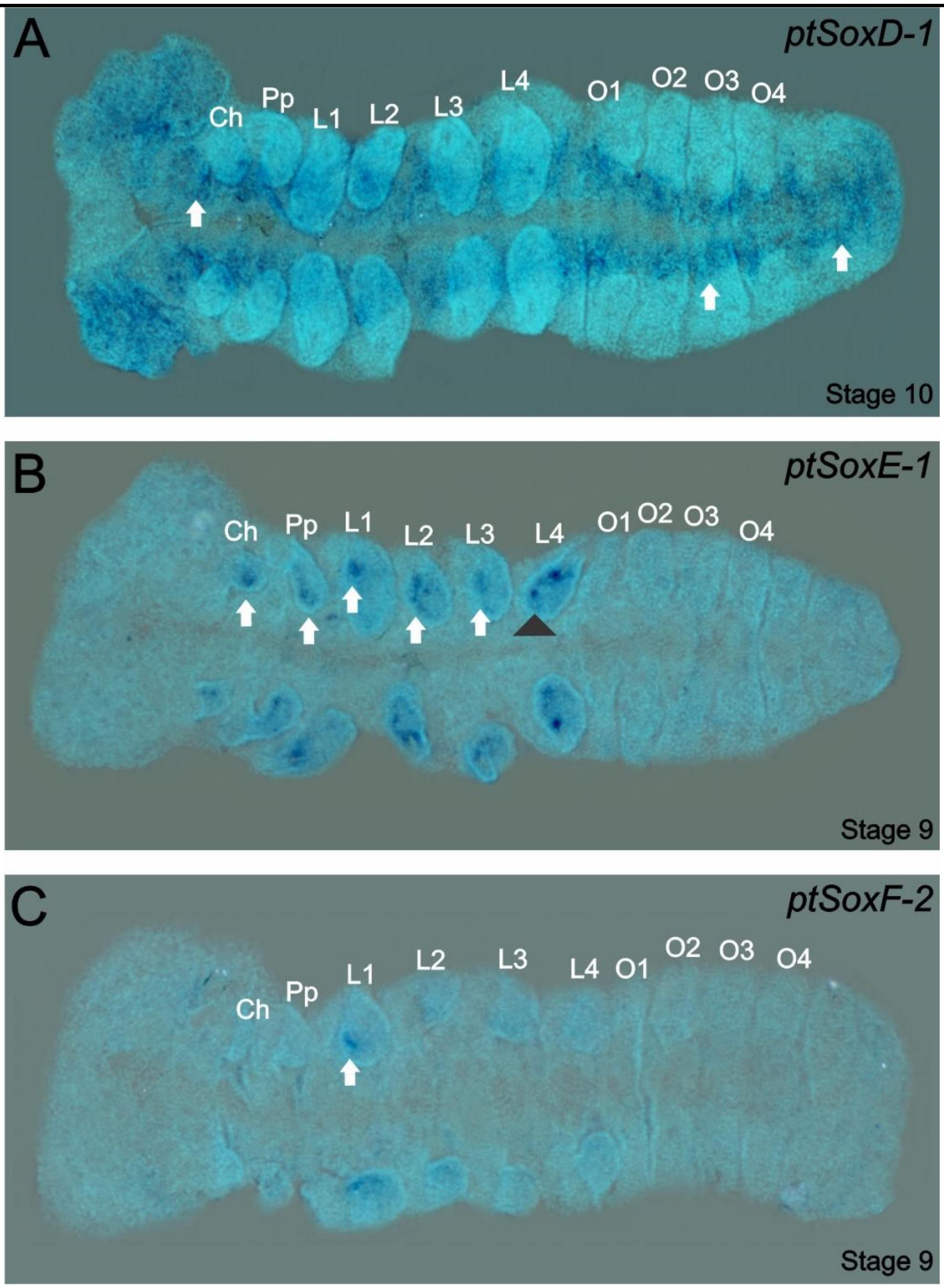


**Figure 3. 8. Expression of Sox D, E and F group orthologues.** In situ hybridization on flatmounted embryos at different stages of development. A) *ptSoxD-1* expression is observed throughout the ventral nerve cord in stage 10 embryos as indicated by the arrows. B) *ptSoxE-1* expression at stage 9 is visible as single dots in the forming chelicerae, broader expression in the pedipalps and L1 to L3 (white arrows), and as two dots in the L4 limb bud as indicated by the black arrowhead. C) The expression of *ptSoxF-2* is only visible in the L1 limb buds at stage 9 (arrows). Ch: chelicerae, L1 – L4: prosomal segments 1 to 4, O1 – O4: opisthosomal segments 1 to 4, Pp: pedipalps; SAZ: segment addition zone. Ventral views are shown for all embryos with the anterior to the left.

## Discussion

### Conservation and divergence of Sox Gene Expression

In this chapter I outlined the identification and expression analysis of Sox genes in the common house spider *P. tepidariorum*. This gene family is conserved throughout metazoans as shown in cnidarians (Magie, 2005), molluscs (Focareta, 2016), different species of arthropod (Wilson & Dearden, 2008; Janssen *et al*, 2017) and vertebrates (Lefebvre, 2010).

First, it was shown by our group work that the genes which belong to the Sox family can be grouped accordingly to the HMG domain (Kamachi, 2013). The amino acid alignment of the HMG domain of the paralogs from the different groups suggests that the duplicated copies for each group have both high (groups B and C) and also low divergence (groups D, E and F). This might explain the differences in expression patterns of the duplicated orthologues, that these aminoacid changes could lead to the binding of different set of proteins, and this can be a strong argument in support of the WGD event that *P. tepidariorum* undergo during its evolutionary story (Schwager *et al*., 2017). Since the Sox group A is vertebrate specific, I will discuss the expression of the orthologues from groups B to F hereafter.

*ptSoxN* is expressed from the beginning of the germ band stage in the anteriormost region of the embryo, in a prominent stripe where the future head will develop (Figure 3.4). This head-specific expression in *P. tepidariorum* is similar to early expression of *SoxN* observed in *D. melanogaster* (Cremazy, 2000) and in *A. mellifera*, where *SoxB1* is expressed in the gastrulation fold and the anterior part of the presumptive neuroectoderm

(Wilson & Dearden, 2008). The anterior-to-posterior pattern of expression found for this orthologue might reflect the differences in ectodermal potential at the germ band stage. The ectoderm at this stage can be divided in epithelial and neuroectodermal, and the gradient of *ptSoxN* expression towards the posterior might reflect this conserved feature seen in both *D. melanogaster* and *A. mellifera.* In summary, epithelial cells are patterned to become neuronal after the activation of *SoxN* expression, in the waves from the anterior to the posterior, and not all at the same time (Buescher, 2002; Overton, 2002; Wilson & Dearden, 2008).

In chelicerates, neurogenic progenitors have been shown to delaminate in clusters of cells rather than single neuroblast-like cells found in dipterans and some hymenopterans (Stollewerk, 2006). However, even given these different modes of neurogenic differentiation, the expression of *SoxN* orthologues suggest this gene performs the same function in chelicerates and insects. Indeed, the recent study by Janssen *et al* (2018), of *T. castaneum*, *E. kanangrensis* and *G. marginata* also shows that the *SoxN* orthologues in these species have widespread and early neuroectodermal expression. Taken together these data clearly support the view that throughout the Bilateria a SoxN class protein is a marker of the earliest stages of neural specification. The lack of expression of *SoxN* after stage 9 in *P. tepidariorum* is probably because after the ventral nerve cord is specified for neural fate, *ptSoxN* expression is reduced and other Sox factors acts patterning the specific neuronal clusters.

### A SoxB1 orthologue has segmentation expression pattern

The second gene that I analysed the expression of was an orthologue of *D. melanogaster Sox21B*. *P. tepidariorum* has two orthologues of this gene, *Sox21b-1* and *Sox21b-2*. This gene is expressed in the oocytes, then reappears in the germ disc at stage 5 as scattered dots in the center (Figure 3.5 – A – C). The expression then localizes in the nascent prosomal segments and in the posterior segment addition zone (SAZ) from stage 6 (Figure 3.5 – D and E). At stage 8.2 expression is observed in the most anterior part of the germ band, which corresponds to the presumptive neuroectoderm in the future head and prosomal segments (Figure 3.5 - F). At stages 9 and 10, strong expression is apparent throughout the ventral nerve cord, similar to *ptSoxN*. Comparing expression in the SAZ at different stages in these fixed preparations suggest that Sox21b-1 may be dynamic in this region (Figure 3.5 – G and H).

In *T. castaneum*, *Sox21b* shows similar expression to *D* early in the SAZ and then in the developing CNS (Janssen *et al*., 2018). In *E. kanangrensis* and *G. marginata,* there is no early *Sox21b* expression during embryogenesis (Janssen *et al*., 2018), however in these species, *D* is expressed during segmentation and then later in the CNS. This suggests that the role of *D* in segmentation in *D. melanogaster* and *T. castaneum* (Clark & Peel, 2018) could extend to *E. kanangrensis* and *G. marginata* while Sox21b has a different role later in development or in adults of these animals. However, in spiders the closely related *Sox21b-1* gene may play the role of early segmentation and late neurogenesis.

Later in development, the expression of *Sox21b-1* in spiders resembles that of *SoxN* in *D. melanogaster* (Niwa, 2016). The expression pattern is similar to that of both *SoxN* and *Dichaete* in *D. melanogaster*, which are expressed in neuroblasts of the neuroectoderm and then differentiating neurons in the ventral nerve cord (Aleksic, 2013) (Figure 17 – G-H). This spider SoxB gene overlaps in expression with *ptSoxN*, as I showed with the double fluorescent in situ hybridization (Figure 3.6). In *Drosophila*, *SoxN* shows a unique lateral expression domain (Buescher, 2002), whereas in the spider, *ptSox21b-1* is more broadly expressed in the lateral part of the ventral nerve cord compared to *ptSoxN.* This might suggest that this orthologue is maintaining the undifferentiated state of the cells, while *ptSoxN* is involved in the patterning of the neuronal ectoderm.

### SoxC to F groups have divergent expression patterns

I found that the gene annotated as *Pt-SoxC-1* has an interesting expression pattern during head development, that becomes restricted to the proneural clusters at stage 9. In *D. melanogaster* the single *SoxC* gene has been shown to play a role in the response to ecdysone at the onset of metamorphosis and has no known role in the embryonic CNS (Ritter, 2013). In contrast, the vertebrate *SoxC* genes (*Sox4, 11* and *12*) play critical roles in the differentiation of post-mitotic neurons, acting like Group B genes, which specify neural progenitors (Tanaka, 2004). In *A. mellifera,* late expression of the SoxC gene was observed in the embryonic cephalic lobes and in the mushroom bodies (Wilson & Dearden, 2008). Interestingly, the expression of SoxC orthologues was observed in the embryonic CNS of other invertebrates (Janssen et al., 2018). Taken together my results and those of others suggest SoxC genes may play a conserved role in aspects of neuronal differentiation, which has been lost in *D. melanogaster*. Interestingly, a comparison of target genes bound by Sox11 (a SoxC orthologue) in differentiating mouse neurons and SoxN in the *D. melanogaster* embryo shows a conserved set of neural differentiation genes, suggesting that in *D. melanogaster* the role of *SoxC* in neurogenesis has been taken over by *SoxN* (Bergsland, 2011).

In the SoxD group, I found similarities to what has been observed for the *Drosophila SoxD* gene, which is also expressed during later stages of embryonic CNS development (Cremazy, 2001) and has been shown to play roles in neurogenesis in the larval CNS (Li *et al*., 2017). While *SoxD* is reported to be ubiquitously expressed in *A. mellifera* embryos, it is also expressed in the mushroom bodies of the adult brain (Wilson & Dearden, 2008). Embryonic brain expression of SoxD orthologues in beetles, myriapods and velvet worms (Janssen et al., 2018), as well as a known role for *SoxD* genes in aspects of vertebrate neurogenesis (Cremazy, 2001; Tanaka, 2004; Lefevbre, 2010: Bergsland, 2011) again suggests conserved roles for *SoxD* during metazoan evolution.

The expression pattern seen for *Pt-SoxE-1* (Figure 3.8 - B) suggests that it is restricted to mechanoreceptors in the limbs of the spider. These receptors are distributed all over the body, but the trichobothria only appear on the extremities of the limbs (Stollewerk, 2001) where they differentiate from PNS progenitors. This is somewhat similar to what is seen in vertebrates, where Group E genes are required in neural crest cells that contribute to the PNS (Bell *et al*., 1997; Stolt, 2010; Kamachi, 2013). The pattern seen in the spider has a striking difference for that is seen in *D. melanogaster,* in which *SoxE* is associated with both endodermal and mesodermal differentiation, with expression observed in the embryonic gut, malpighian tubules and gonad (Loh, 2000), and it has been shown to be required for testis differentiation during metamorphosis

(Nanda *et al*., 2009). Both the *A. mellifera SoxE* genes are also expressed in the testis (Wilson & Dearden, 2008). Janssen and colleagues (2018), observed expression of *SoxE* genes in other invertebrates, associated with limb buds like the spider, but they also detected posterior expression associated with gonadogenesis. These observations are particularly intriguing since the vertebrate *Sox9* gene has a crucial function in testis development (Vidal, 2001).

It is thought that the germ line in *P. tepidariorum* arises during late stage 9, in the opisthosomal segments 2 to 6 in clusters of cells expressing *Pt-vasa* (Schwager, 2015). While I did not observe *SoxE* expression associated with early gonadogenesis in this spider, it remains possible that the spider genes are used later in this process, after the hatching of the juveniles. I suggest this because, while the fly *SoxE* gene is expressed from the earliest stages of gonadogenesis, null mutant phenotypes are not apparent until the onset of metamorphosis (Nanda *et al*., 2009).

Finally, the expression of *ptSoxF-2* is only detected at stage 9, in single foci at the tips of the L1 segment limb buds (Figure 3.8 - C). *SoxF* is a gene expressed in the embryonic PNS (Cremazy, 2001) and plays a role in the differentiation of sensory organ precursors in the fruit-fly (Miller *et al*., 2009), whereas in *A. mellifera* the *SoxF* orthologue is expressed ubiquitously throughout the embryo (Wilson & Dearden, 2008). In *T. castaneum*, *E. kanangrensis* and *G. marginata* (Janssen *et al*., 2018). *SoxF* expression is also associated with the embryonic limbs, again suggesting that this was an ancestral function of this Sox family in the Euarthropoda.

Taken together, these results broaden the understanding of a highly conserved family of transcriptional regulators that play important roles in in metazoan development.

Our analysis indicates that the classification of Sox genes in invertebrates appears to be robust and that genes in all groups have aspects of their expression patterns that suggest evolutionary conservation across the Bilateria.

# Chapter 3

# A SoxB Gene Regulates Segmentation in the Spider

## Introduction

In the previous chapter I showed the expression patterns for the Sox genes in *P. tepidariorum*. One of the SoxB genes, *Sox21b-1*, in particular has interesting expression from the germ disc stage up until late stage 11, which suggests it is involved in the process of segmentation (Figure 3.5), while the expression of others implies they are involved mainly in neurogenesis (Figures 3.4; 3.6; 3.7).

### *Sox21b-1* regulates prosomal and opisthosomal segmentation

To understand the function of the Sox genes during the spider embryogenesis, I designed double-stranded RNA from fragments for all fifteen Sox genes found in the *P. tepidariorum* genome. Four spiders were injected for each fragment (n= 60), and after careful analysis of their embryonic development, I noticed that only the orthologue of *D. melanogaster* Sox21B, namely *Sox21b-1* showed a visible phenotype, so I pursued the analysis of this gene further.

After the injections with one dsRNA fragment against *Sox21b-1*, I observed a phenotype in the late germ disc/early germ band stage in which the cumulus did not migrate properly, and in older embryos there was clearly a truncation of the germband as well as perturbed prosomal segmentation. To analyze that phenotype in more detail, I designed and re-injected two non-overlapping fragments for this gene, one corresponding to the HMG domain and the other corresponding to a region further 3’ (Figure 4.1 - A).
Both fragments were injected in to seven spiders and the phenotypes I observed were consistent between the two non-overlapping *Sox21b-1* fragments I used for the first trial of RNAi injections, as well as the original single dsRNA fragment (Figure 4.1 - B).

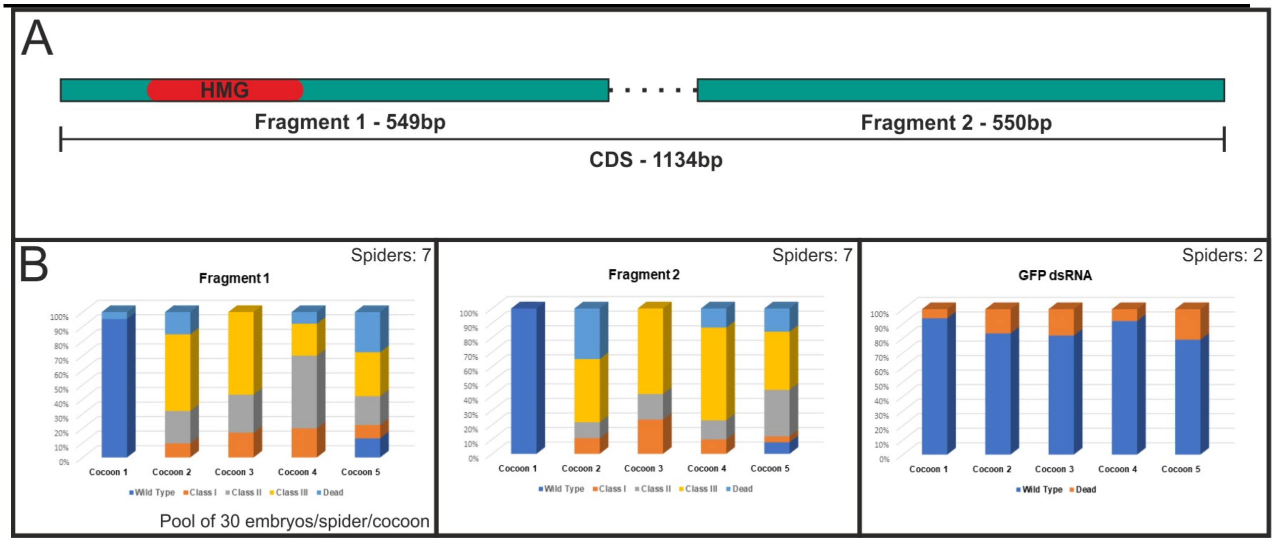


**Figure 4. 1. dsRNA design and phenotypic class frequencies for each *Sox21b-1* fragment and GFP control injections. A**) Two non-overlapping fragments were designed for the *Sox21b1* coding sequence. Fragment 1 corresponds to the HMG conserved domain (549 bp) and fragment 2 has no conserved domains (550 bp). **B**) Frequencies for each fragment, cocoon number and phenotype class. Seven spiders were injected for each Pt-Sox21b-1 fragment and two spiders for the GFP dsRNA controls. For the phenotype class frequencies, 30 embryos per spider per cocoon were pooled, DAPI stained and analysed (total n = 210 for each).

The embryos produced by female spiders injected with dsRNA fragments of *Sox21b-1* exhibited a range of abnormal phenotypes that I subdivided them into three classes. Note that the range and number of phenotypes observed was similar for the two dsRNA fragments and these effects were never observed in cocoons from the dsGFP control injections.

Class I embryos developed a presumptive head region (Figure 4.2 - A-C), as well as normal cheliceral, pedipalpal and first leg bearing (L1) segments in stage 9 (Figure 4.2 - C). However, in Class I embryos the other three leg bearing segments, L2 - L4, as well as all the opisthosomal segments were missing (Figure 4.2 – C). These embryos exhibited a truncated germ band, terminating in disorganised tissue in the region of the SAZ (Figure 4.2 - C). In the case of Class II phenotypes, embryos only differentiated the head region and the cheliceral and pedipalpal segments (Figure 4.2 - D). All leg bearing segments of the prosoma and all opisthosomal segments produced from the SAZ were missing in stage 9 embryos (Figure 4.2 - D). In Class III embryos, the germ band did not form properly from the germ disc at stage 5 (Figure 4.2 - E) and I therefore looked earlier in development to understand how this severe phenotype arose and could affect the posterior stages of development.

I observed that the formation of the primary thickening occurs normally at stage 4 (Akiyama-Oda & Oda, 2006; Pechmann *et al*., 2017). Subsequently the cumulus, the group of mesenchymal cells that arises as the primary thickening at the centre of the germ disc, fails to migrate properly to the rim of the germ disc during stage 5 in Class III phenotype embryos (Figure 4.2 - D). Since migration of the cumulus is required for the transition from germ disc to germ band, this observation, at least in part, explains the subsequent Class III phenotype. In some embryos, I did observe the opening of the dorsal field in stage 6 embryos: therefore, I suggest these embryos later develop Class I and II phenotypes (Figure 4.2 - B-C).

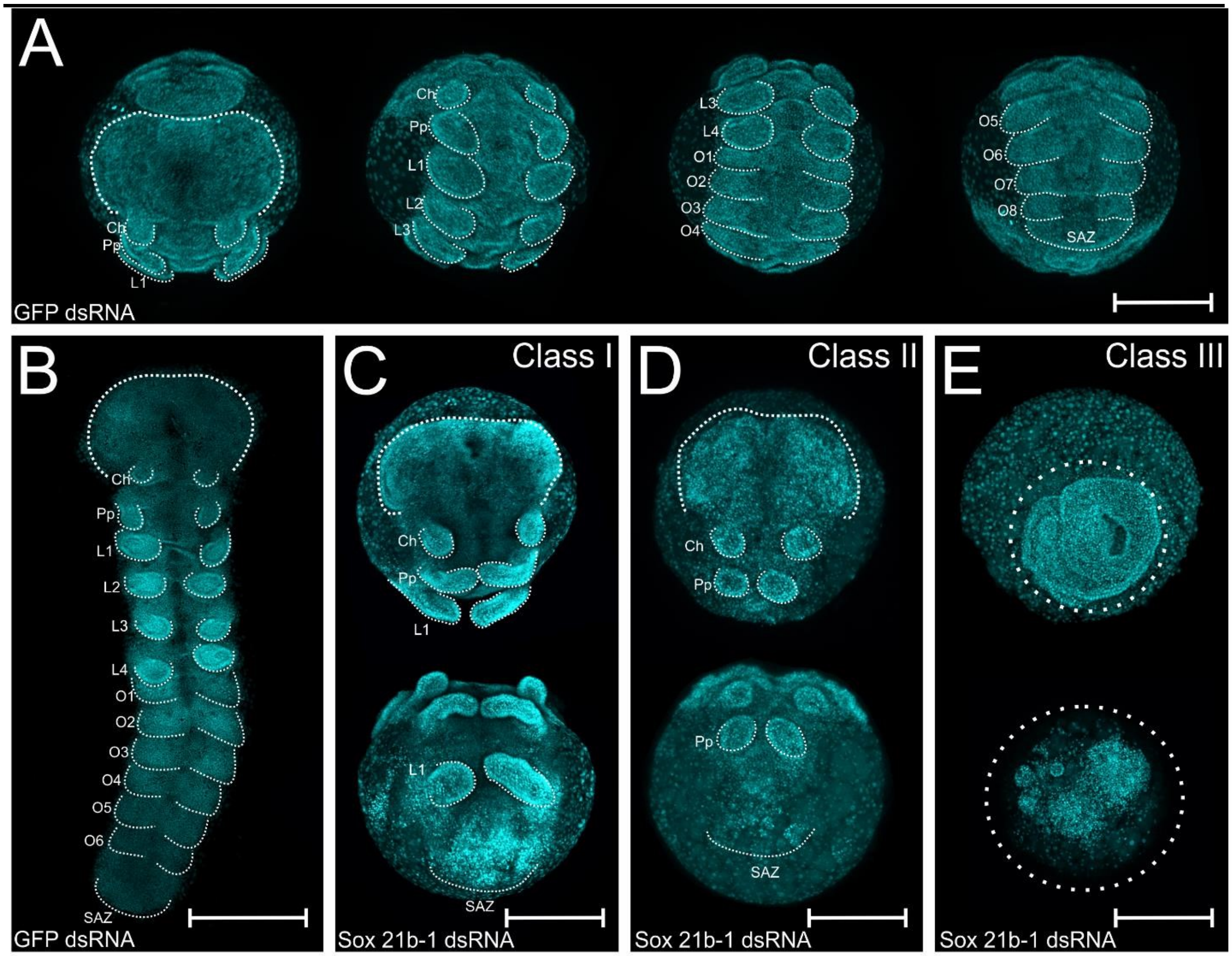


**Figure 4. 2. Embryo phenotypes after *Sox21b-1* parental RNAi knockdown.** Whole mount (**A**) and flat mount (**B**) control embryos at stage 9 stained with DAPI. Stage 9, Class I (**C**), Class II (**D**) and Class III (**E**) phenotypes from *Sox21b-1* knockdown. In the control embryos (A and B), the head, cheliceral (Ch), pedipalpal (Pp), prosomal walking limbs (L1 to L4), opisthosomal segments (O1 to O6) and a posterior SAZ are all clearly visible as indicated. **C**) Class I phenotype embryos show a morphologically normal head, pairs of chelicerae, pedipalps and first walking limbs (Ch, Pp, L1), but a disorganised cluster of cells in the posterior where L2-L4, opisthosomal segments and the SAZ should be. **D**) Class II phenotype embryos consist of fewer cells, but still form a head, chelicerae, pedipalps (Ch, Pp) and a structure resembling the SAZ in the posterior. **E**) Class III embryos exhibit the most severe phenotype, where, after the germ disc stage, the embryo fails to form an organised germ band. Anterior is to the top, scale bars: 150 μm.

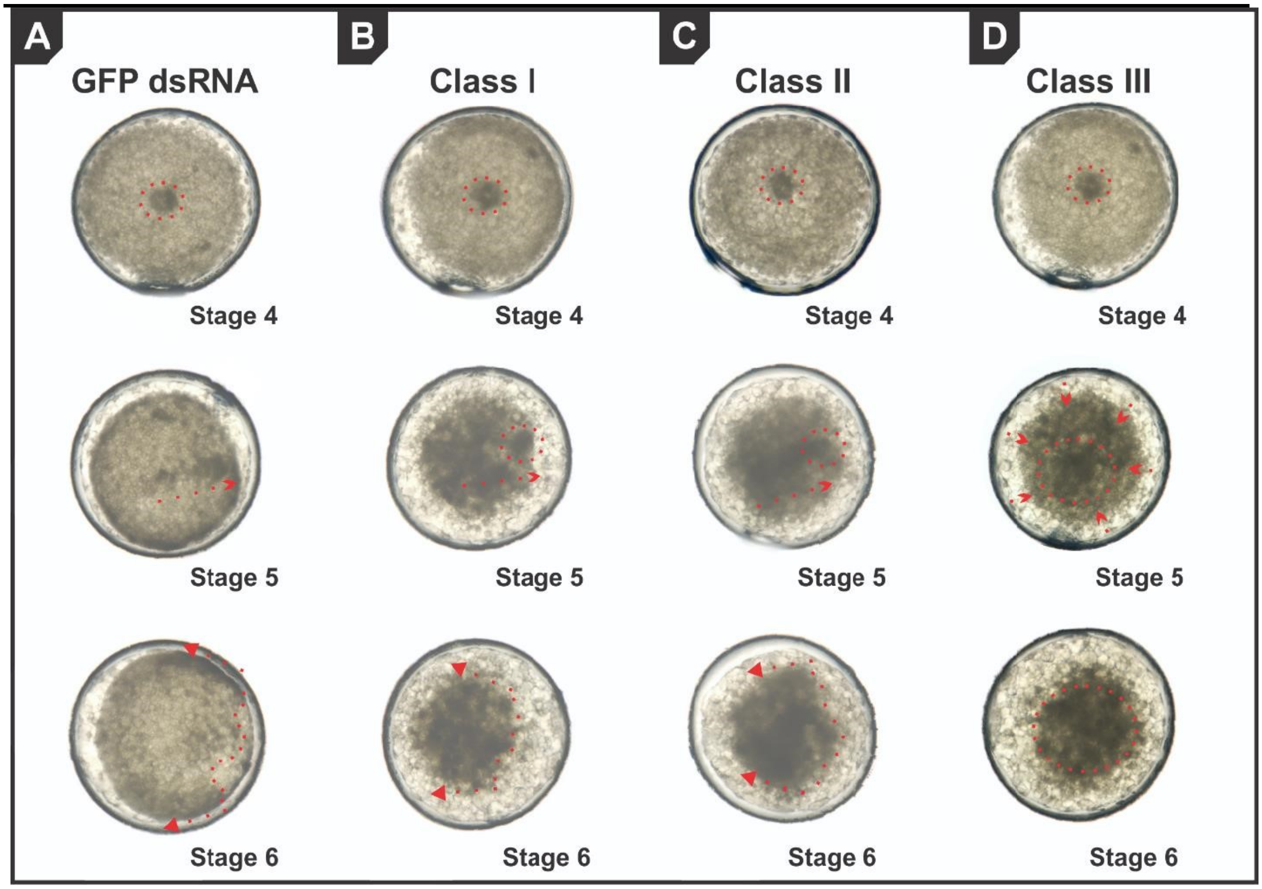


**Figure 4. 3. Snapshots from live imaging of control and *Sox21b-1* knockdown embryos.** Ventral view of the germ disc in **A**) GFP dsRNA control embryos, showing the cumulus formation (red dotted lines), cumulus migration (red dotted arrow) and dorsal field opening (red dotted line and arrows). **B**) Class I *Sox21b-1* knockdown embryos showing cumulus formation, the partial migration of mesenchymal cells and limited dorsal field opening, which is also seen but more severely disrupted in Class II embryos (**C**), and absent in class III (**D**). Anterior is to the left, opposite to the direction of cumulus migration.

The identity of the segments that differentiated in the knockdown embryos was confirmed by expression of *labial* (*lab*) in the pedipalps and L1 and *Deformed-A* (*Dfd*-A) in L1 (Figure 4.4 – A and B).

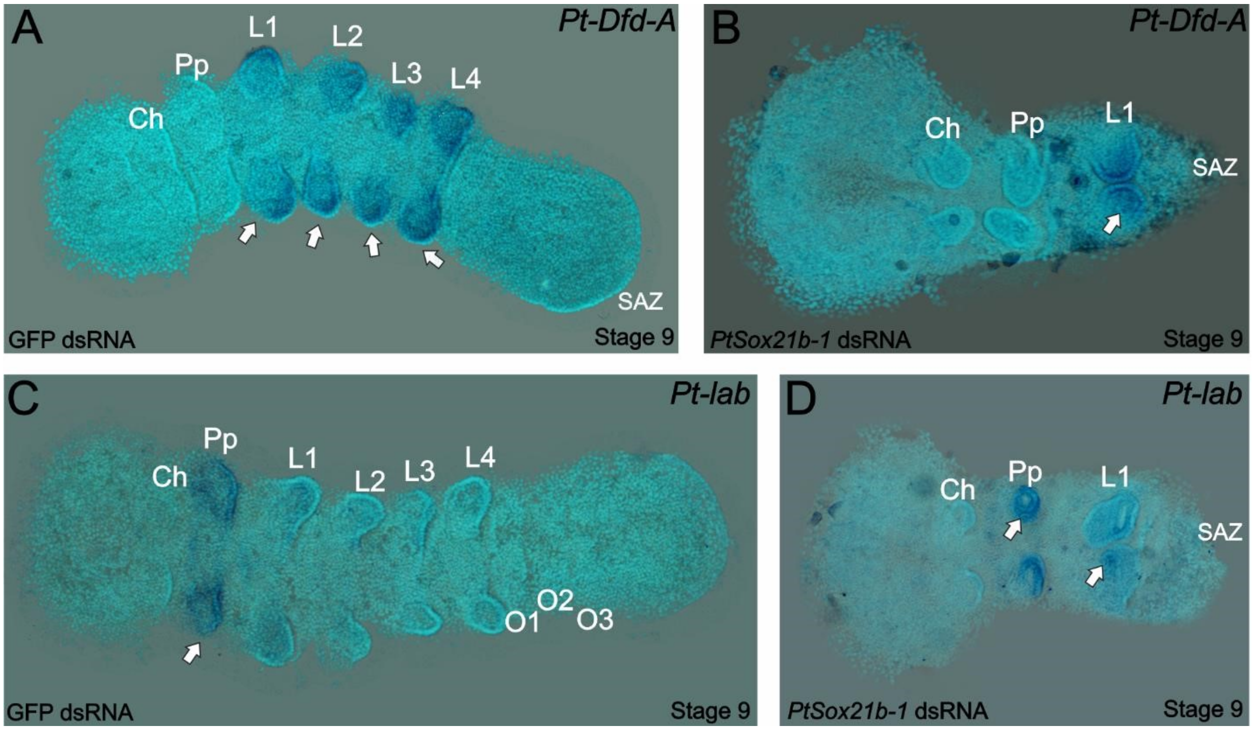


**Figure 4. 4. Homeotic gene expression at stage 9 in *Sox21b-1* pRNAi embryos. A)** Ventral view showing *Pt-Dfd-A* expression in the limb buds of L1 to L4 segments in the control embryos (white arrows). **B**) Expression of *Pt-Dfd-A* is also observed in L1 in *Sox21b-1* pRNAi embryos (n = 9) (white arrow in **B**). **C**) *Pt-lab* is expressed in the pedipalpal segment and faintly in L1 segment in control embryos (white arrow in **D**). In *Sox21b-1* pRNAi embryos, *Pt-lab* expression can still be observed in the pedipalpal and L1 segments (n = 10) (white arrows in **D**). DAPI stained nuclei are shown in cyan and the membrane marker alpha-Tubulin in red. Anterior is to the left in all panels.

## *Sox21B-1* is involved in cell maintenance during spider embryogenesis

I next examined the effect of *Sox21b-1* depletion on cell death and proliferation at stages 5 and 9 in knockdown and control embryos using antibodies against Caspase-3 and phosphorylated Histone 3 (PHH3) (Figure 4.5). At the germ disc stage there is no detectable cell death in control embryos (n = 10), but I observed some small clusters of apoptotic cells in the *Sox21b-1* knockdown embryos (n = 10) (Figure 4.5 – A and B). At stage 9, a few cells expressed Caspase-3 in the posterior-most part of the SAZ (Figure 4.5 - C), but I did not observe cell death in this region of *Sox21b-1* knockdown embryos (Figure 4.5 - D). However, I did observe pronounced cell death in the head extraembryonic layer of the same embryos (n = 10) (Figure 4.5 - D).

Expression of PHH3 at stages 5 and 9, indicated that *Sox21b-1* knockdown embryos show decreased cell proliferation compared to controls (n = 10 for each) (Figure 4.5 - E-H). Interestingly the cells were also clearly larger in *Sox21b-1* knockdown embryos compared to controls, which may reflect perturbed cell proliferation (Figure 4.5 - E-H). Thus, my functional analysis shows that *Sox21b-1* regulates cell proliferation and the transition from radial to axial symmetry. Moreover, *Sox21b-1* is involved in two different segmentation mechanisms in spiders: it seems to be required for the correct differentiation of the head and trunk segments, but also for the formation of the SAZ and subsequent production of opisthosomal segments.

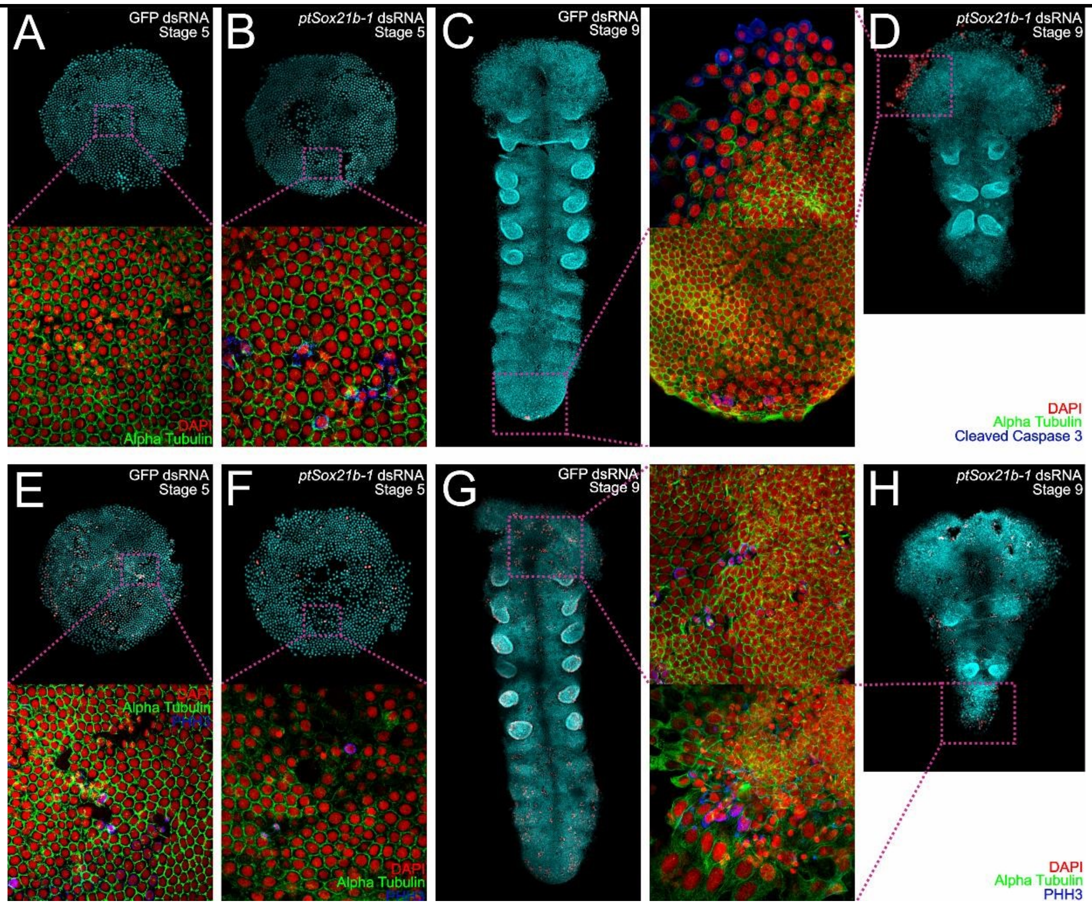


**Figure 4. 5. Cell death and cell proliferation in *Sox21b-1* knockdown embryos.** Ventral view of stage 5 control embryos stained for Cleaved-Caspase3 (**A**) and PHH3 (**E**). Cell death is not detectable in control embryos, but a high level of proliferation can be seen. In *Sox21b-1* knockdown embryos, clusters of cells undergoing cell death can be found (**B**), as well as a decrease in proliferation in the knockdown embryos compared to controls (n = 15 for each staining) (**F**). Embryos at stage 9 stained for Cleaved-Caspase 3 (**C**) and PHH3 (**G**) show that only a small amount of cell death occurs in the SAZ, and that there is proliferation detectable throughout the entire embryo. Cell death is visible in the head extraembryonic layer in Sox21b-1 pRNAi embryos (**D**), and less proliferation is detected in stage 9 knockdown embryos (n = 15 for each staining). Anterior is to the top. Magnifications are 100X and 400x respectively.

## Effects of *Sox21b-1* knockdown on the germ disc and mesoderm formation

In *P. tepidariorum, decapentaplegic* (*dpp*) and *Ets4* are required for cumulus formation (Akiyama-Oda & Oda, 2006; Pechmann *et al*., 2017). To investigate if *Sox21b-1* is involved in the formation of this cluster of cells I assayed the expression of *dpp* and *Ets4* in *Sox21b-1* RNAi knockdown embryos. However, I observed that both genes were expressed normally, and cumulus formation was unaffected in the embryos analysed. Note that in this experiment the embryos for the three classes were pooled (Figure 4.6 – E and F).

The rim of the spider germ disc develops into the head structures and is regulated in part by *hh*, while the mesodermal and endodermal layers of the head are specified by the mesendodermal gene *forkhead* (*fkh*) (Akiyama-Oda & Oda, 2006; Feitosa, 2017). To investigate if anterior expression of *Sox21b-1* is involved in the formation of the head rudiment and differentiation of the mesodermal and endodermal layers, I assayed the expression of *fkh* and *hh* in class I and II *Sox21b-1* knockdown embryos.

*hh* is expressed in the ectoderm at the rim of the germ disc (Figure 4.6 - D) (Kanayama, 2011) and remains unaffected by *Sox21b-1* knockdown (Figure 4.6 - H). *fkh* is also expressed in cells around the rim, as well as in the centre of the germ disc in mesendodermal cells (Figure 4.6 - C). In *Sox21b-1* knockdown embryos both *fkh* expression domains are lost (Figure 4.6 - G) and it therefore appears that *Sox21b-1* is required for the specification of mesodermal cells in the germ disc of spider embryos.

In both spiders and flies, *twist* (*twi*) is involved in mesoderm specification (Yamasaki, 2005) and I therefore examined the expression of this gene after *Sox21b-1* knockdown. To investigate if the loss of *fkh* in stage 5 embryos leads to defects in the formation of the secondary internal layers, I evaluated the late expression of *twi* in both control and stage 9 *Sox21b-1* knockdown embryos.

*twi* is expressed in the visceral mesoderm of the limb buds from L1 to L4, in the opisthosomal segments O1 to O4, and in an anterior mesodermal patch in the central part of the developing head in control embryos (Figure 4.6 - I) (Yamazaki, 2005). While the head expression persists in *Sox21b-1* class I embryos, strikingly, expression in all the limb and opisthosomal segments appears to be lower or absent (Figure 4.6 - J).

This apparent loss of *fkh* expression in stage 5 could be explained by a reduction of the endomesodermal layers in the older stages, but *twi* expression is still normal in the head region. Therefore, I analysed orthogonal projections of control and *Sox21b-1* knockdown embryos to evaluate if in the cellular level there was a decrease in the layer formation. In the germ disc of stage 5, when *fkh* expression commences, I observed invaginating cells forming a second layer (Figure 4.6 - K), however, in *Sox21b-1* knockdown embryos I found a lower number of invaginating cells, which exhibit bigger nuclei compared to the controls (Figure 4.6 - L).

A similar effect was observed at stage 9. In the anterior-most region of the embryo three layers of cells can be identified in control embryos (Figure 4.6 - M), but in *Sox21b1* knockdown embryos the formation of these layers is perturbed, and number of cells seems to be reduced in secondary layers (Figure 4.6 - N). These data suggest that the segmentation in the prosomal region is ectodermal and occurs even upon a reduction of the internal most layers of the embryo. This maintenance of the ectodermal layer by *PtSox21b-1* could be another important role for this gene in the segment formation of the spider.

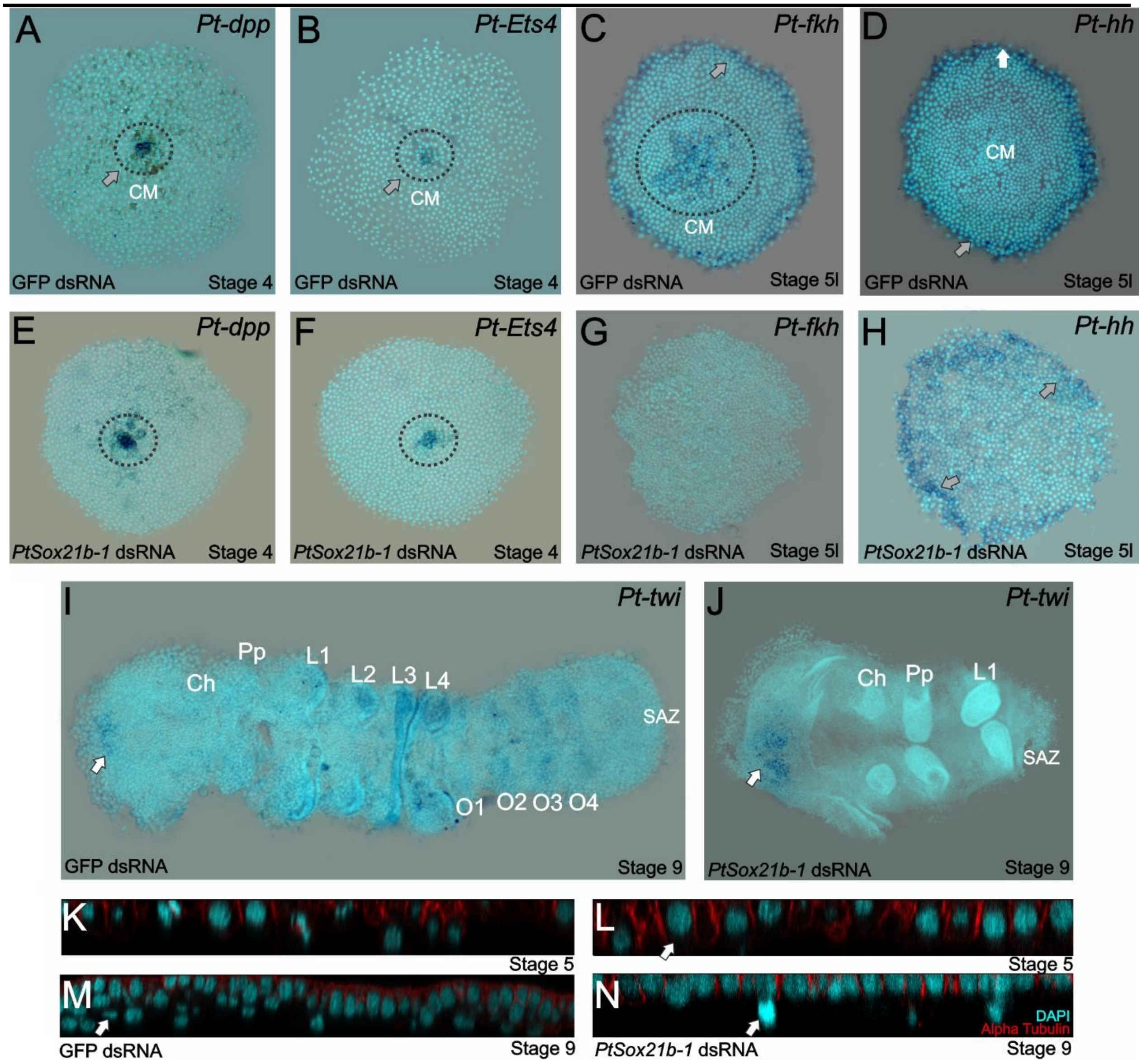


**Figure 4. 6. Gene expression in control and *Sox21b-1* knockdowns at the germ disc stage and during mesoderm formation.** *Pt-dpp* (A) and *Pt-Ets4* (B) are expressed in the forming cumulus (CM) in the centre of the germ disc at stage 4 (grey arrow and dotted circle). This expression is unaffected by knockdown of *Sox21b-1* (**E** and **F**) (n= 30 for each gene). **C**) *Pt-fkh* is expressed at the rim and centre of the germ disc at late stage 5 (grey arrow and dotted circle in C), but expression is lost in *Sox21b-1* embryos (n= 30) (**G**). *Pt-hh* expression at the rim of the germ disc (**D**) is normal in *Sox21b-1* knockdown embryos (**H**) (grey arrows). **I**) The mesodermal marker *Pt-twi* is expressed in the anterior-most medial region of the head, limb buds of L1 to L4, and with a striped pattern in the O1 to O4 segments. **J**) In Sox21b-1 knockdown embryos, only the head expression is maintained (n= 14) (white arrow in **J**). **K-L** show orthogonal projections of the cumulus (stage 5) and the head (stage 9) at 40x magnification of whole mount control embryos (left panels) and Sox21b-1 knockdown embryos (right panels), respectively. In control embryos the formation of subectodermal layers are visible, which are lost in the knockdown embryos. Ch: Chelicerae; HL: Head Lobes; L1 to L4: Prosomal leg bearing segments; O1 to O5: Opisthosomal segments; SAZ: Segment Addition Zone. Anterior is to the left in stage 9 embryos.

## Effects of *Sox21b-1* knockdown on segmentation

In *P. tepidariorum*, formation of the SAZ and production of posterior segments requires the Wnt8 and Delta-Notch signalling pathways (McGregor, 2008; Schönauer *et al*., 2016). Interactions between these pathways regulate *hairy* (*h*) and, via *caudal (cad*)*,* the expression of pair-rule gene orthologues including *even skipped* (*eve*) (McGregor, 2008; Schönauer *et al*., 2016). To better understand the loss of opisthosomal segments I observed in *Sox21b-1* knockdown embryos, I analysed the expression of *Dl, Wnt8, h* and *cad* in these embryos compared to controls.

*Dl* is expressed at stages 5 and 6 in the forming SAZ, in the region of the L4 primordia, and in the presumptive head (Figure 4.7 - A). Subsequently at stage 9, *Dl* expression is visible in clusters of differentiating neuronal cells and oscillates in the SAZ - an expression pattern associated with the sequential addition of new segments (Figure 4.7 - B) (Oda *et al.*, 2007). In *Sox21b-1* knockdown embryos, *Dl* expression is not detected at stage 5 (Figure 4.7 - C) and is absent in the posterior at stage 9 (Figure 4.7 - D). However, expression in the anterior neuroectoderm appears to be normal up to the pedipalpal segment, although neurogenesis is apparently perturbed in the presumptive L1 segment (Figure 4.7 - D) (Oda *et al*., 2007). This suggests that the ectoderm up to the L1 segment differentiates normally, but the formation of the SAZ and posterior segment addition controlled by *Dl* is lost upon *Sox21b-1* knockdown.

As mentioned in results chapter 1, *Pt-h* seems to be involved in the regulation of segment addition in *P. tepidariorum* (McGregor *et al.,* 2008A; Pechmann *et al.*, 2009). At stage 9, *h* expression resembles *Dl* in the ventral neuroectoderm up to the opisthosomal segment 2 (O2), as well as in the SAZ (Figure 4.7 - G). Similar to what is seen to *Dl* expression in late stage 5 *Sox21b-1* knockdown embryos, the expression of *Pt-h* is also lost throughout the entire germ disc (Figure 4.7 - F). In addition, in Class I phenotype embryos at stage 9, the expression of *h* is completely absent in the tissue posterior to the pedipalpal segment (Figure 4.7 - H). *Pt-h* is a downstream regulator of Notch signaling, thus the similarity with the *Pt-Dl* expression in *So21b-1* knockdown embryos might have been expected.

*Wnt8* is initially expressed at stage 5 in the centre and at the rim of the germ disc (Figure 4.7 - I). At stage 9, striped expression of *Wnt8* is seen from the head to the posterior segments and in the posterior cells of the SAZ (Figure 4.7 - K). Knockdown of *Sox21b-1* results in the loss of *Wnt8* expression in late stage 5 embryos (Figure 4.7 - J). At stage 9, *Wnt8* expression is observed in the cheliceral, pedipalpal and first walking limb segments of *Sox21b-1* knockdown embryos, but no expression is detected in the remaining posterior cells (Figure 4.7 - L). Consistent with the loss of *Dl* and *Wnt8*, *cad* expression is also lost in stage 5 and stage 9 *Sox21b-1* knockdown embryos (Figure 4.7 - O-P).

To look at the effect of *Sox21b-1* knockdown on segmentation in more detail I examined the expression of *engrailed* (*en*) and *hh*. At stage 9, *en* is expressed segmentally from the cheliceral to the O3 segment in wild type embryos (Figure 4.7 - Q). However, in *Sox21b-1* knockdown embryos, expression of *en* was only observed in the cheliceral, pedipalpal and L1 segments, consistent with the loss of all the more posterior segments (Figure 4.7 - R). *hh* has a similar expression pattern to *en* at stage 9, except it exhibits an anterior splitting wave in the cheliceral segment and is also expressed earlier in opisthosomal segments and in the SAZ (Figure 4.7 - S). Upon *Sox21b-1* knockdown, *hh* is only detected in shortened stripes in the cheliceral and pedipalpal segments (Figure 4.7 - T).

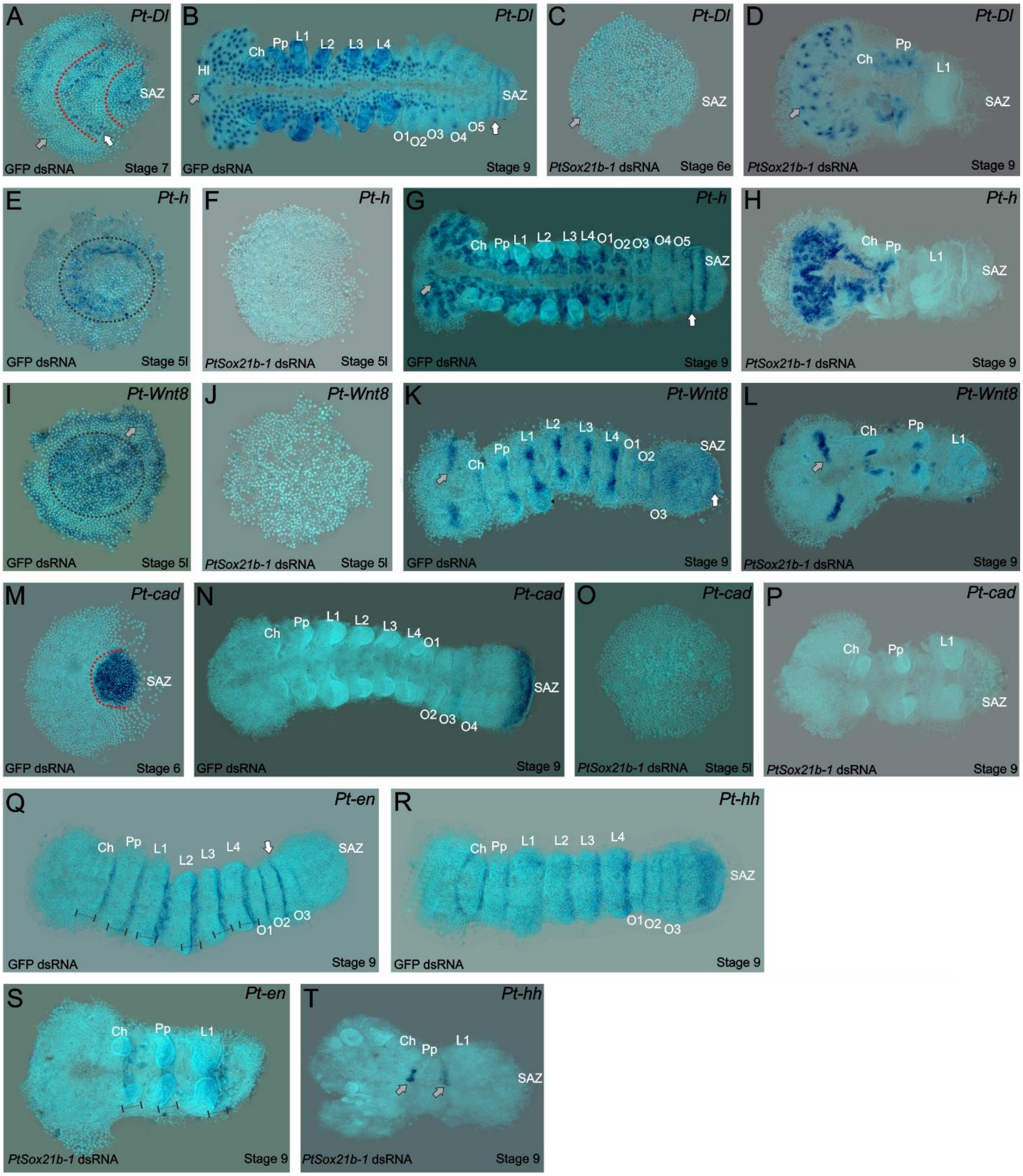


**Figure 4. 7. Expression of segmentation genes in *Sox21b-1* pRNAi embryos. A** and **B**) *Pt-Dl* expression at late stage 6/early stage 7 appears to be dynamic in the SAZ and is also observed in the presumptive head region and prosoma of the embryo (red dotted lines and grey arrows).

**B**) At stage 9, *Pt-Dl* expression is seen in the SAZ (white arrow) but is restricted to the clusters of proneural differentiation in the anterior region of the embryo (grey arrow in the head lobes). **C**) In *Sox21b-1* knockdown embryos, *Pt-Dl* expression is not detectable in late stage 5/early stage 6 embryos (grey arrow) but can still be observed in the anterior ventral neuroectoderm at stage 9 up to the pedipalpal segment (n = 17 and n = 14 for stage 5 and 9, respectively) (**D**). *Pt-h* expression at stage 5 in control embryos is seen at the rim and in the centre of the germ disc (black dotted circle in **E**), which is lost in *Sox21b-1* knockdown embryos (**F**). At stage 9, *Pt-h* expression resembles *Pt-Dl*, both in the control and *Sox21b-1* knockdown embryos (**G** and **H**) (n = 15 for both stages). *Pt-Wnt8* expression is similar to *Pt-h* in stage 5 control embryos (black dotted circle in the centre, grey arrow to the rim) and is also lost in *Sox21b-1* knockdown embryos (n = 11) (**I** and **J**). Control embryos at stage 9 show the expression of *Pt-Wnt8* in the medial region of the head (grey arrow), and in distal parts of each segment up to the SAZ (white arrow) (**K**). In *Sox21b-1* knockdown embryos at the same stage, the brain (grey arrow), cheliceral and pedipalpal expression is still present, but the posterior expression is lost (n = 17 for each stage) (**L**). *Pt-cad* is expressed in the SAZ at late stage 5/early stage 6 embryos (**M**), which persists throughout to stage 9 control embryos (**N**). However, *Pt-cad* expression is lost upon Sox21b-1 knockdown (n= 20 for each stage) (**O** and **P**). *Pt-en* expression is present in the posterior of each segment (black lines in **Q**), and in cheliceral, pedipalpal and L1 segments in *Sox21b-1* knockdown embryos at stage 9 (n = 10) (**S**). *Pt-hh* expression in control embryos at stage 9 is seen in the posterior of each segment and in the SAZ (**R**). When *Sox21b-1* is knocked-down, *Pt-hh* embryos show expression in the middle posterior of the cheliceral and pedipalpal segments (n = 8) (**T**). Ch: Chelicerae; HL: Head Lobes; L1 to L4: Prosomal leg bearing segments; O1 to O5: Opisthosomal segments; SAZ: Segment Addition Zone. Anterior is to the left in stage 9 embryos.

These observations indicate that *Sox21b-1* is required for the formation of the prosomal L1 – L4 segments and acts upstream of Wnt8 and Delta-Notch signalling to regulate the formation of the SAZ and the subsequent production of posterior segments. In support of this regulatory hierarchy, I observed that *Sox21b-1* expression is still detected in the posterior regions of the truncated embryos produced by RNAi knockdown of either *Dl* or *Wnt8* (Figure 4.8).

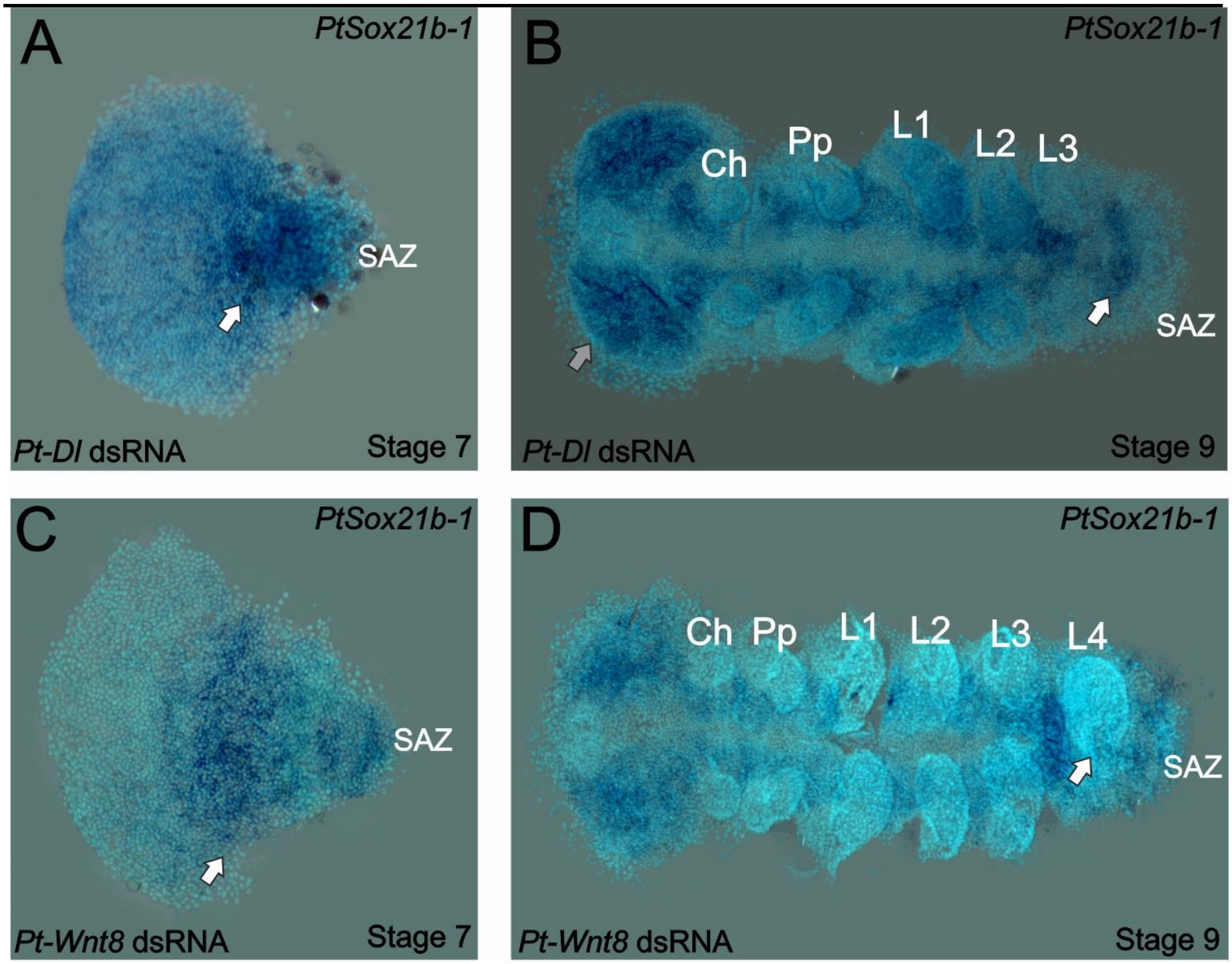


**Figure 4. 8. Expression of *Sox21b-1* in *Dl* and *Wnt8* pRNAi embryos.** Ventral view of stage 7 and 9 knockdown embryos for *Pt-Dl* (**A** and **B**) and *Pt-Wnt8* (**C** and **D**). In knockdown embryos for both *Pt-Dl* and *Pt-Wnt8*, *Sox21b-1* is still expressed at the germ band stage (**A** and **C**), in a dynamic pattern in the remaining SAZ cells, and in the forming segments in the presumptive prosoma of the embryo (white arrows). In stage 9 *Pt-Wnt8* knockdown embryos, *Sox21b-1* remains highly expressed in the ventral nerve cord (**D**). *Pt-Dl* knockdown embryos lack the posterior L4 segment (white arrow), but brain formation appears normal (grey arrow) (**B**). *Pt-Wnt8* embryos show a fusion of the L4 limb buds, and *Sox21b-1* is still expressed in the remaining SAZ cells (**D**). Anterior is to the left in all panels.

Taken together, my analysis of *P. tepidariorum Sox21b-1* parental RNAi depleted embryos reveals an important role for this Group B Sox gene in both the gap-like segmentation of the prosoma, as well as posterior segment formation from the SAZ.

These experiments further emphasise the critical role this class of transcription factors play in arthropod segmentation.

The results presented in this chapter are included in a manuscript entitled "A SoxB gene acts as an anterior gap gene and regulates posterior segment addition in the spider *P. tepidariorum*", that was submitted in April 2018.

## Discussion

## A SoxB gene is required for two different mechanisms of spider segmentation

The Sox (Sry-Related High-Mobility Group box) gene family encodes transcription factors that regulate many important processes underlying embryonic development of metazoans (Sinclair, 1990; Lefevbre, 2010). One such gene, *D,* is expressed in a gap gene-like pattern and is involved in regulating the segmentation cascade in *D. melanogaster* (Russell, 1996; Clark & Peel, 2018). Recently, the analysis of the expression of *D* in the flour beetle *T. castaneum* strongly suggests it also plays a role in short germ segmentation, further supported by knockdown of the *D* orthologue in *Bombyx mori,* which resulted in the loss of posterior segmentation (Nakao, 2018).

I injected dsRNA specifically for each orthologue out of the 15 Sox genes in *P. tepidariorum*, but the only gene that resulted in a phenotype was *Pt-Sox21b-1*. One cannot be sure of the exact reasons why only one out of fifteen knockdowns presented a phenotype – all the primary dsRNAs were designed outside the HMG domain, and so it is possible the redundancy of the Sox genes (Reiprich, 2015) may underlie these apparent lacks of effects in the knockdown assays. In *D. melanogaster*, knockdown of *D* and *SoxN* were only achieved when a combination of targets for these two Sox genes were mixed before the injections (Steven Russell, personal communication).

Here I show that, while *D* is very likely not to be involved in spider segmentation (Chapter 2 of this thesis - Paese *et al.,* 2018), the closely related SoxB gene, *Sox21b-1*, regulates formation of both prosomal and opisthosomal segments. In the prosoma *Sox21b-1* has a gap gene like role and is required for the specification of L1-L4 segments (Figure 3.5), resembling the roles of *hb* and *Dll* in prosomal segmentation in this spider

(Schwager, 2009; Pechmann, 2011) and, at least superficially, gap gene function in *Drosophila*.

In *Drosophila* the gap genes regulate pair-rule gene expression, and while the orthologues of *eve* and *runt* are not expressed in the prosoma of spiders (Damen, 2000; Schönauer *et al.*, 2016), our results indicate that *Sox21b-1* is required for the expression of *h* and the generation of leg bearing prosomal segments (Figure 4.7 - E).

In the posterior, *Sox21b-1* knockdown perturbs SAZ formation and consequently results in truncated embryos, lacking all opisthosomal segments. Therefore, *Sox21b-1* regulates development of the SAZ, and our observations indicate this is at least in part through roles in organising the germ layers and specification of mesoendodermal cells during stages 5 and 6. This is supported by the loss of *fkh* expression upon *Sox21b-1* knockdown, which is required for mesoderm and endoderm formation in both spiders and insects (Holmberg, 2008; Kormish, 2010). Moreover, the subsequent dynamic expression of *Sox21b-1* in the SAZ after stage 6 is suggestive of a role in segment addition.

These findings about *Sox21b-1* provide an important new insight into the gene regulatory network (GRN) underlying the formation of the SAZ and the sequential addition of segments from this tissue. I have shown that *Sox21b-1* acts upstream of Wnt8 and Delta-Notch signalling in this GRN and is required for the activation of these important signalling pathways during posterior development (Figure 4.9). Further work is needed to determine if Group B Sox genes, such as *D* and *Sox21b-1*, occupy a similar position in the regulatory hierarchy for posterior segmentation in other short-germ arthropods. This could provide important new insights into the evolution of the regulation of segmentation in arthropods since a Wnt-Delta-Notch-Cad regulatory cassette was probably used ancestrally in arthropods to regulate posterior development (Janssen, 2004; McGregor, 2009; Brena, 2012). Interestingly, SoxB genes also cooperate with Wnt and Delta-Notch signalling in various aspects of vertebrate development including the patterning of neural progenitors and maintenance of the stem state in the neuroepithelium (Kormish, 2010; Koch, 2017).

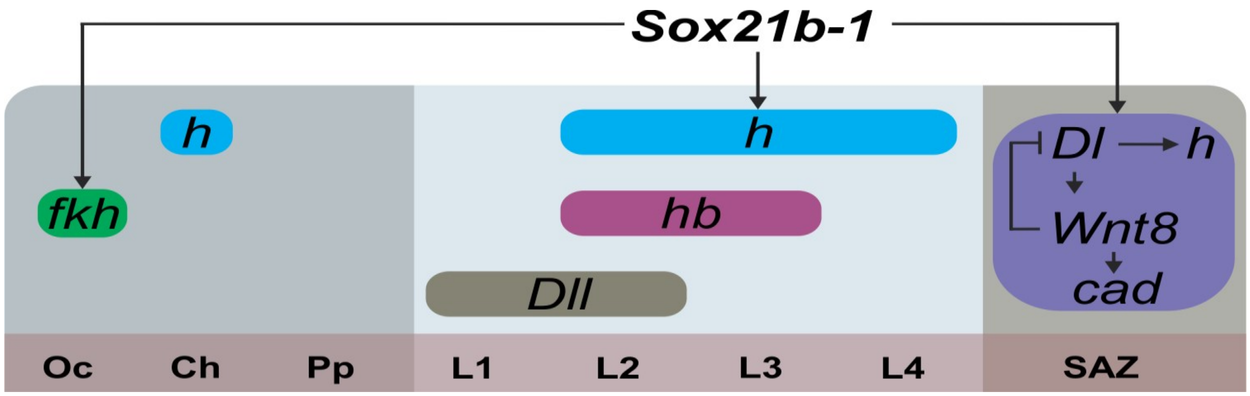


**Figure 4. 9. Summary of the updated Gene Regulatory Network.** The interaction of *Sox21b-1* is presented in relation to genes involved in spider embryogenesis. I found that *fkh* expression requires *Sox21b-1* in the most anterior part of the head (OC, Ch, Pp segments). *Distal-less*, *hunchback* and *hairy*, involved in prosomal limb differentiation (L1-L4) are also positively regulated by *Sox21b-1*. Our results suggest that *Sox 21b-1* controls the expression of *hairy* in segments L2 to L4. The molecular control of segmentation in the SAZ involves a feedback loop between *Dl* and *Wnt8*, which acts upstream of *cad* and also controls the dynamic expression of *hairy*. I can infer from our results that *Sox21b-1* acts upstream of these genes in the SAZ.

### *Sox21b-1* exhibits highly pleiotropic phenotypes during early spider embryogenesis

My results show that *Sox21b-1* is not only involved in segmentation but is maternally supplied and regulates cell division in the early germ disc, as well as probably the transition from radial to axial symmetry during germ band formation. Further experiments with *Sox21b-1* are required to fully elucidate the mechanisms by which it affects these early functions. Furthermore, while spider head development is less affected than trunk segmentation by knockdown of *Sox21b-1*, it is clear from my experiment that *Sox21b-1* regulates cell fate in this region. Interestingly, Sox2 is involved with the neuromesodermal fate choice in mice and *Dichaete* has a role in embryonic brain development in *Drosophila* (Sanchez-Soriano, 1998): consequently, SoxB genes may play an ancestral role in the patterning of the head ectoderm and mesoderm in metazoans (Zhong, 2011).

## The evolution of Sox21b-1

The evolution and diversification of Group B Sox genes in insects is not fully resolved due to difficulties in clearly assigning orthologues based on the highly conserved HMG domain sequence (Wegner, 1999; Overton, 2002). However, despite these ambiguities it is clear that the *Dichaete* and *Sox21b* class genes in all arthropods examined to date are closely related and likely arose from a duplication in the common ancestor of this phylum [see Zhong *et al.* (2011) for discussion]. Note that in insects *Dichaete*, *Sox21a* and *Sox21b* are clustered (Wegner, 1999), however, while *Dichaete* and *Sox21a* are clustered in *P. tepidariorum,* the *Sox21b* paralogs are dispersed in the genome of this spider (Paese *et al.,* 2018).

I believe it is highly significant that two very closely related SoxB genes are involved in segmentation in both the spider *P. tepidariorum* and in insects, indicating/suggesting an ancient role for this subfamily of Sox genes in invertebrates. Given the close similarity between the HMG domains of Sox21b and Dichaete, it is possible that in some lineages the *D* orthologue assumed the segmentation role, whereas in others it was Sox21b. In spiders, *Wnt8* is involved in posterior development while in other arthropods this role is played by *Wnt1/wg* (McGregor *et al.*, 2008A), and therefore the evolution of *Sox21b-1* via WGD may have led to the co-option to different genes and developmental systems drift of the GRN for posterior development.

As mentioned before, the spider contains an additional related SoxB gene, *Sox21b*-2, that possibly arose as part of the whole genome duplication event in the ancestor of arachnopulmonates over 400 million years ago (Schwager *et al.,* 2017). When compared the HMG aminoacidic sequences, Sox21b-1 and Sox21b-2 have a total of 12 mismatches in a 59 aa sequence (Figure 3.2). I therefore suggest that while Sox21b-1 may have been co-opted to acquire the *D* function in *P. tepidariorum*, Sox21b-2 might have no function in the control of segmentation, but perhaps a later function in the maintenance of different processes. Thus, it will be interesting to examine the roles of Sox genes in other chelicerates, including those that did not undergo a genome duplication, to have a broader comparison in the evolution of segmentation in arthropods.

# Chapter 4

# Identification of Putative Regulatory Regions for Spider Manipulation

## Introduction

The ability to achieve mRNA knockdown with the RNA interference technique was a great breakthrough for the studies of gene function in biological systems, and it has been widely applied in developmental biology in a range of organisms (Weiss, 1999; Kim, 2009). In summary, synthesized double-stranded RNA is injected into females or early embryos, which triggers the RNA III endonuclease Dicer, which in turn recognizes the target mRNA and will therefore start a complex of degradation of this transcript (Thumecke, 2018). This technique was used successfully in many studies in the spider with both embryonic and parental delivery (injection in the female opisthosoma) (Schoppmeier, 2001; AkiyamaOda & Oda, 2003; McGregor *et al*., 2008; Pechmann *et al*., 2009; Schwager *et al*., 2009; Schönauer *et al*., 2016).

However, this technique only allows transient knockdown of the mRNA. After a few hours, all the RNAi levels will be lower, and the transcription of genes whose levels were decreased by the Dicer enzyme will now come back to a normal state (Boultros, 2008). This might explain why some genes in *P. tepidariorum* don't show any phenotype when knocked-down with parental RNAi (e.g. pair-rule and segment polarity orthologues) or the effect wears off. Therefore, it is important to establish approaches whereby gene function analysis can also be made with the endogenous targeting of the organisms' DNA directly (Gilles, 2015).

Such targeting was made possible by techniques such as Zinc-finger nucleases (ZFNs), Transcription activator–like effector nucleases (TALENs) and most recently with the CRISPR/Cas RNA-guided nucleases (RGNs) (Kim, 2009). Using pathways intrinsic to the cell, these techniques allow the specific locus to be mutated by either nonhomologous end-joining (NHEJ), which creates small indels at the site of repair, or by the insertion/translocation of a construct, by the homologous-directed recombination (HDR) pathway, that copies a template that was inserted by complementary homology, (Szostak, 1983; Moore & Haber, 1996). These genomic modifications are now used routinely in established model organisms, such as mouse (Capechi, 2005; Burgio, 2018), the roundworm *Caenorhabditis elegans* (Rieckher *et al*., 2009) and the dipteran insect *D. melanogaster* (Ringrose, 2009; Bassett, 2014). Furthermore, these techniques have been adapted to a few other research organisms, such as *Bombyx mori* (Thomas *et al*., 2002), *Aedes aegypti* (Pinkerton *et al*., 2000), Tribolium (Gilles, 2015) and others.

The CRISPR/Cas9 nuclease is a system derived from the bacterium

*Streptococcus pyogenes* (Jinek *et al*., 2012) and consists of both protein and RNA. The Cas9 endonuclease recognizes a 19 to 22 nucleotide-long guide-RNA (gRNA) that contains the protospacer adjacent motif (PAM) at its 5' end (Gilles, 2015) (summarized in Figure 5.1).

COPYRIGHTED IMAGE REMOVED FROM ELECTRONIC VERSION

**Figure 5. 1. Schematic representation of the CRISPR/Cas9 system.** The Cas9 endonuclease recognizes the PAM in the gRNA sequence and creates a protein-RNA complex that cleaves the DNA in a complementary way, creating indels that are both repaired by NHEJ or HDR. Adapted from Jinek *et al*., 2012.

The CRISPR injection mix must consist of the Cas9 (delivered as protein or mRNA), the gRNAs that is designed directed to the region of interest in the host genome, and a tracRNA that contains the *S. pyogenes* sequence that will make the Cas9 complex (Jinek *et al* 2012). The efficiency of genomic modifications that can be achieved by CRISPR/Cas9 depends on the delivery method used and the ease of screening. For example, in mice oocytes injected with the CRISPR-mix and a GFP tracer oocytes can then be selected for the fluorescence and surgically implanted into the ovaries (Burgio, 2018). Similarly, in the crustacean *Parhyale hawaiensis*, the synthetic gRNA is injected with the Cas9 protein into 1 or 2 cell stage blastoderm, enabling germ line transformation (Martin *et al*., 2016). *D. melanogaster* embryos contain a yolk-rich blastoderm that will cellularize after 1h post fertilization. The Cas9 protein, guide RNA and donor DNA are injected in the posterior part of the embryo, and due to the syncytial nature of the early blastoderm, the injection mix efficiently is spread throughout all the blastomeres leading to modification of somatic cells as well as the germline (Bassett *et al*., 2014).

However, it has been shown in *Drosophila*, *T. castaneum* and other arthropod species exhibiting centrolecithal eggs, that the efficiency of the CRISPR/Cas9 technique increases when the Cas9 protein and the gRNAs are transcribed endogenously with the aid of a species-specific promoter (Gilles, 2015; Huang *et al*., 2017).

There are three types of RNA polymerase (RNApol) classified by which products they transcribe (Figure 5.2). In the following I focus on RNApols II and III. The main difference between these two enzymes is that RNA polymerase II catalyses the DNAdirected transcription of mRNA, which will then be translated into protein (Cramer, 2004), whereas RNA polymerase III transcribes DNA that encodes for small-RNAs, such as small-hairpin RNAs, microRNAs and gRNAs (Figure 5.2) (Ma *et* al., 2014; Huang *et al*., 2017).

Examples of promoters used for transcription with RNApol II are usually genes that are expressed ubiquitously, such as the beta-actin (Ng, 1989), heat-shock proteins (hsp) (Pavlopoulos *et al*., 2009) and tubulins (Gloster *et al*., 1994). The functionality of these promoters was confirmed by the expression of endogenous GFP and mCherry in the embryonic cells (Gilles, 2015). For the transcription of small RNAs with the RNApol III enzyme, promoters such as H1 and U6 are the most commonly used (Ma *et* al., 2014; Gilles, 2015), and the functionality of these is evaluated by the level of genomic modification efficiency.

COPYRIGHTED IMAGE REMOVED FROM ELECTRONIC VERSION

**Figure 5. 2. Table showing the differences amongst the types of RNA polymerases.** In eukaryotes, three types of RNA polymerases are responsible for the transcription of different types of RNA. Briefly, the nucleolar RNApol I transcribe ribosomal RNAs, whereas the nucleuslocalized RNApol II and III transcribes both pre-mRNA and small RNAs, respectively (adapted from Ma *et al*., 2014).

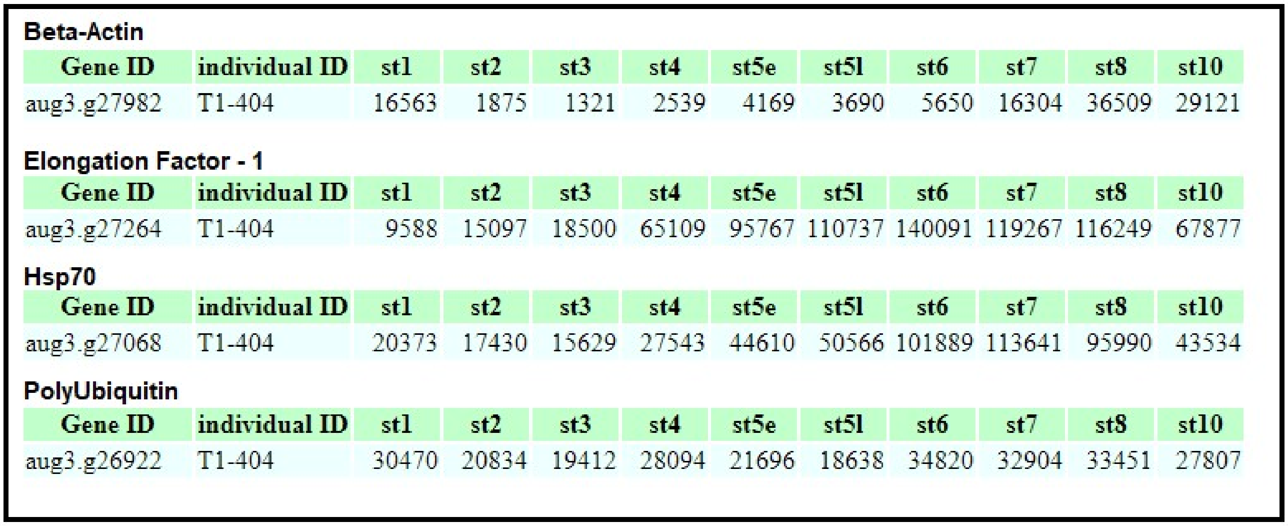
## Establishing tools for genome editing and other applications in *P. tepidariorum*

Injection of synthetic gRNAs together with Cas9 mRNA or protein respectively has been tried in our laboratory previously but was unsuccessful (Schönauer, unpublished data). Therefore, with the aim of producing the Cas9 and guide RNAs at high levels endogenously in spider embryos, I will outline the results of the identification and isolation of putative regulatory regions of various genes in this chapter. Identification of promoter regions is laborious work (Pavlopoulos, 2009) and requires transcriptomic and genomic resources, which were fortunately already available for *P. tepidariorum* (Posnien *et* al., 2014; Schwager *et al*., 2017).

## Characterization and isolation of RNApol II-associated promoters

I started the identification of promoters that could be used to drive gene expression in spider embryos by targeting known ubiquitously expressed genes that have been used successfully to drive gene expression in other organsisms. For this, I carried out a tBLASTn search in the *P. tepidariorum* genome browser for *D. melanogaster* orthologues of the following genes: *beta-actin* (NP_727048), *elongation-factor-1* (NP_477375), *heatshock protein 70* (NP_731651) and *poly-ubiquitin* (AEW12073). Due to the ancestral genome duplication in this spider (Schwager *et al.,* 2017), some of the orthologues have three or more copies of each gene. Therefore, I only followed-up the genes that exbited an expression read-count higher than 5.000 for the embryonic stages 1 to 5 (Figure 5.3). This data was obtained from the publically available *P. tepidariorum* transcriptome (Iwasaki-Yokosama, 2018). This allowed me to identify highly expressed orthologues of *beta-actin*, *elongation-factor-1*, *heat-shock protein 70* and *poly-ubiquitin* in the spider genome, which I present individually below.

Note that I also identified a 16 kb 5' untranslated region (UTR) for *vasa* that potentially contains promoters for this gene (aug3.g5162), but I experienced technical difficulties with in the isolation of fragments over 10 kb by PCR. For this reason, I focused on the promoters of the other genes, but future analysis will need to be made in order to characterize this regulatory region of this gene because it has great potential for germ line transformations (Krøvel, 2002; Tilgner *et al*., 2010).



**Beta-Actin**

| Gene ID | individual ID | st1 | st2 | st3 | st4 | st5e | st5l | st6 | st7 | st8 | st10 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| aug3.g27982 | T1-404 | 16563 | 1875 | 1321 | 2539 | 4169 | 3690 | 5650 | 16304 | 36509 | 29121 |

**Elongation Factor - 1**

| Gene ID | individual ID | st1 | st2 | st3 | st4 | st5e | st5l | st6 | st7 | st8 | st10 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| aug3.g27264 | T1-404 | 9588 | 15097 | 18500 | 65109 | 95767 | 110737 | 140091 | 119267 | 116249 | 67877 |

**Hsp70**

| Gene ID | individual ID | st1 | st2 | st3 | st4 | st5e | st5l | st6 | st7 | st8 | st10 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| aug3.g27068 | T1-404 | 20373 | 17430 | 15629 | 27543 | 44610 | 50566 | 101889 | 113641 | 95990 | 43534 |

**PolyUbiquitin**

| Gene ID | individual ID | st1 | st2 | st3 | st4 | st5e | st5l | st6 | st7 | st8 | st10 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| aug3.g26922 | T1-404 | 30470 | 20834 | 19412 | 28094 | 21696 | 18638 | 34820 | 32904 | 33451 | 27807 |

**Figure 5. 3. Table with the genes ID and read counts for embryonic stages 1 to 10.** Transcriptomic read counts for *beta-Actin*, *ef-1*, *hsp70* and *polyUbiquitin*. The read-counts are per million. The data was collected from the BRH Data Resources.

### *Beta-actin*

The scaffold containing the *beta-actin* gene (aug3.g27982, scaffold 247) annotated in *P. tepidariorum* has 2818 nucleotides upstream of the transcription start codon, in which 183 bp of these was annotated as 5' UTR from transcriptomic data (yellow highlighted nucleotides in figure 5.4). This entire sequence was then analyzed in the software Neural Network Promoter Prediction (BDGP - Berkeley), a program that r identifies putative promoter regions. The only sequence match was a 50 bp sequence from the nucleotide 2710 to the nucleotide 2760 has a prediction of 0.99 (Figure 5.4). Since this putative promoter region is close to the start of the coding sequence there are potential enhancer sequences upstream of this position. Therefore, I chose to isolate the entire 2818 bp region from genomic DNA of the spider with oligonucleotides specifically designed to terminate one nucleotide before the start codon of the *beta-actin* CDS and cloned it upstream of the reporter gene GFP in plasmid pCR4-TopoII with restriction-directed cloning.

This generated plasmid pCR4_Actin_eGFP_SV40PolyA that is now ready to be injected in embryos of different stages and to determine if it is able to drive ubiquitous GFP expression.

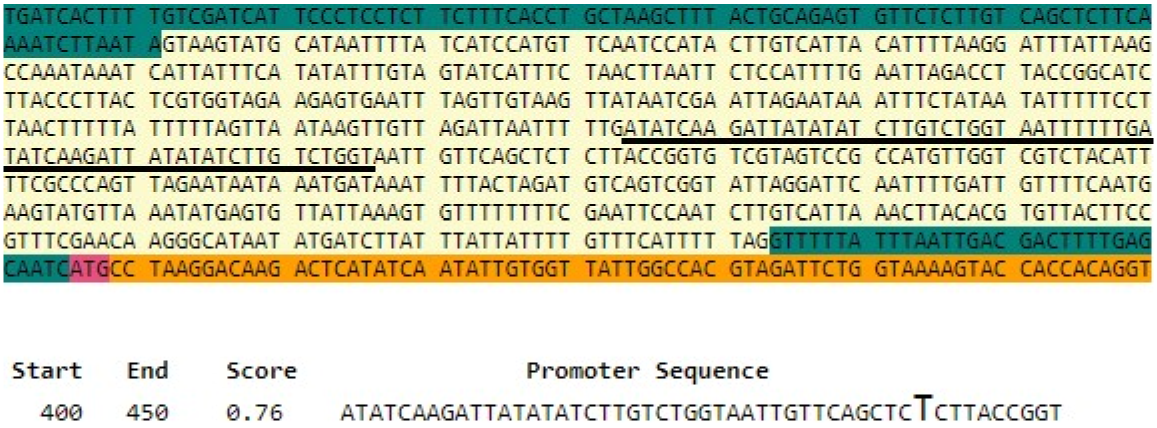
```
GTTGGAAAGG TAGTCACAGA CTCATACCAT ATATGGTAAA GATTGAGAAC ATAGCTTCGC TTAGGTACAC GTGCCATAAA
AGACCGGCAT CACTTCTTTT AAATCGCTGT TTCTTGCTGG TTCTAGTAGT GGTGTGTTCT CTGTGTTGTT AGTTTTCTTT
CGTGTAAGTT AGAAATTTGC TGAATATCTT CACTATTTGG ATATTTATTT GATAATCAAG TGCGGGTGAG CTTAAACAAA
AAGATATACT AATTTTATTT AAATTTATAA ACAGAAAAAT TTATTTGTAA ATTTTCTTAC TGAAATTTGA CAGCTCAGAA
TTCAATTCTG ACATTTTAGG GCGTGTGCCC CCAGTGATGA CGTTGCTTCA AATTTAAAAC CATTTTACTA TCTCTAGAAA
TTTTATACAG ATTTGTATTT TCTACTAATA TATATATTTT ATATACTTTC GTAAATTTAA TTTCTTTTTT AATATTTAAA
TGCTTGTATA ATTTATCTAG ATAAAGTTAA TGCTATTTAA AATTAAGTTG CGATCTGCTA TGCTTTAAAA GAAATTGTTC
ATGTTTTAAT TTCATTAAAT TAAAATTGTA TTACCTAGAA ATAATTACTA ATAAACGTAA CTGTTGTTGA TAAAATCTTT
GAGTTAATTT TAATTTGTAT GTAGGTTTCA ATTTCTGTTG AATATTGATG AAGTTGCTAG ATTTTAAAGA TTTATTTTAT
TTGTTAATCT CTTTTAAGAG ATGTAATTAG AGAGTTTGAT TTTGTCTAAG TTTCTTTATT AAGTTAGGAG TGATTTGTGT
ATTATGAAAT AGATATTATA ATAATGATTT ATAATCTTAA TCTCATGAAG TGTTAAATTG TTTAATTATA TTGCTGAATA
ATTAATTTCG TGTTGTAATG TAAAATATGT AATATAAAAT TTATTTAGTC GCAATTTGAA TGTTTAATAT TAAGGATTAC
ATTGAAGCTA ATAAATGTGG GTATTTCTTA AACAATTTTG ATGTAATCTG GGTGTGTGGA TTTAAATATA ATTATTAAAA
TTTTCTTGTG GTATCAAAAA TATATCTATT GACAATCGTT ATTTCAAAAT GTTTTTTTTT TTTCGCTTAC TAATTTAATT
CAGTTTGTTA AATGGTCGCT GATTACTTTA AAACAGGCTT ATGTTAGATA ATCGTGTCTG ATTTCTCGCT GTATTTTTAA
GTAATTTAAA TTTATTTCTA TGCACTATTA GCAGAATGTA TGTTTCGAAT GAAATTGTTT CTGCTATGGA AACTCTCGTT
CAAATAGTTT TATTGCGTAA TATTTTGCTT TGGCATGATA AATACGTAAA CATTTATGAA ACCTAATTTT TAAATTTTAT
CCGAAGATTC TATCCCCCCC CCCCCCAAGA AACAAACTAG TTTTTTTTAA AATTAAAATA TCTAACATTG TTTCTTAGTT
TGAATTTCTA ACTATTGATT AGTTTTCCAA TATATTTTAT TATGTTTACA TATATAAACC TTGTTTAAAA AGCTAATCCT
TAATTTTATG TCATTCCCGT ACCAGAGTAT TGTTCAGAAT TTTGTCATTT TTTAAAGTGA GAAAGTAGCT CGCGAATTCG
CAATTTGCGA TCCTCCCCCT TTTTTTTTTA AGTAGAAGGT GTAAGATTCT CTGTATATAT CTTTATCTTT TTAGCTATAT
TATCACTGGA CAATTAGTAT GAAAATATTT GGGTAGTGAA ATTTAAATTG CATTGGCGTA ATTTAAATTG AACAAATGTT
GTTAAATGAG ATTTATAATT ATCTGATACA TTAATCTTAT GGACTAGTTC CCTCCCCCCA CGCACAGGCT AAAGTGTGGC
TTCGTTTGAA ATTAACGTTG TGATTCCCCC CCCCCTTTTT TTTTTCTTTC CTTGCCTACA GATACTGTGC CGCGATGCAC
GCTTTGCGTT AACCGGTTAA TTCATATTTT AAATTTCCTC GCAGAAATTG AGTTAATATA ATTTAATAAG AACTCTTGGT
TGGAAAACTG CTTTTAACTT ATAGTTTTTA ATCCTTGAAT TCCGTTCGGT AAATTTTAGT TCACCCTTGA ATTCCGTTTG
ATGGGCAAAT GTTAAAATTA ATATCTTGGT GATTTATTTT TATAATTTAT GGAAAGTAGC CTAAAATTAT TCAGAAGCAG
GAGTTAATTT ATTGTCTATT CGTTAACTTC TTCGAATTTT ATTTCTCTTG AAATTATTAA AAGTTCAATA TTTTAAGGGG
AACCTTAGTG TTTTGGAATT CTAAATTTCG TAAAATCTAT TCAGAGTAAC ATCAAAAGCG TAGTCAACGG GTTTTTCATG
CTCAGTGCTA AAATTCCAAA AACATGTCTC ATTTGTTTAA ATCAAATTGG CCAATTCGAA AAATTGCTAT TAGAGCCTGT
CAAATCATTA CTGAGGAAGG CAGGATGATA ATCAAATTAG TATGTCTGAG ATTTGTTCCA ACTTCATAGA TCTCAGGGCA
GTATTTAAAA TTTCTATAAT TTTGCATAGT GTTGGGTCAT TAGTTTTTTA TTTCGTTTTT GCGTAGTTGT AGATTTGCGA
CCCATAATGT GATCATTCAC TTAATTTTAG AATATTAAAG CAACTACAGT TGCAGAAGAT GGTTTAAGTT GTGCTATTTA
ACCATTTTGG TGTGATAATT GCAGTTATCT GCTTCCCATG CAGCCTCCCT CTCCCTAACA GCTAGGCTAA ACTCTATATA
TTAAAAGGGA CATAAGTTTT CTTGAAGAGT CGTTTAATTG TTTTTTTGAA ATGCTTAATT TTCATCTTTC TCTTACAGGC
TCAAAAAACT AAACCAAAAT GTGTGACGAC GATGTTGCCG CTCTTGTTGT AGACAATGGA TCCGGTATGT GCAAGGCCGG
```

| Start | End | Score | Promoter Sequence |
|---|---|---|---|
| 2710 | 2760 | 0.99 | AACTCTATATATTAAAAGGGACATAAGTTTTCTTGAAGAGTCGTTTAATT |

**Figure 5. 4. Sequence of the 5' UTR from *P. tepidariorum beta-actin*.** The 2818 nucleotides that comprise the region upstream of the *beta-actin* CDS. Marked in yellow are the nucleotides annotated as 5'UTR, in between those are intronic regions that can contain regulatory sequences (underlined in black). Marked in purple is the methionine (first codon that is translated), followed by the first nucleotides that belongs to the *beta-actin* coding sequence (marked in orange).

### *Elongation Factor-1*

In the scaffold containing the annotated EF-1 gene of *P. tepidariorum* (aug3.g27264, scaffold 207), I identified a 725 bp region upstream of the start codon, with two regions annotated as 5’ UTR that span a total of 123 bp. Note that the BLAST search was used only for the identification of the orthologue, and not for the analysis of the putative promoter region. A 50 bp promoter was predicted in the intron using the BDGP program with a score of 0.76 (Figure 5.5). This fragment was isolated in the same way as *beta-actin*, and cloned upstream of the GFP fragment, in a plasmid named pCR4_EF1_eGFP_SV40PolyA.



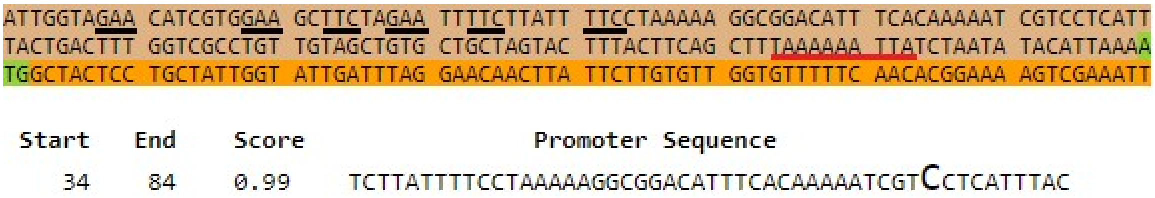
```
TGATCACTTT TGTCGATCAT TCCCTCCTCT TCTTTCACCT GCTAAGCTTT ACTGCAGAGT GTTCTCTTGT CAGCTCTTCA
AAATCTTAAT AGTAAGTATG CATAATTTTA TCATCCATGT TCAATCCATA CTTGTCATTA CATTTTAAGG ATTTATTAAG
CCAAATAAAT CATTATTTCA TATATTTGTA GTATCATTTC TAACTTAATT CTCCATTTTG AATTAGACCT TACCGGCATC
TTACCCTTAC TCGTGGTAGA AGAGTGAATT TAGTTGTAAG TTATAATCGA ATTAGAATAA ATTTCTATAA TATTTTTCCT
TAACTTTTTA TTTTTAGTTA ATAAGTTGTT AGATTAATTT TTGATATCAA GATTATATAT CTTGTCTGGT AATTTTTTGA
TATCAAGATT ATATATCTTG TCTGGTAATT GTTCAGCTCT CTTACCGGTG TCGTAGTCCG CCATGTTGGT CGTCTACATT
TTCGCCCAGT TAGAATAATA AATGATAAAT TTTACTAGAT GTCAGTCGGT ATTAGGATTC AATTTTGATT GTTTTCAATG
AAGTATGTTA AATATGAGTG TTATTAAAGT GTTTTTTTTC GAATTCCAAT CTTGTCATTA AACTTACACG TGTTACTTCC
GTTTCGAACA AGGGCATAAT ATGATCTTAT TTATTATTTT GTTTCATTTT TAGGTTTTTA TTTAATTGAC GACTTTTGAG
CAATCATGCC TAAGGACAAG ACTCATATCA ATATTGTGGT TATTGGCCAC GTAGATTCTG GTAAAAGTAC CACCACAGGT
```

```
Start   End   Score             Promoter Sequence
  400   450   0.76    ATATCAAGATTATATATCTTGTCTGGTAATTGTTCAGCTCTCTTACCGGT
```

**Figure 5. 5. Sequence of the 5’ UTR from *P. tepidariorum elongation factor-1*.** A 725 bp region upstream of the start codon (marked in purple) was isolated from the scaffold that contains the *elongation factor-1* gene. In blue are marked the 5’ UTR regions of the scaffold, followed by an intron. The CDS is marked in orange after the start codon (ATG). Below is the predicted promoter region (50 bp underlined in black), that contains the highlighted T as the putative TSS.

### *Heat-shock protein 70*

One of the heat-shock proteins found by BLAST in the *P. tepidariorum* genome (aug3.g27068, scaffold 191) is predicted to have a small 5’ UTR 159 nucleotides, and the promoter prediction showed a single but specific sequence (Figure 5.6). I also found six predicted Heat Shock Elements (HSE) in this small sequence (GAA – nn – TCC repeats), and a robust TATA-box 30 bp upstream of the start codon (Figure 5.6). The fragment was cloned upstream of the GFP as for the other genes, in a plasmid named pCR4_Hsp70_eGFP_SV40PolyA.



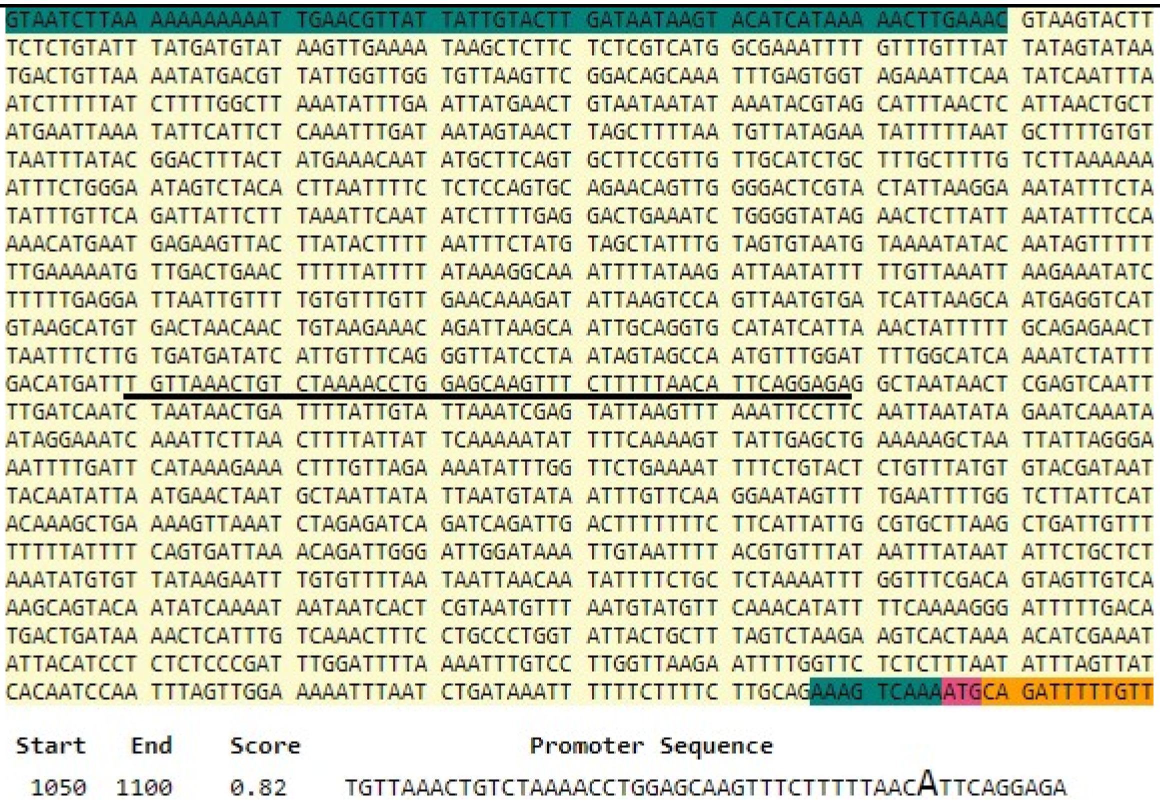
```
ATTGGTAGAA CATCGTGGAA GCTTCTAGAA TTTTCTTATT TTCCTAAAAA GGCGGACATT TCACAAAAAT CGTCCTCATT
TACTGACTTT GGTCGCCTGT TGTAGCTGTG CTGCTAGTAC TTTACTTCAG CTTTAAAAAA TTATCTAATA TACATTAAAA
TGGCTACTCC TGCTATTGGT ATTGATTTAG GAACAACTTA TTCTTGTGTT GGTGTTTTTC AACACGGAAA AGTCGAAATT

 Start    End     Score                  Promoter Sequence
   34      84     0.99     TCTTATTTTCCTAAAAAGGCGGACATTTCACAAAAATCGTCCTCATTTAC
```

**Figure 5. 6. Sequence of the 5' UTR from *P. tepidariorum heat shock protein 70*.** In light brown is marked the 5' UTR of the *hsp70* gene. The HSE are underlined in black (GAA – NN – TTC), and the TATA-box underlined in red. The start codon is marked in green, and the CDS in orange.

### *PolyUbiquitin*

The BLAST search for the spider *polyUbiquitin* gene identifed a scaffold (aug3.g26922, scaffold 180) containing 1985 bp upstream of the start codon of this gene (Figure 5.7). The BGDP promoter search suggests that there is a putative promoter located in the intronic region between nucleotides -1050 and -1100 (Figure 5.7). An A-T rich region was also identified around 30 bp upstream of the start codon that could be a TATA-box (Figure 5.7). The 1985 bp fragment was isolated and cloned upstream of the GFP coding sequence, in a plasmid named pCR4_pUb_eGFP_SV40PolyA.



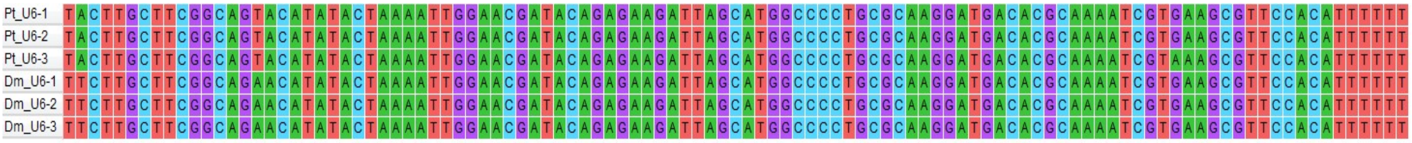
```
GTAATCTTAA AAAAAAAAAT TGAACGTTAT TATTGTACTT GATAATAAGT ACATCATAAA AACTTGAAAC GTAAGTACTT
TCTCTGTATT TATGATGTAT AAGTTGAAAA TAAGCTCTTC TCTCGTCATG GCGAAATTTT GTTTGTTTAT TATAGTATAA
TGACTGTTAA AATATGACGT TATTGGTTGG TGTTAAGTTC GGACAGCAAA TTTGAGTGGT AGAAATTCAA TATCAATTTA
ATCTTTTTAT CTTTTGGCTT AAATATTTGA ATTATGAACT GTAATAATAT AAATACGTAG CATTTAACTC ATTAACTGCT
ATGAATTAAA TATTCATTCT CAAATTTGAT AATAGTAACT TAGCTTTTAA TGTTATAGAA TATTTTTAAT GCTTTTGTGT
TAATTTATAC GGACTTTACT ATGAAACAAT ATGCTTCAGT GCTTCCGTTG TTGCATCTGC TTTGCTTTTG TCTTAAAAAA
ATTTCTGGGA ATAGTCTACA CTTAATTTTC TCTCCAGTGC AGAACAGTTG GGGACTCGTA CTATTAAGGA AATATTTCTA
TATTTGTTCA GATTATTCTT TAAATTCAAT ATCTTTTGAG GACTGAAATC TGGGGTATAG AACTCTTATT AATATTTCCA
AAACATGAAT GAGAAGTTAC TTATACTTTT AATTTCTATG TAGCTATTTG TAGTGTAATG TAAAATATAC AATAGTTTTT
TTGAAAAATG TTGACTGAAC TTTTTATTTT ATAAAGGCAA ATTTTATAAG ATTAATATTT TTGTTAAATT AAGAAATATC
TTTTTGAGGA TTAATTGTTT TGTGTTTGTT GAACAAAGAT ATTAAGTCCA GTTAATGTGA TCATTAAGCA ATGAGGTCAT
GTAAGCATGT GACTAACAAC TGTAAGAAAC AGATTAAGCA ATTGCAGGTG CATATCATTA AACTATTTTT GCAGAGAACT
TAATTTCTTG TGATGATATC ATTGTTTCAG GGTTATCCTA ATAGTAGCCA ATGTTTGGAT TTTGGCATCA AAATCTATTT
GACATGATTT GTTAAACTGT CTAAAACCTG GAGCAAGTTT CTTTTTAACA TTCAGGAGAG GCTAATAACT CGAGTCAATT
TTGATCAATC TAATAACTGA TTTTATTGTA TTAAATCGAG TATTAAGTTT AAATTCCTTC AATTAATATA GAATCAAATA
ATAGGAAATC AAATTCTTAA CTTTTATTAT TCAAAAATAT TTTCAAAAGT TATTGAGCTG AAAAAGCTAA TTATTAGGGA
AATTTTGATT CATAAAGAAA CTTTGTTAGA AAATATTTGG TTCTGAAAAT TTTCTGTACT CTGTTTATGT GTACGATAAT
TACAATATTA ATGAACTAAT GCTAATTATA TTAATGTATA ATTTGTTCAA GGAATAGTTT TGAATTTTGG TCTTATTCAT
ACAAAGCTGA AAAGTTAAAT CTAGAGATCA GATCAGATTG ACTTTTTTTC TTCATTATTG CGTGCTTAAG CTGATTGTTT
TTTTTATTTT CAGTGATTAA ACAGATTGGG ATTGGATAAA TTGTAATTTT ACGTGTTTAT AATTTATAAT ATTCTGCTCT
AAATATGTGT TATAAGAATT TGTGTTTTAA TAATTAACAA TATTTTCTGC TCTAAAATTT GGTTTCGACA GTAGTTGTCA
AAGCAGTACA ATATCAAAAT AATAATCACT CGTAATGTTT AATGTATGTT CAAACATATT TTCAAAAGGG ATTTTTGACA
TGACTGATAA AACTCATTTG TCAAACTTTC CTGCCCTGGT ATTACTGCTT TAGTCTAAGA AGTCACTAAA ACATCGAAAT
ATTACATCCT CTCTCCCGAT TTGGATTTTA AAATTTGTCC TTGGTTAAGA ATTTTGGTTC TCTCTTTAAT ATTTAGTTAT
CACAATCCAA TTTAGTTGGA AAAATTTAAT CTGATAAATT TTTTCTTTTC TTGCAGAAAG TCAAAATGCA GATTTTTGTT
```

```
 Start   End    Score              Promoter Sequence
  1050   1100   0.82    TGTTAAACTGTCTAAAACCTGGAGCAAGTTTCTTTTTAACATTCAGGAGA
```

**Figure 5. 7. Sequence of the 5' UTR from *P. tepidariorum polyUbiquitin*.** The 5' UTR annotated sequence is marked in dark green, with a big intronic region in between the two fragments. Purple marked is the start codon, followed by the orange marked CDS. The putative promoter sequence contains the highlighted A as the TSS (underlined in black).

## Characterization and isolation of RNApol III-associated promoters

Nucleotide BLAST search for the U6 snRNA in *P. tepidariorum* was carried out with the *D. melanogaster* U6-1 (NR_002081), U6-2 (NR_002082) and U6-3 (NR_002083) promoter sequences. These three U6 snRNA's are each 107 bp, and three regions were found in the spider genome that show more than 99% similarity with the *Drosophila* sequences (Figure 5.8). I thus annotated these regions as Pt_U6-1 (scaffold 1617), Pt_U6-2 (scaffold 523) and Pt_U6-3 (scaffold 687).



```
Pt_U6-1  TACTTGCTTCGGCAGTACATATACTAAAATTGGAACGATACAGAGAAGATTAGCATGGCCCCTGCGCAAGGATGACACGCAAAATCGTGAAGCGTTCCACATTTTTT
Pt_U6-2  TACTTGCTTCGGCAGTACATATACTAAAATTGGAACGATACAGAGAAGATTAGCATGGCCCCTGCGCAAGGATGACACGCAAAATCGTGAAGCGTTCCACATTTTTT
Pt_U6-3  TACTTGCTTCGGCAGTACATATACTAAAATTGGAACGATACAGAGAAGATTAGCATGGCCCCTGCGCAAGGATGACACGCAAAATCGTAAAGCGTTCCACATTTTTT
Dm_U6-1  TTCTTGCTTCGGCAGAACATATACTAAAATTGGAACGATACAGAGAAGATTAGCATGGCCCCTGCGCAAGGATGACACGCAAAATCGTGAAGCGTTCCACATTTTTT
Dm_U6-2  TTCTTGCTTCGGCAGAACATATACTAAAATTGGAACGATACAGAGAAGATTAGCATGGCCCCTGCGCAAGGATGACACGCAAAATCGTGAAGCGTTCCACATTTTTT
Dm_U6-3  TTCTTGCTTCGGCAGAACATATACTAAAATTGGAACGATACAGAGAAGATTAGCATGGCCCCTGCGCAAGGATGACACGCAAAATCGTGAAGCGTTCCACATTTTTT
```

**Figure 5. 8. Alignment of the *D. melanogaster* and *P. tepidariorum* predicted U6 snRNAs.** Alignment of the 107 bp region of the snRNAs. There is an A to T change at nucleotide 2, a T to A at nucleotide 15 in the three U6s in *P. tepidariorum*, and a G to A at nucleotide 98 in the spider U6-3 compared to *D. melanogaster*.

I then focused my analysis on the region upstream of the putative TSS (the G marked with the arrow in figure 5.9) for these three sequences. Strikingly, U6-1 and U62 have a conserved TATA-box 32 bp upstream of the TSS (underlined in black in figure 5.9), which is completely absent in the U6-3. Therefore, I focused on the isolation of the U6 1 and 2. I designed specific primers than spanned a 650 bp region upstream of the TSS and amplified this fragment from the genomic DNA of *P. tepidariorum*. This was based in other studies, where the isolation of the U6 promoters never spanned regions larger than 600 bp (Bassett, 2014; Gilles, 2015).

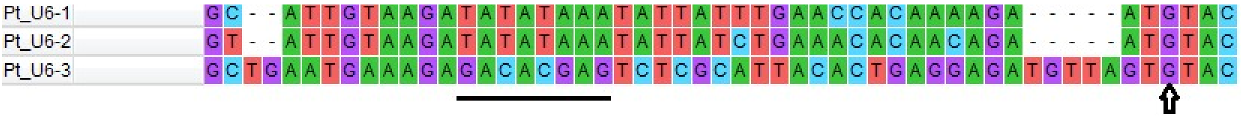
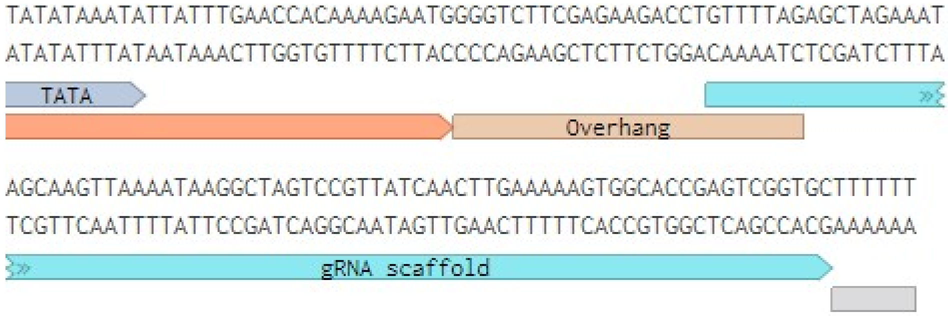


**Figure 5. 9. Alignment of the *P. tepidariorum* predicted U6 regulatory regions.** Alignment of the three putative U6 snRNA's of the spider. Underlined in black is the region identified as the TATA-box, 32 bp upstream of the TSS (black arrow in the G). The putative regulatory regions for the U6-1 and 2 have a striking high homology.

To generate a plasmid that can drive the endogenous expression of the gRNAs, I then cloned the 650 bp fragments from U6-1 and U6-2 upstream of the tracRNA complex.

This region contains a BbsI overhang (for insertion of the specific guide with a single Gibson assembly cloning), a 76 bp *S. pyogenes* specific tracRNA (which will complex with the Cas9 enzyme) and 6 T's that are the RNApol III transcription stop signal (Figure 5.10). The plasmids were named pSFD1-U6:1_BbsI_gRNA and pSFD1-

U6:2_BbsI_gRNA, respectively.



```
TATATAAATATTATTTGAACCACAAAAGAATGGGGTCTTCGAGAAGACCTGTTTTAGAGCTAGAAAT
ATATATTTATAATAAACTTGGTGTTTTCTTACCCCAGAAGCTCTTCTGGACAAAATCTCGATCTTTA

AGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTT
TCGTTCAATTTTATTCCGATCAGGCAATAGTTGAACTTTTTCACCGTGGCTCAGCCACGAAAAAA
```

**Figure 5. 10. Sequence of the pSFD1-U6:1_BbsI_gRNA promoter and tracRNA region.** The 650 bp putative promoter region for the U6-1 of the spider was cloned upstream an overhang that contains the BbsI restriction enzyme site, a gRNA scaffold that will transcribe the tracRNA, and 6 repeated T nucleotides, that are the RNApol III transcription stop signal.

### Micro-injection of spider embryos

As a proof of principle that spider embryos can be inhjeced with genetic material to facilitate gene expression and potentially germ line transformation, following the protocol of the Oda group (Kanayama *et al*., 2010), I injected a capped mRNA corresponding to the coding sequence of GFP into1, 2 and 4-nuclei stage *P. tepidariorum* embryos (without chorion removal). This resulted in all of the blastomeres expressing of GFP in stage 5 embryos (e.g. Figure 5.11). How many died? – i.e. give some idea of the efficiency.

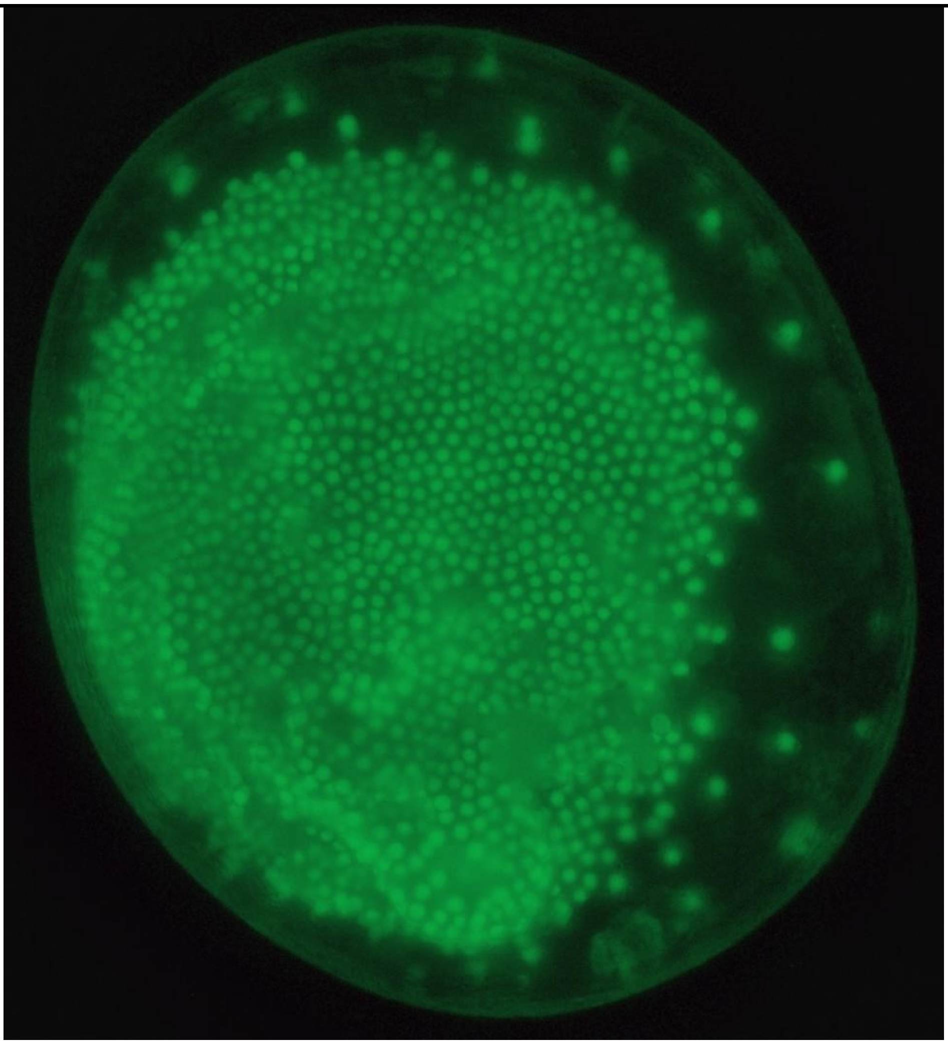


**Figure 5. 11. Injection of eGFP mRNA in early stage spider embryos.** 2-cell stage embryos were injected with capped mRNA for the eGFP coding sequence, and it is visible the strong expression of this construct in late stage 5 embryos.

This result was confirmed later by Pechmann in his 2016 paper on germ disc development in *P. tepidariorum*, in which the injections of FITC and Lyn-GFP at earlier stages gave rise to the expression of these throughout the embryos. This success means that there is great potential for new technologies like CRISPR/Cas9 to be efficiently delivered into the embryos.

I also injected embryos at a range of early stages with another marker – a dsRED that contains a nuclear localization signal (NLS) at the N-terminal. The injections into 4, 8 and 16 cell stage embryos (Figure 5.12) demonstrated the ability of the constructs to be transcribed in daughter cells of the injected blastomeres.

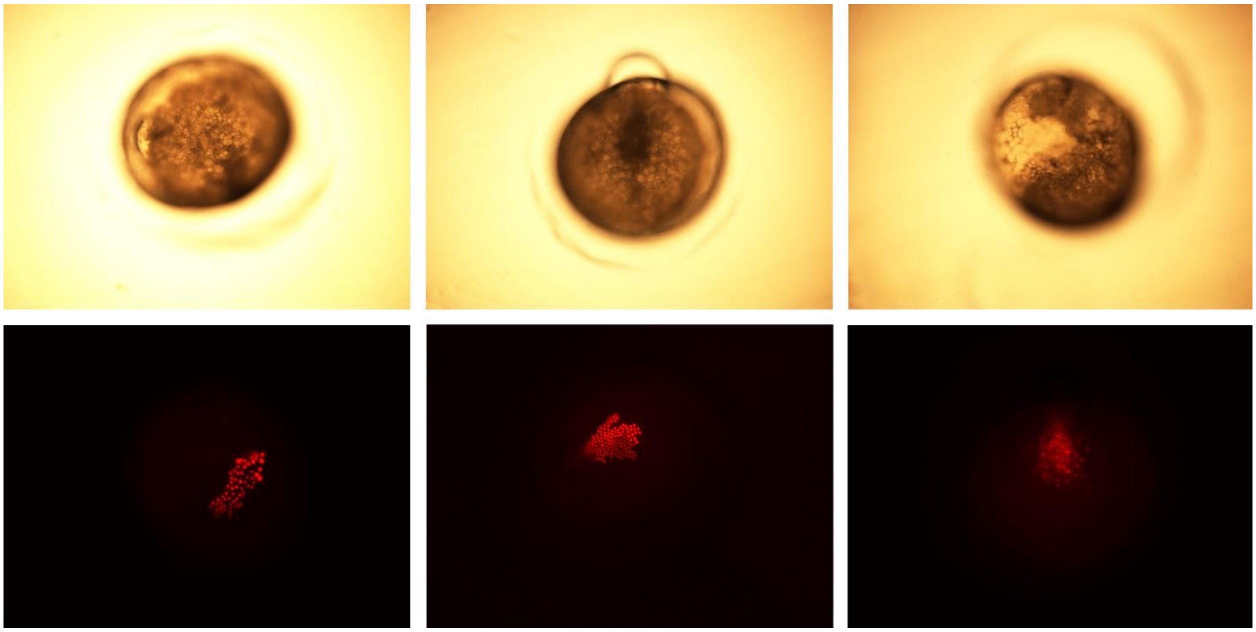


**Figure 5. 12. Embryos injected with fluorescent mRNA.** From left to right on top: Bright-field pictures of stage 5 and 7 embryos that were injected at 4, 8 and 16 cell stages. Lower pictures show the dsRed expression of the RNA that was injected at each stage.

I next asked if embryos injected prior to the 16 cells stage could give drive ubiquitous expression in the embryo. Note that Kanayama (2010) states that embryos before the 4-cell stage are unable to drive expression – or don't survive the injection (see figure 5.11).

Figure 5.13 shows an embryo that was injected at 1 cell stage with the same invitro transcribed RNA (dsRed) at nearly 128-cells stage. Therefore, it is possible to inject early embryos to give rise to expression in all blastomeres at later stages.

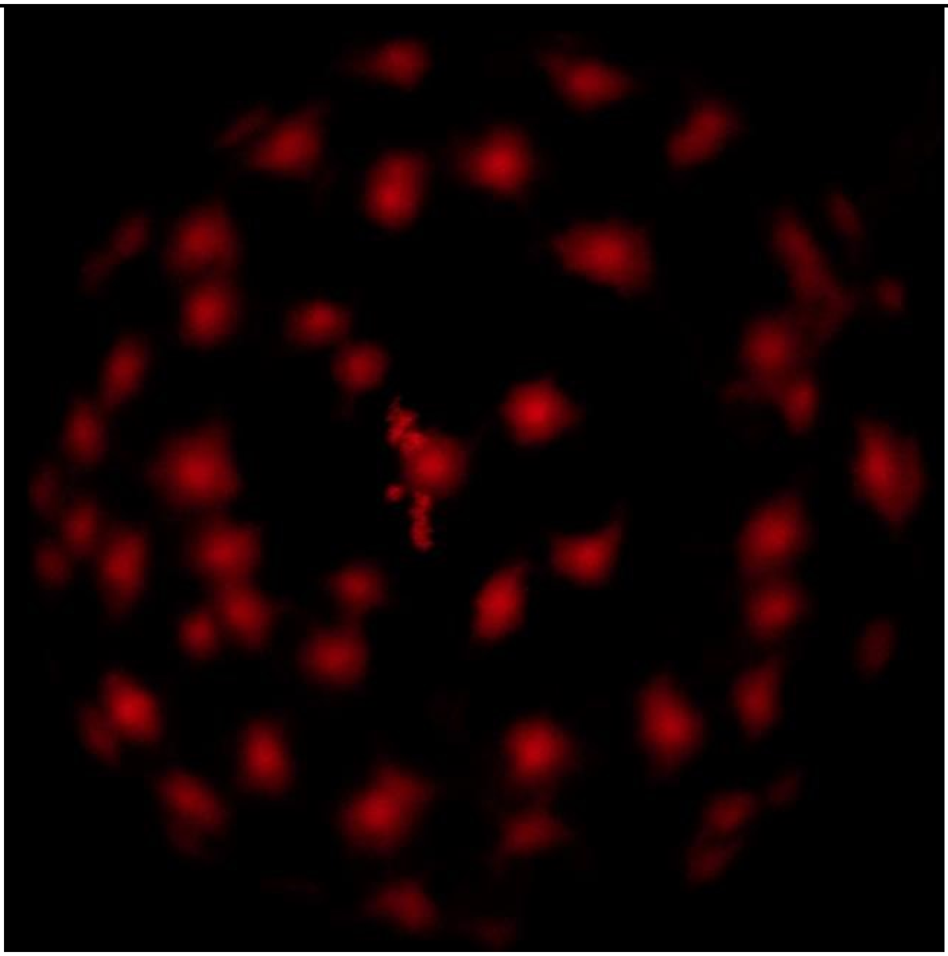


**Figure 5. 13. Embryos injected with dsRed mRNA.** Fluorescent (red channel) snap picture of an embryo injected at 1 cell stage with in-vitro transcribed mRNA for the dsRed with nuclear localization. The embryo is at 128 cell stage, and all the cells in the blastoderm are expressing the injected capped-mRNA.

## Discussion

It has been a standard protocol for spider studies to use RNA interference both embryonically and parentally to assay the function of specific genes during embryonic development (Schoppmeier, 2001; Akiyama-Oda & Oda, 2003). However, this approach has its pros and cons, depending on the gene to be studied, and it is not fully understood why some RNAi mediated gene knockdowns are not efficient in the spider, whether injected parentally or embryonically. For example, in my experience with parental RNAi, I found that it was very efficient for Sox21b-1, but I also injected the Hes-like and Sox both parentally and embryonically, and I never observed a phenotype. Boutros & Ahringer (2008) discussed the use of RNAi for screening in different organisms, and they clearly state that there are disadvantages such as off-target effects, incomplete penetrance and variability caused by the decrease of the dsRNA over developmental time. However, we don’t really understand why it works for some genes bit not others in *P. tepidariorum*.

*P. tepidariorum* embryogenesis starts with intralecithal cleavages until around the 16-cell stage (Kondo, 1969; Anderson, 1973; Kanayama, 2010), which means that there are no membranes in between the single syncytial nuclei. This allows the injection of different RNA’s in the embryo that will be processed during the development. The embryos at early stages of development are very delicate, with the chorionic membrane still being formed, so if there is a dehydration caused by the injection procedure, the survival rate will be low (Suzuki, 1995; Kanayama, 2010). Therefore, injections at slightly later stages (8 or 16 cell stages) that will produce a clonal cell inheritance of the dsRNA and mRNA more efficiently is now common in *P. tepidariorum* (Kanayama, 2010; Schönauer *et al.*, 2016; Pechmann, 2016; Pechmann, 2017). However, as I showed in the results, injections in early stages, when there are only 1 or 2 nuclei, successfully gave rise to cells expressing GFP, which is a proof of principle that germ line transformations can be achieved in this organism.

However, to generate a transgenic line and better understand the function of a complete knockout of a specific gene, it would be desirable to be able to generate complete knock out of the gene of interest – even to compare to RNAi results. Thus, I focused my efforts on trying to continue Anna Schonauer's work in establishing the CRISPR/Cas9 technique in the common house spider. Several efforts were previously made with synthetically transcribed gRNAs, with the Cas9 delivered as mRNA or protein at different concentrations, but with no success.

For this reason, I followed the approach that has been used for several other arthropod species to implement the CRISPR technique with a high efficiency: plasmid engineering that allows the expression of the Cas9 endonuclease from an endogenous RNAPol II and gRNAs from RNAPol III promoters (Gilles, 2015; Sun *et al*., 2017).

Starting with RNAPol II putative regulatory regions, I isolated the region upstream of the start codon (ATG) for the *beta-actin* gene (aug3.g27982) comprising 2818 bp and with putative regions of transcriptional initiation and promoter. This is an orthologue of the *D. melanogaster* cytoskeletal actin gene, that is ubiquitously expressed from early stages of development, which is consistent with the high expression levels that I identified in the transcriptome of the spider for this gene in early stages of embryogenesis (Chung & Keller, 1990; Quitschke *et al*., 1990). In the fruit fly, this gene has a proximal and a distal promoter (Chung & Keller, 1990), but analysis of the sequences of introns 2 and 3 did not show any putative regulatory regions that could be homologous to the distal promoter seen in *Drosophila*.

Elongation-Factor-1 belongs to an enzymatic complex that delivers tRNAs to the ribosomes, transcribed from when fertilization occurs. It has a strong promoter used in cell lines and mammalian studies and is considered to be stronger than the cytomegalovirus promoter (CMV) (Tokushige *et al,* 1997; Teschendorf *et al*., 2002).

Heat-shock proteins contain the most studied inducible promoters amongst both animals and plants (Amin, 1988). The constructs that untilise heat shock promoters have the ability of inducing the expression of transcripts at a specific time of the development (Shogi, 2008; Pavlopoulos, 2009). Also, these regulatory regions are relatively easy to identify by the fact that they contain HSE – specific nucleotides (GAA – nn – TTC) (Pavlopoulos, 2009; Schinko, 2012). Recap what you identified as well here – very briefly.

The polyubiquitin protein is one of the many ubiquitin-complex genes, responsible for different aspects of organelle and membrane production. Thus, it has the characteristic of being expressed at high levels throughout the entire life of an organism (Maekawa *et al*., 2008; Anderson *et al*., 2010). It's been used as a constitutive promoter in different research organisms (Garbarino, 1995; Maekawa *et al*., 2008).

The fragments that I isolated for the putative regulatory regions of the genes discussed above were cloned upstream of an eGFP coding sequence and the SV40 RNAPol II transcription stop signal (poly A). If these putative promoters are suitable for driving the expression of the GFP, the technology can then be modified to drive the expression of the Cas9 protein endogenously, and also of different types of constructs (e.g. live imaging for the SAZ morphology or neurogenical progenitors differentiation).

In regard to the identification and isolation of RNAPol III putative regulatory regions, I focused my efforts on the U6 snRNAs, instead of trying to find orthologues for the H1 and 7SK in *P. tepidariorum*, because U6s are the most widely used and most strongly expressed amongst them all (Cummins, 2011; Gao *et al*., 2017).

The use of the U6 promoters for gRNAs transcription has been used in different organisms, especially in arthropods such as *Bombyx mori* (Ma *et al*., 2013), *D. melanogaster* (Ren *et al*., 2013), *Plutella xylostella* (Huang *et al*., 2017), *Tribolium castaneum* (Gilles, 2015), and others. Most arthropods where this technique has been established exhibit a great amount of yolk in their eggs (see Anderson, 1973 for a summary of egg types), and the improvement of the CRISPR technique was helped with the use of endogenously-transcribed Cas9 and gRNAs. Strikingly, there is still currently no study in chelicerates evidencing CRISPR/Cas9 working even with minimal efficiency. This might be due to a response from the cell-cycle proteins against the Cas9 protein, in order to protect the correct development of the embryo in chelicerates. For example, it was seen recently in human cell lines that the p53 protein acts against the Cas9, and when this protein was knocked-out, the efficiency of the CRISPR technique was increased to 80% (Ihry, 2018). Therefore, intrinsic mechanisms might have evolved in order to defend the cells from external factors. The other explanation might be that the Cas9 protein is not responsive in the chelicerates cells, and a codon optimization is needed before the injection of this protein in the embryos (Weninger *et al*., 2016).

However, it is likely that CRISPR/Cas9 should work in spiders and my efforts in identifying putative promoters for endogenous expression in the spider embryos will provide new powerful tools for genomic editing in spiders. Assay for Transposase-

Although I generated all the plasmids, I did not have the time for the injections. My hope is that the current members of the lab, who are interest in functional studies in spiders can pursue this work with the aid of the plasmids that I designed, and possible CRISPR/Cas9 a routine tool to understand the development of chelicerates.

# General Discussion

## Results Summary

In this thesis, I focused my efforts on the further characterization of two conserved metazoan gene families: Hes and Sox. I started with the isolation of four hes-like genes and analysis of their expression in the embryos of the spider *P. tepidariorum* (Chapter 1). This contributed to a better understanding of the similarities and differences in the expression of these genes among species with segmented bodies. To further extend my studies of genes involved in embryogenesis, I expanded my analysis to the fifteen Sox factors present in the spider genome (Chapter 2). This investigation provided insights into how such genes are expressed in the embryonic development of the spider in particular during segmentation and neurogenesis, but functional essays were needed. Thus, I tried to knockdown all 15 genes with RNAi. However, I only observed a phenotype for a SoxB gene, *Sox21b-1*, which probably arose via WGD. This gene is expressed in the prosoma, SAZ and neuroectoderm. I therefore, focused my analysis on the regulation of *Sox21b-1*, in order to better understand its role in the spider development (Chapter 3). I found that this gene regulates a range of processes including segmentation. The experiments I have done in my PhD also led me to pursue a more efficient method for functional analysis in spiders, and in particular to generate better tools to apply CRISPR/Cas9 which is growing in use in developmental biology studies in a wide range of organisms (Gilles, 2015). Therefore. in Chapter 4 I described the characterization of putative regulatory regions that can be used to drive endogenous expression of different factors, particularly Cas9 protein. I also explored the injecting markers into very early embryos. Finally, here in this general discussion chapter I will highlight the main findings of my work and propose future directions for the study of developmental biology in organisms such as the chelicerates.

## Spider Hes-like and Sox orthologues and WGD

My characterisation of the Hes-like and Sox orthologues in the *P. tepidariorum* genome, and their expression demonstrates the visible consequences of a WGD, which was recently described for this species (Schwager *et al*., 2017). Many genes encoding transcription factors have been retained as duplicates in spiders (and scorpions) following this WGD. For example, a high rate of retention of duplicated orthologues of homeobox genes was shown by Leite *et al* (2018). This study also showed that several of the duplicated genes acquired new expression patterns (neofunctionalization) and or had been subject to subfunctionalization when compared the expression of their single copy orthologues in another chelicerate that did not have an ancestral WGD, the harvestmen *Phalangium opilio*.

For the gene families I studied, I found that some of the duplicated Sox factors such as *PtSox21a-1*, *PtSox21a-2*, *PtSoxC-2*, *PtSoxD-1*, *PtSoxE-1* and *PtSoxF-1,* and the single-copy *Pt-D* of show no detectable expression in the stages that were analysed. The lack of expression in the embryos means that they are likely to be expressed at other developmental stages or in adults and therefore this may represent subfunctionalization and/or neofunctionalisation after the WGD. Indeed, in comparison, the expression of *PtSoxC-1, PtSoxD-2* and *PtSoxF-2* appears to be conserved because it resembles what is seen for the SoxC, D and F orthologues in other insects, which the expression is solely in the ventral nerve cord (Sox C and D) or in the mechanoreceptors in the limbs (SoxF), (Janssen *et al*., 2018; Paese *et al*., 2018). However, the expression of *PtSoxE-2* appears to have diverged when compared to insects (Janssen *et al*., 2018). While *G. marginata* and *T. castaneum* have genes of the SoxE group expressed in the undifferentiated gonadal tissue, the counterpart in the spider is shown to be expressed in the PNS, similar to what is seen for *sox9* in vertebrates (Paese *et al*., 2018).

Interestingly, while I observed a high rate of retention of duplicated Sox genes consistent with homeobox genes and other regulatory factors like microRNAs (ref), I found that the hes-like genes were only present as single copies. Therefore, it appears that either by chance or because of their specific functions and perhaps dosage, duplicated copies of these genes have been lost since WGD. It would be interesting to study if this is also the case in other spiders and scorpions more generally.

## A Sox factor regulates segmentation with Delta-Notch and Wnt signalling

I described the involvement of a SoxB group orthologue (*PtSox21b-1*) during segmentation in *P. tepidariorum*. In other arthropods (long and short germ band) that have not undergone a WGD, there seems to be a similar role for the SoxB gene *D* (Russell *et al*., 1996; Clark & Peel, 2018), and the expression patterns seen for other insects and the onychophoran *E. kanagrensis* suggests that *D* and *Sox21B* are involved with segmentation in these species (Janssen *et al*., 2018). Strikingly, the gene that I identified as *Pt-D* shows no detectable expression in the stages observed and the RNAi experiments did not result in a phenotype. This is an indication that after WGD *PtSox21b1* may have acquired an ancestral *D* function in segmentation in spiders, and this analysis could be much improved if other arachnids were studied in this regard, especially those that have not gone a WGD, like the harvestmen (Leite *et al*., 2018).

The novelty in my work is not only in characterizing another gene that belongs to the gene regulatory network underlying this process, but in demonstrating that this transcription factor acts upstream of both Notch and Wnt signalling. It has been demonstrated that these signallings have important roles in spider segmentation (Oda *et al*, 2007; McGregor *et al*., 2008) furthermore, these two pathways are upstream of the factor *caudal*, and thus trigger the pair-rule gene orthologues *even-skipped* and *runt* in the mechanism of sequential addition of segments (McGregor *et al.,* 2009; Schonauer *et al*., 2016). My work shows that a Sox gene likely activates the expression of this GRN.

The involvement of Delta-Notch and Wnt signalling has been demonstrated for the cockroach *P. americana* (Chesebro, 2013), and the expression pattern of the *delta* orthologues in *T. castaneum* and *S. maritima* are a tantalizing suggestion that these genes also regulate segment addition in other short germ arthropods (Aranda, 2008; Chipman, 2008). Therefore, it will be interesting in the future to see if a SoxB genes plays a similar role upstream of Delta-Notch and Wnt signalling in these and other arthropod species.

### A gap-gene like function for *Sox21b-1*?

Loss of L2-L4 upon knockdown of Sox21b-1 in *P. tepidariorum* also resembled a gap like phenotype. And while previous studies have demonstrated that there are major differences in the regulation of the prosoma versus segment addition from the opisthosoma in spiders, I have found that Sox21b-1 is involved in segmentation of both tagma (McGregor *et* al., 2008; Pechmann *et al*., 2009; Schwager *et al*., 2009; Schonauer *et al*., 2016).

In the anterior of the prosoma, it is known that *orthodenticle* and *hedgehog* are important regulators of the head patterning in *P. tepidariorum*, a process that involves the splitting wave of *Pt-hh* from the pedipalpal segment (Pechmann *et al*., 2009; Kanayama, 2011). In *PtSox21b-1* RNAi embryos, the expression of *Pt-otd* is completely lost from stage 5 embryos (data not shown), but *Pt-hh* is still expressed at the rim of the germ disc in stage 5, and in the L1 segment at stage 9. In *D. melanogaster* and *S. maritima*, genes such as *orthodenticle* and *hedgehog* are essential for head patterning, but the involvement of Sox factors has not yet been described (Hunnekuhl, 2017). Therefore, it would be interesting to determine if the interplay between these factors and the regulation by a SoxB gene is a conserved feature in arthropod head development.

I also suggest that would be interesting to further analyse the apparent gap-like function of *PtSox21b-1* in the spider. The genes l*distal-less* and *hunchback* have been shown to give rise to a gap-like phenotype after knockdown by RNAi in *P. tepidariorum* (Schwager *et al*., 2009; Pechmann *et al*., 2011). *Pt-Dll* knockdown embryos lose the L1 segment, while *Pt-hb* knockdown removes the L1 and L2 segments. Since knockdown of *PtSox21b-1* leads to the loss of segments from either L1 (Class I phenotype) or the pedipalpal segment (Class II phenotype), this suggests this Sox factor may regulate the two aforementioned gap genes as well as the prosomal expression of hairy in segments L2-L4. It will be interesting to test this further in the future to better understand prosomal segmentation.

### Secondary layers formation is decreased in Sox21b-1 knockdowns

Another interesting effect of the *Sox21b-1* knockdown in *P. tepidariorum* regards germ layer formation. In mammalians, *sox2*, a SoxB group gene, is responsible for the maintenance of the ectodermal and neural progenitors (Zhang, 2014). Furthermore, *sox2* is dependent on *tbx6* in order to induce the epiblast to form the caudal neural plate and the paraxial mesoderm (Takemoto *et al*., 2011), having thus a role in the differentiation of ectodermal progenitors to the EMT process, and then a control in the mesodermdifferentiating genes.

I suggest that *PtSox21b-1* is also upstream of mesodermal and endodermal differentiating genes in *P. tepidariorum*. This was shown by the loss of *Pt-fkh* expression in stage 5 embryos, and also apparent loss of *Pt-twi* in several regions of the embryo with the exception of anterior regions (Figure 4.6). Indeed, orthogonal projections show perturbation of secondary layers in stage 9 embryos (Figure 4.6 – K-N). At the cellular level, it was shown that the cells have a bigger nucleus and there is less cell division. So far in the literature, this is the first description of a SoxB gene regulating various aspects of segmentation in arthropods, and also that this control might be due to the intrinsic potential that this Sox group have in undifferentiated cells. Thus, functional analysis of SoxB orthologues in other species should be carried out to increase our knowledge of how these processes evolved in the different clades, and how they might have been coopted to control the process of layer formation in arthropods.

## Future Directions

Chelicerates have been used for different aspects of research for a long time, and for embryology in particular (Montgomery, 1909; Holm, 1954; Abzhanov, 2001; McGregor *et al*., 2009; Hilbrant, 2012). Nowadays, studies of the common house spider *P. tepidariorum* are facilitated by the availability of genomic and transcriptomic resources

(Posnien *et al*., 2014; Schwager *et al*., 2017), and it was shown that the lineage leading to this species was subject to a WGD. There is much still to be learned about the development of this spider and how this compares to other animals as well as the consequences of WGD. While excellent tools already exist to study gene expression and function in this spider, including the ability to deliver in-vitro transcribed constructs like dsRNA and mRNA into the eggs (Schonauer *et al*., 2016; Pechmann, 2016), new tools need to be developed to utilise this organism further.

CRISPR/Cas9 has already proven to be a key breakthrough in functional genetics that has been applied to a range of different species in to tackle questions in the field of developmental biology (Gilles, 2015; Sun *et al*., 2017). The efficiency of this procedure is improved when the Cas9 protein and the sgRNAs expression is driven from endogenous promoters. I focused on the identification of a series of different promoters to produce constructs with the potential to drive the expression in spider embryos. Once this technique can be applied in *P. tepidariorum*, the understanding of regulation of duplicated genes (e.g. Sox family members) can be investigated further because it will be possible to tag the endogenous locus with fluorescent proteins, and thus evaluate their roles in the spider embryogenesis as well as to study their cis-regulatory regions.

This thesis further evidences that the WGD in chelicerates led to ontogenetic changes. However, this can only be understood if there are further studies of different species of this clade. Therefore, sequencing and analysing species such as the harvestmen and other arachnids that have not undergone a round of genomic duplication will provide important insights in the regulation of segmentation and neurogenesis, as well as many other features of embryogenesis in these animals.

# Material and Methods

## 2.1 Embryological Procedures

If not stated otherwise, all procedures were carried out according to the manufacturer's guidelines.

### Spider Culture, Husbandry and Embryology

The *P. tepidariorum* culture is kept in a temperature and humidity-controlled room (25°C and 50%, respectively) at Oxford Brookes University. Spiderlings (after their first moults) were kept in Petri dishes that contained a piece of moist Whatman paper to maintain the humidity and fed with *D. melanogaster vestigial* mutants. After a few weeks, when the sex can be morphologically determined, the juvenile spiders were transferred into individual containers capped with foam plugs (Greiner Bio-One), which contain wet coconut substrate (acquired from local pet shops) to maintain the humidity and temperature. At this stage the female spiders were fed with small crickets (*Achaeta domestica* – band 2 – Livefoods direct), whereas males were still fed with flies. Once the spiders reach sexual maturity, the female was transferred to a new vial, together with a male to enable mating. Fertilized *P. tepidariorum* females can produce up to two cocoons per week for at least two months. However, to produce high quality embryos, we only collected the first five cocoons of each female.

To visualise and stage embryos, a few embryos from each cocoon were transferred to Terasaki plates (Greiner Bio-One) under Halocarbon Oil 700 (Sigma-Aldrich – H8998), which results in the otherwise opaque embryos becoming transparent. Embryos were staged according to Mittmann and Wolff (2012).

### Fixation of Embryos and Ovaries

Embryos ranging from 1 nuclei blastoderm up to stage 12 were fixed as described in AkiyamaOda and Oda (2016), with minor modifications. Embryos before stage 5 were dechorionated with commercial bleach and fixed for 2 h in a two-layer fixative mix (1 ml 37% formaldehyde + 1 ml 2 x PEMS buffer ((0.1 M PIPES, 2 mM EGTA, 1 mM MgSO4, pH 6.9)) and 2 ml heptane) in 20 ml scintillation vials. Embryos from stage 5 up to stage 12, were dechorionated in the same way, but were then covered with a shallow layer of Halocarbon Oil 700 and their vitelline membranes pierced with tungsten needles in the order to speed up the penetration of the fixative. This allows a decrease in the fixation time to 35 min, which helps to minimise autofluorescence that causes background signal in the stainings, and for easy removal of the yolk granules upon flat mounting in older stage embryos.

For the dissection of ovaries, virgin females or mated spiders which have produced a cocoon in the last 2 days, were anesthetized with $CO_2$ and the ovaries dissected in dishes containing PBS-t. The ovaries were fixed in 4% formaldehyde in PBS for 2 h.

To stop fixations, embryos and ovaries were washed in 0.1% PBS-t (PBS supplemented with 0.1% Tween-20) three times, and depending on the subsequent usage, gradually transferred to 100% MeOH (chromogenic *in situ* hybridization) or 100% EtOH (fluorescent *in situ* hybridization and immunohistochemistry).

## 2.2 Molecular Biology Procedures

**Identification and annotation of *P. tepidariorum* Hes, Sox and other genes** tBLASTN searches of the *P. tepidariorum* genome were performed with the bHLH domain of *D. melanogaster* for the HES genes (UniProtKB – P14003) and the HMG domain of mouse Sox2 (UniProtKB - P48432) at http://bioinf.uni-greifswald.de/blast/P./blast.php. Four members of the HES gene family, namely HES-1 (formerly annotated as *hairy* in McGregor *et al*, 2008A), HES-2, HES-4 and HES-7 and fourteen Sox family genes were identified in the spider transcriptome. In addition, two bHLH-containing genes annotated as *acute-scute-lethal* complex (Ash1 and Ash2) were identified.

To identify the phylogenetic relationships of *P. tepidariorum* Sox genes, the HMG domains of *Anopheles gambiae*, *Mus musculus*, *D. melanogaster*, *P. tepidariorum* and *S. mimosarum* Sox genes were aligned with ClustalW (Paese *et al*., 2018). Phylogenetic analysis was performed in RAxML, with support levels estimated by implementing the rapid bootstrap algorithm (1000 replicates) (Stamatakis *et al*., 2008), under the PROTGAMMALG model of amino acid substitution, which was identified as best fitting using a custom Perl script from the Exelixis Lab website (https://sco.h-its.org/exelixis/web/software/raxml/hands_on.html).

Putative regulatory regions of ubiquitously expressed genes were retrieved with the Neural Network Promoter Prediction (BDGP – University of California Berkeley - http://www.fruitfly.org/seq_tools/promoter.html) using a minimal promoter score of 0.9 (scores are from 0 to 1, with higher being more reliable), and the Genomatix Gene2Promoter software (http://www.genomatix.de).

### Genomic DNA, Total RNA Extraction and cDNA synthesis

Embryos ranging from stage 1 to stage 12 were carefully placed in 1.5 ml Eppendorf tubes, and both genomic DNA or total RNA was extracted with the RecoverAll Total Nucleic Acid Isolation kit (ThermoFisher Scientific) and eluted in 20 µl of $dH_2O$. 1 µg of the eluted total RNA was used to synthesize the complementary DNA with the QuantiTech Reverse Transcription kit (Qiagen).

### Gene Isolation from cDNA and genomic DNA

Fragments of genes were isolated with PCR for *in situ* hybridization probe synthesis or doublestranded RNA synthesis (all oligonucleotide sequences used are listed in Table 1). Genomic DNA was used for the isolation of putative regulatory regions. The underlined sequences for dsRNA synthesis correspond to the T7 promoter sequence, which enables in-vitro transcription of PCR fragments without the need for molecular cloning. The underlined sequences for the putative promoters are the overhangs used for Gibson Assembly cloning.

**Table 1. List of primers used in this research.**

| Oligonucleotide name | Sequence (5’ to 3’) |
|---|---|
| **In situ Hybridization probes - Cloning** | |
| **Ash1_fw** | ATG ACG ACG CTA ACA GTC TTA GAA AAT TTA |
| **Ash1_rv** | TTA AGA AAG CCA CGT TGT GAA GTC GA |
| **Ash2_fw** | ATG GCT TCC TTG ACG CTT CTG AAT |
| **Ash2_rv** | TCA AAA CCA ATT GGC GAA ATC CAT CAA AT |
| **Caudal_fw** | CCC ATG CGG AGT TAT GGA CA |
| **Caudal_rv** | GTC CTG GTT CTG CCT GGA TT |
| **Dfd_A_fw** | CCC CTG TAA GTT ATG GCC C |
| **Dfd_A_rv** | AGC ACT GGG TTG CTG TTT CT |
| **Dpp_fw** | ATG CGC CAG CGC ATT TGG GCT |
| **Dpp_rv** | ACG GCA ACC ACA TCC TTC AAC AAC |
| **Delta_fw** | CTG TCG TTT GGG TTG GCA AG |

| | |
|---|---|
| **Delta_rv** | CCC CAT TGA GGC ATG GTT CT |
| **Engrailed_fw** | ATG ATA CCA ATG AGA ACT CGA |
| **Engrailed_rv** | CCA TTA ATT GCA ATG CCA GT |
| **Ets4_fw** | AGG TCC ACC TCC CTA TGT |
| **Ets4_rv** | ACG CTC AAC GTC ACA GGA |
| **Fkh_fw** | CAT GCC CAT GTC CCT CAA C |
| **Fkh_rv** | AAG CGT TTT TGG CGC CTT AG |
| **Hairy_fw** | AAA TAC GGC CAC AGT CAG GG |
| **Hairy_rv** | ATC CGA GCT TAT GCT CAC CG |
| **Hedgehog_fw** | GTG CCT GGC CGC ATT AGT G |
| **Hedgehog_rv** | TGA GTC ACC ATC GAA ACA TC |
| **Labial_fw** | GGA CAA CTA CGT GCA GGA CA |
| **Dichaete_fw** | AGC TGA AAC AGA CGC TCC TC |
| **Dichaete_rv** | AAG AAAGAC GCC GAA CGA GT |
| **Sox21a_fw** | CTT GGA GGC ACA GGT TTT GC |
| **Sox21a_rv** | TTT TCA GCA GGA TCC ACC CC |
| **Sox21b_fw** | GCA TAT GCT GCA CCC TCA ATC |
| **Sox21b_rv** | ACC TGC AAT TAA TGA AGC GCC |
| **SoxN_fw** | AAA GCA GCA GGA TGA GCT AC |
| **SoxN_rv** | GCT GAG AGG CTA AGG TGC AA |
| **SoxB1_fw** | CGT ATT GGT GGT GAG GTG CT |
| **SoxB1_rv** | TAT GGT TTG GGC TCA GGC AG |
| **SoxC1_fw** | CAC TGT TCG TCC ATG AGG CT |
| **SoxC1_rv** | GGT TGT TAC AGT GGG AGC GA |
| **SoxC2_fw** | TTG CAG TGG TAG GTT GAG GC TGG |
| **SoxC2_rv** | GGG AGC TCT TCG ACA TA |
| **SoxD1_fw** | GTG GAC GGG GCT TTA GTC AA |
| **SoxD1_rv** | CGT CCC CGA CGA AAA CCT AA |
| **SoxD2_fw** | CGG GGA ACA CTG ATG TTG GA |

| | |
|---|---|
| **SoxD2_rv** | GCG CTG CAC ATG AAA GAA CA |
| **SoxE1_fw** | CGG ACT CTG TGG TAC CG |
| **SoxE2_rv** | AAG CAG AAC GTC TCC GAC AG |
| **SoxF1_fw** | TTG AAG CAG CAG AGA GAG CC |
| **SoxF1_rv** | |
| **SoxF2_fw** | TGG CGA TCA CGT TAA GAG GC |
| **SoxF2_rv** | GTG GGA AAG TAC GTC GGA GG |
| **Sox21b1_fw** | AGG CCT TCA AAT CAA CGA GAC T |
| | AGC GCA GAA GTA ACT GAT GGA |
| **Sox21b1_rv** | ATG CAA GCT CCG CAA ATC GTA CAA AA |
| **Twist_fw** | TTA CAT CTG TAA TGG CAT GCC ACG |
| **Twist_rv** | ACG TTA GGA CGA ATC CAC TG |
| **Wnt8_fw** | CTG GGC TCT CTG AAC CTG |
| **Wnt8_rv** | CTA TGC AGA CAG CGT TGC TAT TG |
| **Hairy (HES-1)_fw** | AAA CGT CGA AGG GCT CGA A |
| **Hairy (HES-1)_rv** | GGT CAC TTG ATG CAG GAC T |
| **HES_2_fw** | AAA ACG AAG ACG AGC ACG C |
| **HES_2_rv** | GAA TTC ACC GTT CGG CAA C |
| **HES_4_fw** | GCG CAC GTG AAG TTG ATC TC |
| **HES_4_rv** | CAA CTC GCT GCT GAC AAT GG |
| **HES_7_fw** | TCG GCT AAA CCG ACA CCT ATG |
| **HES_7_rv** | GGC TGT TGA AGA GTC TGG CT |

**dsRNA Synthesis**

| | |
|---|---|
| **Dl_dsRNA_fw** | TAATACGACTCACTATAGG ATGTAAGCGAGTTCTGGACTCAAGACA |
| **Dl_dsRNA_rv** | TAATACGACTCACTATAGG CACGTTCCTCCATTAGAGCACGGCTTG |
| **GFP_dsRNA_fw** | TAATACGACTCACTATAGG CGTGTCCGGCGAGGGCGAGGG |
| **GFP_dsRNA_rv** | TAATACGACTCACTATAGG AGGACCATGTGATCGCGCT |

| | |
|---|---|
| **Sox21b1_dsRNA_F1_fw** | TAATACGACTCACTATAGG ATGCAAGCTCCGCAAATCGTAC |
| **Sox21b1_dsRNA_F1_rv** | TAATACGACTCACTATAGG AGAAGAGGCAGGATAGCCGC |
| **Sox21b1_dsRNA_F2_fw** | TAATACGACTCACTATAGG TCAAGTGTCTGGATCAGCAGC |
| **Sox21b1_dsRNA_F2_rv** | TAATACGACTCACTATAGG TTACATCTGTAATGGCATGCCAC |
| **Wnt8_dsRNA_fw** | TAATACGACTCACTATAGG CTATGCAGACAGCGTTGCTATTG |
| **Wnt8_dsRNA_rv** | TAATACGACTCACTATAGG GGTGAAATTTCATTGTAGATTAGCTGG |

**Putative Regulatory Region Isolation**

| | |
|---|---|
| **Act5_eGFP_fw** | TTGCTGGTTCTAGTAGTGGT ATGGTGAGCAAGGGCGAGG |
| **Act5_eGFP_rv** | TATTAAATTTTTAAAGTCAT TTACTTGTACAGCTCGTCCATGC |
| **Pub_eGFP_fw** | AAAATTTGATTTTTTTACACAGGGC ATGGTGAGCAAGGGCGAGG |
| **Pub_eGFP_rv** | CTTTACTGATTCATTTTGGTCAGCC TTACTTGTACAGCTCGTCCATGC |
| **Tub_eGFP_fw** | AAGAAAACTCTATTTCAAATCAACT ATGGTGAGCAAGGGCGAGG |
| **Tub_eGFP_rv** | TATGAACAAGACATCTTTTTGAAAA TTACTTGTACAGCTCGTCCATGC |
| **Hsp_eGFP_fw** | AAAAAATTATCTAATATACATTAAA ATGGTGAGCAAGGGCGAGG |
| **Hsp_eGFP_rv** | TCACAAAGCATAGTTCGGAATGATC TTACTTGTACAGCTCGTCCATGC |
| **Vasa_eGFP_fw** | CTTAGTTTCTCATCAATCACCTAAA ATGGTGAGCAAGGGCGAGG |
| **Vasa_eGFP_rv** | TTTTTATAGTCAATAAACAGTATCT TTACTTGTACAGCTCGTCCATGC |

For the Polymerase Chain Reaction (PCR) a HiFi Polymerase (PCR Biosystems) was used in the following mixture:

| | |
|---|---|
| 5 x PCRBio Buffer | - 10 µl |
| 10 µM Forward Primer | - 2 µl |
| 10 µM Reverse Primer | - 2 µl |
| Template DNA (100 to 500 ng) | - x |

| | |
|---|---|
| HiFi Polymerase | - 0,5 µl |
| $dH_2O$ | - to 50 µl |

PCR program:

| | | |
|---|---|---|
| 1x | | |
| 95 °C | 1 min | Initial denaturation |
| 25 to 35x | | |
| 95 °C | 15 sec | Denaturation |
| 55 °C to 65°C | 15 sec | Annealing |
| 72 °C | 30 sec/kb | Extension |
| 4 °C | indefinite | Cool down |

The final product was loaded on a 1% agarose gel, the specific band extracted and purified with the NucleoSpin Gel and PCR cleanup kit (Machery-Nagel).

### Molecular Cloning and Gibson Assembly

The purified PCR product was used for molecular cloning with the pCR4-TOPO kit (Invitrogen, Life Technologies), transformed into OneShot TOP10 chemically competent cells (Invitrogen) and plated on Luria-Bertani (LB) plates supplemented with both Kanamycin or Ampicillin. For rapid ligation into the vector without the need to use a ligase enzyme, Gibson Assembly (New England Biolabs) was used instead. An equimolar mixture of the PCR product and the digested vector in the Gibson assembly mix was incubated for 60 min at 50 °C and subsequently transformed into OneShot TOP10 cells, according to the manufacturer’s protocol.

### Colony-Check PCR and Overnight Cultures

8 to 10 colonies were picked with a 10 µl pipette tip and placed into PCR tubes containing a 10 µl OneTaq 2 x PCR MasterMix (New England Biolabs) and the products were loaded in a 1% agarose gel. Positive colonies were grown in liquid LB medium (100 µg/ml), substituted with the required antibiotic, overnight at 37 °C. Mini or Midi preparations of the plasmid were made with EZNA Plasmid Mini Kit I (VWR) following the manufacturer’s protocol and sent for Sanger sequencing (Eurofins Genomics). The sequences were later analysed with both ClustalW or in the Benchling platform.

### Antisense Labelled RNA Probe Synthesis

RNA probes were labelled with Digoxigenin (Roche) or Fluorescein (Roche) and synthesised using either T7 or T3 polymerases following the manufacturer’s protocol (Roche). The RNA probes were cleaned up using lithium chloride precipitation and resuspended in 25 µl of $dH_2O$.

### Whole Mount *in situ* Hybridization

To visualise transcript expression in *P. tepidariorum* whole mount embryos, *in situ* hybridization was performed with previously described protocols with minor modifications (Akiyama-Oda & Oda, 2003; Schönauer, 2016). Minor modifications for to the fixation process is described above, but the post-fixations and proteinase K treatment were also omitted, and antibody incubation was limited to 2 h. Fluorescent *in situ* hybridization was performed following Oda (2016). However, Tyramide Signal Amplification (TSA) was performed with TSA kits from PerkinElmer (TSA

Fluorescein and TSA Cyanine). Poly-L-lysine (Sigma-Aldrich) coated coverslips were used for flat-mounting embryos. Nuclei were stained by incubating embryos in 1 µg/ml 4-6-diamidino-2phenylindol (DAPI) in PBS with 0.1% Tween-20 for 15 min.

### Immunohistochemistry

Immunostaining was carried out following Schwager *et al.* (2015) with minor modifications: antibodies were not pre-absorbed prior to incubation and the concentration of Triton was increased to 0.1%. The following primary antibodies were used: mouse anti-α-Tubulin DM1a (Sigma) (1:50), rabbit α cleaved caspase 3 (Cell Signaling - 9661) (1:200) and rabbit Antiphospho-Histone H3 (Ser10) (Merck Millipore - 06-570). For detection the following secondary antibodies were used: donkey anti-mouse IgG Alexa Fluor 555 (Invitrogen) and goat anti-rabbit

Alexa Fluor 647 (Invitrogen). The counterstaining was carried out by incubation in 1 µg/ml 4-6diamidino-2-phenylindol (DAPI) in PBS + Triton 0,1% for 20 minutes.

### dsRNA synthesis and Parental RNA interference

Double stranded RNA (dsRNA) for parental RNA interference was synthesized according to Akiyama-Oda & Oda (2003) and injected following the standard protocol from Schönauer *et al.* (2016). Two non-overlapping fragments of *P. tepidariorum Sox21b-1* were isolated from the 1134 bp coding sequence of the gene: fragment 1 spanning 549 bp and fragment 2 covering 550 bp. Double stranded RNA for *P. tepidariorum Dl* (853 bp), *Wnt8* (714 bp) and the coding sequence of GFP (720 bp) as used previously (Akiyama-Oda & Oda, 2006), were transcribed using the same method. Synthesis of double stranded RNA was performed using the MegaScript T7 transcription kit (Invitrogen). After purification the dsRNA transcripts were annealed in a water bath starting at 95°C and slowly cooled down to room temperature. dsRNA was injected at 2.0 µg/µl in the opisthosoma of adult females every two days, with a total of five injections (n = 7 for each dsRNA;

n= 2 for GFP controls). The injected spiders were mated after the second injection and embryos from injected spiders were fixed for gene expression and phenotypic analyses at three different time points: stage 4 (cumulus formation), stage 5 late (germ disc with migrating cumulus) and stage 9 (head and limbs bud formation).

### Imaging, Live Imaging and Image Analysis

For imaging of flat-mounted embryos after *in situ* hybridisation an AxioZoom V16 stereomicroscope (Zeiss) equipped with an Axiocam 506-Mono and a colour digital camera were used. Immunostained embryos were imaged with Zeiss LSM 800 or 880 with Airyscan confocal microscopes. For live imaging, embryos were aligned on heptane glue coated coverslips and submersed in a thin layer of halocarbon oil. Bright-field live imaging was performed using an AxioZoom V16 stereomicroscope, while fluorescence live imaging was performed with confocal microscopes. Image stacks were processed in Fiji (Schindelin *et al.,* 2012) and Helicon Focus (HeliconSoft). Image brightness and intensity was adjusted in Corel PhotoPaint X5 (CorelDraw) and Fiji.

# References

**Abzhanov, A., Popadic, A., Kaufman, T. C.** (2001). Chelicerate Hox genes and the homology of arthropod segments. *Evol & Dev* **1**, 77-89.

**Akam, M.** (1989). Drosophila development: making stripes inelegantly. *Nature* **28**, 2823.

**Akam, M. E., Martinez-Arias, A.** (1985). The distribution of Ultrabithorax transcripts in Drosophila embryos. *EMBO J* **4**, 1689-700.

**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**, 1735-47.

**Akiyama-Oda, Y., Oda, H.** (2006). Axis specification in the spider embryo: dpp is required for radial-to-axial symmetry transformation and sog for ventral patterning. *Development* **133**, 2347-57.

**Akiyama-Oda, Y., Oda, H.** (2016). Multi-color FISH facilitates analysis of cell-type diversification and developmental gene regulation in the P. spider embryo. *Dev Growth Differ* **58**, 215-224.

**Aleksic, J., Ferrero, E., Fischer, B., Shen, S. P., Russell, S.** (2013). The role of Dichaete in transcriptional regulation during Drosophila embryonic development. *BMC Genomics* **14**, 861.

**Amin, J., Ananthan, J., Voellmy, R.** (1988). Key features of heat shock regulatory elements. *Mol Cell Biol* **8**, 3761-9.

**Anderson, D. T.** (1973). Embryology and Phylogeny in Annelids and Arthropods, **1**, 1510.

**Anderson, M. A., Gross, T. L., Myles, K. M., Adelman, Z. N.** (2010). Validation of novel promoter sequences derived from two endogenous ubiquitin genes in transgenic Aedes aegypti*. Insect Mol Biol* **19**, 441-9.

**Aranda, M., Marques-Souza, H., Bayer, T., Tautz, D.** (2008). The role of the segmentation gene hairy in Tribolium. *Dev Genes Evol* **218**, 465–477.

**Baker, N. E., Brown, N. L.** (2018). All in the family: proneural bHLH genes and neuronal diversity. *Development*, **145**.

**Balavoine, G.** (2014). Segment formation in Annelids: patterns, processes and evolution. *Int J Dev Biol* 6, 469-83.

**Bao, Y., Xu, F., Shimeld, S. M.** (2017). Phylogenetics of Lophotrochozoan bHLH Genes and the Evolution of Lineage-Specific Gene Duplicates. *Genome Biol Evol* **9**, 869886.

**Bassett, A. R., Liu, J. L.** (2014). CRISPR/Cas9 and genome editing in Drosophila. *J Genet Genomics* **41**, 7-19.

**Bell, D. M., Leung, K. K., Wheatley, S. C., Ng, L. J., Zhou, S., Ling, K. W., Sham, M. H., Koopman, P., Tam, P. P., Cheah, K. S.** (1997). SOX9 directly regulates the type-II collagen gene. *Nat Genet* **16**, 174-178.

**Bergsland, M., Ramskold, D., Zaouter, C., Klum, S., Sandberg, R., Muhr, J.** (2011). Sequentially acting Sox transcription factors in neural lineage development. *Genes and Development* **25**, 2453-2464.

**Bier, E.; Vaessin, H.; Younger-Sheperd, S.; Jan, L. Y.; Jan, Y. N.** (1992). deadpan, an essential pan-neural gene in Drosophila, encodes a helix-loop-helix protein similar to the hairy gene product. *Genes Dev* **6**, 2137-51.

**Boutros, M., Ahringer, J.** (2008). The art and design of genetic screens: RNA interference. *Nat Rev Genet* **9**, 554-66.

**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**, 239-255.

**Brena, C., Akam, M.** (2012) The embryonic development of the centipede Strigamia maritima. *Dev Biol* **363**, 290-307.

**Buenrostro, J. D., Wu, B., Chang, H. Y., Greenleaf, W. J.** (2015). ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Current Protocols in Molecular Biology* **21**, 1–21.

**Buescher, M., Hing, F. S., Chia, W.** (2002). Formation of neuroblasts in the embryonic central nervous system of Drosophila melanogaster is controlled by SoxNeuro *Development* **129**, 4193-4203.

**Burgio, G.** (2018). Redefining mouse transgenesis with CRISPR/Cas9 genome editing technology. *Genome Biol* **19**, 27.

**Capecchi, M. R.** (2005). Gene targeting in mice: functional analysis of the mammalian genome for the twenty-first century. *Nat Rev Genet* **6**, 507-512.

**Carroll, S. B., Laughon, A., Thalley, B. S.** (1988). Expression, function, and regulation of the hairy segmentation protein in the Drosophila embryo. *Genes Dev* **2**, 883-90.

**Chang, W. H., Lai, A. G.** (2018). Genome-wide analyses of the bHLH superfamily in crustaceans: reappraisal of higher-order groupings and evidence for lineagespecific duplications. *R Soc Open Sci* **5**, 172433.

**Chaw, R. C., Vance, E., Black, S. D.** (2007). Gastrulation in the spider Zygiella x-notata involves three distinct phases of cell internalization. *Dev Dyn* **236**, 3484-95.

**Chipman, A. D., Akam, M.** (2008). The segmentation cascade in the centipede Strigamia maritima: involvement of the Notch pathway and pair-rule gene homologues. *Dev Biol* **319**, 160-9.

**Cho, J. H., Klein, W. H., Tsai, M. J.** (2007). Compensational regulation of bHLH transcription factors in the postnatal development of BETA2/NeuroD1-null retina. *Mech Dev* **124**, 543-50.

**Choe, C. P. and Brown, S. J.** (2007). Evolutionary flexibility of pair-rule patterning revealed by functional analysis of secondary pair-rule genes, paired and sloppypaired in the short-germ insect, Tribolium castaneum. *Dev Biol* **302**, 281-294.

**Choe, C. P., Miller, S. C. and Brown, S. J.** (2006). A pair-rule gene circuit defines segments sequentially in the short-germ insect Tribolium castaneum. *Proc Natl Acad Sci U S A* **103**, 6560-6564.

**Chung, Y. T., Keller, E. B.** (1990). Regulatory elements mediating transcription from the Drosophila melanogaster actin 5C proximal promoter. *Mol Cell Biol* **10**, 206–216.

**Chylinski, K., Jinek, M., Fonfara, I., Hauer, M., Doudna, J. A., Charpentier, E.** (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816-821.

**Cinquin, O.** (2007). Understanding the somitogenesis clock: what's missing? *Mech Dev* **124**, 501-17.

**Claparede, E.** (1862). Recherches sur l'evolution des Araignees.

**Clark, E., Akam, M.** (2016). Odd-paired controls frequency doubling in Drosophila segmentation by altering the pair-rule gene regulatory network. *eLife* **15**, e18215.

**Clark, E., Peel, A. D.** (2018). Evidence for the temporal regulation of insect segmentation by a conserved sequence of transcription factors. *Development* **145**, 155580.

**Cooke, J., Zeeman, C.** (1976). A clock and wavefront model for control of the number of repeated structures during animal morphogenesis. *Journal of Theoretical Biology* **58**, 455–476.

**Cramer, P.** (2004). RNA polymerase II (Pol II) is the central enzyme that catalyses DNAdirected mRNA synthesis during the transcription of protein-coding genes. *Advances in Protein Chemistry*.

---

**Cremazy, F., Berta, P., Girard, F.** (2000). SoxNeuro, a new Drosophila Sox gene expressed in the developing central nervous system. *Mech Dev* **93**, 215-219.

**Cremazy, F., Berta, P., Girard, F.** (2001). Genome-wide analysis of Sox genes in Drosophila melanogaster. *Mech Dev* **109**, 371-375.

**Cui, Y.** (2005). Hairy is a cell context signal controlling Notch activity. *Dev Growth Differ* **47**, 609-25.

**Cummins, D. M., Tyack, S. G., Doran, T. J.** (2011). Characterisation and comparison of the chicken H1 RNA polymerase III promoter for short hairpin RNA expression. *Biochem Biophys Res Commun* **416**, 194-8.

**Damen, W. G.** (2002). Parasegmental organization of the spider embryo implies that the parasegment is an evolutionary conserved entity in arthropod embryogenesis. *Development* **129,** 1239-50.

**Damen, W. G., Tautz, D.** (1999). Abdominal-B expression in a spider suggests a general role for Abdominal-B in specifying the genital structure. *J Exp Zool* **15**, 85-91.

**Damen, W. G., Weller, M., Tautz, D.** (2000). Expression patterns of hairy, even-skipped, and runt in the spider Cupiennius salei imply that these genes were segmentation genes in a basal arthropod. *Proc Natl Acad Sci U S A* **25**, 4515-9.

**Damen, W. G.** (2007). Evolutionary conservation and divergence of the segmentation process in arthropods. *Dev Dyn* **236**, 1379-91.

**Davis, G. K. and Patel, N. H.** (1999). The origin and evolution of segmentation. *Trends Cell Biol* **9**, 68-72.

**Dearden, P. K.** (2015). Origin and evolution of the enhancer of split complex. BMC *Genomics* **16**, 712.

**Dray, N., Tessmar-Raible, K., Le Gouar, M., Vibert, L., Christodoulou, F., Schipany, K., Guillou, A., Zantke, J., Snyman, H., Behague, J., et al.** (2010). Hedgehog signaling regulates segment formation in the annelid Platynereis. *Science* **329**, 339-342.

**Feitosa, N. M., Pechmann, M., Schwager, E. E., et al** (2017) Molecular control of gut formation in the spider P. tepidariorum. *genesis* **55**: 23-33.

**Focareta, L, Cole, A. G.** (2016). Analyses of Sox-B and Sox-E Family Genes in the Cephalopod Sepia officinalis: Revealing the Conserved and the Unusual. *PLoS ONE* **11**, e0157821.

**Frasch, M., Hoey, T., Rushlow, C., Doyle, H. and Levine, M.** (1987). Characterization and localization of the even-skipped protein of Drosophila. *EMBO J* **6**, 749-759.

**Frohnhöfer, H. G., Lehmann, R., Nüsslein-Volhard, C.** (1986). Manipulating the anteroposterior pattern of the Drosophila embryo. *J Embryol Exp Morphol* **97**, 16979.

**Gao, Z., Harwig, A., Berkhout, B., Herrera-Carrillo, E.** (2017). Mutation of nucleotides around the +1 position of type 3 polymerase III promoters: The effect on transcriptional activity and start site usage. *Transcription* **8**, 275-287.

**Garbarino, J. E., Oosumi, T., Belknap, W. R.** (1995). Isolation of a polyubiquitin promoter and its expression in transgenic potato plants. *Plant Physiol* **109**, 13718.

**Gazave, E., Guillou, A. and Balavoine, G.** (2014). History of a prolific family: the Hes/Hey-related genes of the annelid Platynereis. *Evodevo* **5**, 29.

**Gibb, S., Maroto, M., Dale, J, K.** (2010). The segmentation clock mechanism moves up a notch. *Trends Cell Biol* **20,** 593–600.

**Gilles, A. F., Schinko, J. B. and Averof, M.** (2015). Efficient CRISPRmediated gene targeting and transgene replacement in the beetle Tribolium castaneum. *Development* **142**, 2832-2839.

**Gloster, A., Wu, W., Speelman, A., Weiss, S., Causing, C., Pozniak, C., Reynolds, B., Chang, E., Toma, J. G., Miller, F. D.** (1994). The T alpha 1 alpha-tubulin promoter specifies gene expression as a function of neuronal growth and regeneration in transgenic mice. *J Neurosci* **14**, 7319-30.

**Hannibal, R. L., Patel, N.** (2003). What is a segment? *EvoDevo* **4**, 35.

**Hartenstein, V., Stollewerk, A.** (2015). The evolution of early neurogenesis. *Dev Cell* **23**, 390-407.

**Haubensak, W., Attardo, A., Denk, W., Huttner.** (2004). Neurons arise in the basal neuroepithelium of the early mammalian telencephalon: a major site of neurogenesis. *Proc Natl Acad Sci U S A* **101**, 3196-201.

**Heenan, P., Zondag, L., Wilson, M. J.** (2016). Evolution of the Sox gene family within the chordate phylum. *Gene* **575**, 385-392.

**Holm, Å.** (1952). Experimentelle Untersuchungen über die Entwicklung und Entwicklungsphysiologie des Spinnenembryos. *Zool BiDr* **29**, 293–424.

**Holmberg, J., et al.** (2008) SoxB1 transcription factors and Notch signaling use distinct mechanisms to regulate proneural gene function and neural progenitor differentiation. *Development* **135**, 1843-1851.

**Huang, Y., Wang, Y., Zeng, B., Liu, Z. et al** (2017). Functional characterization of Pol III U6 promoters for gene knockdown and knockout in Plutella xylostella. *Insect Biochem Mol Biol* **89**, 71-78.

**Huang, Y., Wang, Y., Zeng, B., Liu, Z., Xu, X., Meng, Q., Huang, Y., Yang, G., Vasseur, L., Gurr, G. M., You, M.** (2017). Functional characterization of Pol III U6 promoters for gene knockdown and knockout in Plutella xylostella. *Insect Biochem Mol Biol* **89**, 71-78.

**Hughes, C. L., Kaufman, T. C.** (2002). Exploring myriapod segmentation: The expression patterns of even-skipped, engrailed, and wingless in a centipede. *Dev Biol* **247**, 47–61.

**Hülskamp, M., Tautz, D.** (1991). Gap genes and gradients--the logic behind the gaps. *Bioessays* **13**, 261-8.

**Hunnekuhl, V. S., Akam, M.** (2017). Formation and subdivision of the head field in the centipede Strigamia maritima, as revealed by the expression of head gap gene orthologues and hedgehog dynamics. *Evodevo* **23**, 18.

**Ihry, R. J., Worringer, K. A., Salick, M. R., Frias, E., Ho, D., Theriault, K., Kommineni, S., Chen, J., Sondey, M., Ye, C., Randhawa, R., Kulkarni, T., Yang, Z., McAllister, G., Russ, C., Reece-Hoyes, J., Forrester, W., Hoffman, G. R., Dolmetsch, R., Kaykas, A.** (2018). p53 inhibits CRISPR-Cas9 engineering in human pluripotent stem cells. *Nat Med*.

**Ingham, P.** (1988). Interactions between the pair-rule genes runt, hairy, evenskipped and fushi tarazu and the establishment of periodic pattern in the Drosophila embryo. *Development* **104**, 51-60.

**Ingham, P. W.** (1985). A clonal analysis of the requirement for the trithorax gene in the diversification of segments in Drosophila. *J Embryol Exp Morphol* **89**, 349-65.

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

**Janssen, R., Andersson, E. S. B., Fowler, W., Höök, L., Leyher, J., Landström, E., 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.

**Janssen, R., Prpic, N. M., Damen, W. G. M.** (2004) Gene expression suggests decoupled dorsal and ventral segmentation in the millipede Glomeris marginata (Myriapoda: Diplopoda). *Dev Biol* **269**, 84-104.

**Jędrzejowska, I., Kubrakiewicz, I. J.** (2007). The Balbiani body in the oocytes of a common cellar spider, Pholcus phalangioides (Araneae: Pholcidae). *Arthropod Struct Dev* **36**, 317-326.

**Jennings, B. H., Tyler, D. M., Bray, S. J.** (1999). Target specificities of Drosophila enhancer of split basic helix-loop-helix proteins. *Mol Cell Biol* **19**, 4600-10.

**Jennings, B. H., Wainwright, M., Ish-Horowicz, D.** (2008). Differential in vivo requirements for oligomerization during Groucho-mediated repression. *EMBO Rep* **9**, 76-83.

**Kageyama, R., Masamizu, Y., Niwa, Y.** (2007). Oscillator mechanism of Notch pathway in the segmentation clock. *Dev Dyn* **236**, 1403-9.

**Kageyama, R., Niwa, Y., Isomura, A., Gonzalez, A. and Harima, Y.** (2012). Oscillatory gene expression and somitogenesis. *Wiley Interdiscip Rev Dev Biol* **1**, 629-641.

**Kainz, F., Ewen-Campen, B., Akam, M., Extavour, C. G.** (2011). Notch/Delta signalling is not required for segment generation in the basally branching insect Gryllus bimaculatus. *Development* **138**, 5015-26.

**Kamachi, Y., Kondoh, H.** (2013). Sox proteins: regulators of cell fate specification and differentiation. *Development* **140**, 4129-4144.

**Kanayama, M., Akiyama-Oda, Y. and Oda, H.** (2010). Early embryonic development in the spider Achaearanea tepidariorum: Microinjection verifies that cellularization is complete before the blastoderm stage. *Arthropod Struct Dev* **39**, 436-445.

**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**, 500.

**Kim, D. H., Rossi, J. J.** (2008). RNAi mechanisms and applications. *Biotechniques* **44**, 613-16.

**King, N., et al.,** (2008). The genome of the choanoflagellate Monosiga brevicollis and the origin of metazoans. *Nature* **451**, 783-788.

**Kingsley, J. S.** (1892). The embryology of Limulus. *J. Morphol* **7**, 35-68.

**Knipple, D. C., Seifert, E., Rosenberg, U. B., Preiss, A., Jäckle, H.** (1985). Spatial and temporal patterns of Krüppel gene expression in early Drosophila embryos. *Nature* **317**, 40-4.

**Koch, F., et al** (2017) Antagonistic Activities of Sox2 and Brachyury Control the Fate Choice of Neuro-Mesodermal Progenitors. *Dev Cell* **42**, 514-526.

**Kondo, A.** (1969) The fine structures of the early spider embryo. *Sci Rep Tokyo Kyoiku Daigaku Sec B* **207**, 47-67.

**Kormish, J. D., Sinner, D., Zorn, A. M.** (2010) Interactions between SOX factors and Wnt/β-catenin signaling in development and disease. *Dev Dyn* **239**, 56-68.

**Kornberg, T., Siden, I., O'Farrell, P. and Simon, M.** (1985). The engrailed locus of Drosophila: in situ localization of transcripts reveals compartment-specific expression. *Cell* **40**, 45-53.

**Krøvel, A. V., Olsen, L. C.** (2002). Expression of a vas::EGFP transgene in primordial germ cells of the zebrafish. *Mech Dev* **116**, 141-50.

**Lassiter, R. N., Stark, M. R., Zhao, T., Zhou, C. J.** (2014). Signalling mechanisms controlling cranial placode neurogenesis and delamination. *Dev Biol* **389**, 39-49.

**Ledent, V., Vervoort, M.** (2001). The basic helix-loop-helix protein family: comparative genomics and phylogenetic analysis. *Genome Res* **11**, 754-70.

**Lefebvre, V.** (2010). The SoxD transcription factors – Sox5, Sox6, and Sox13 – are key cell fate modulators. *Int J Biochem Cell Biol* **42**, 429–432.

**Leimeister, C., Externbrink, A., Klamt, B., Gessler, M.** (1999). Hey genes: a novel subfamily of hairy- and Enhancer of split related genes specifically expressed during mouse embryogenesis. *Mech Dev* **85**, 173-7.

**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., McGregor, A. P.** (2018). Homeobox gene duplication and divergence in arachnids. *Mol Biol Evol*.

**Levine, M.** (2008). A systems view of Drosophila segmentation. *Genome Biol* **9**, 207.

**Lewis, J.** (1996). Neurogenic genes and vertebrate neurogenesis. *Curr Opin Neurobiol* **6**, 3-10.

**Li, A., Hooli, B., Mullin, K., Tate, R., Bubnys, A., Kirchner, R., Chapman, B., Hofmann, O., Hide, W., Tanzi, R. E.** (2017). Silencing of the Drosophila ortholog of Sox5 leads to abnormal neuronal development and behavioural impairment. In: *Hum Mol Gen,* 1-36.

**Liu, P. Z. and Kaufman, T. C.** (2004a). hunchback is required for suppression of abdominal identity, and for proper germband growth and segmentation in the intermediate germband insect Oncopeltus fasciatus. *Development* **131**, 15151527.

**Liu, P. Z. and Kaufman, T. C.** (2004b). Kruppel is a gap gene in the intermediate germband insect Oncopeltus fasciatus and is required for development of both blastoderm and germband-derived segments. *Development* **131**, 4567-4579.

**Liu, P. Z. and Kaufman, T. C.** (2005). even-skipped is not a pair-rule gene but has segmental and gaplike functions in Oncopeltus fasciatus, an intermediate germband insect. *Development* **132**, 2081-2092.

**Liu, P. Z., Kaufman, T. C.** (2005). Short and long germ segmentation: unanswered questions in the evolution of a developmental mode. *Evol Dev* **7**, 629-46.

**Loh, S. H. Y., Russell, S.** (2000). A Drosophila group E Sox gene is dynamically expressed in the embryonic alimentary canal. *Mech Dev* 93, 4.

**Ma, H., Wu, Y., Dang, Y., Choi, J., Zhang, J., Wu, H.** (2014). Pol III Promoters to Express Small RNAs: Delineation of Transcription Initiation. *Mol Ther Nucleic Acids* **3**, e161.

**Ma, S., Chang, J., Wang, X., Liu, Y., Zhang, J., Lu, W., Gao, J., Shi, R., Zhao, P., Xia, Q.** (2014). CRISPR/Cas9 mediated multiplex genome editing and heritable mutagenesis of BmKu70 in Bombyx mori. *Sci Rep* **27**, 4489.

**Maekawa, T., Kusakabe, M., Shimoda, Y., Sato, S., Tabata, S., Murooka, Y., Hayashi, M.** (2008). Polyubiquitin promoter-based binary vectors for overexpression and gene silencing in Lotus japonicus. *Mol Plant Microbe Interact* **21**, 375-82.

**Magie, C. R., Pang, K., Martindale, M. Q.** (2005). Genomic inventory and expression of Sox and Fox genes in the cnidarian Nematostella vectensis. *Dev Genes Evol* **215**, 618-30.

**Martin, A., Serano, J. M., Jarvis, E., Bruce, H. S., Wang, J., Ray, S., Barker, C. A., O'Connell, L. C., Patel, N. H.** (2016). CRISPR/Cas9 Mutagenesis Reveals Versatile Roles of Hox Genes in Crustacean Limb Specification and Evolution. *Curr Biol* **26**, 14-26.

**Massari, M. E., Murre, C.** (2000). Helix-loop-helix proteins: regulators of transcription in eucaryotic organisms. *Mol Cell Biol* **20**, 429–40.

**McDonald, P. M., Struhl, G.** (1986). A molecular gradient in early Drosophila embryos and its role in specifying the body pattern. *Nature* **11**, 537-45.

**McGregor, A. P., Hilbrant, M., Pechmann, M., Schwager, E. E., Prpic, N. M., Damen, W. G.** (2008B). Cupiennius salei and Achaearanea tepidariorum: Spider models for investigating evolution and development. *Bioessays* **30**, 487-98.

**McGregor, A. P., Pechmann, M., Schwager, E. E., Feitosa, N. M., Kruck, S., Aranda, M., Damen, W. G.** (2008). Wnt8 is required for growth-zone establishment and development of opisthosomal segments in a spider. *Curr Biol* **18**, 1619-23.

**McGregor, A. P., Pechmann, M., Schwager, E. E., Feitosa, N. M., Kruck, S., Aranda, M., Damen, W. G.** (2008A). Wnt8 is required for growth-zone establishment and development of opisthosomal segments in a spider. *Curr Biol* **28**, 1619-23.

**McGregor. A. P., Pechmann, M., Schwager, E. E., Damen, W. G.** (2009) An ancestral regulatory network for posterior development in arthropods. *Commun Integr Biol* **2**, 174-176.

**McKimmie, C., Woerfel, G., Russell, S.** (2005). Conserved genomic organisation of group B Sox genes in insects. *BMC Genetics* **6**, 21-26.

**Miller, S. W., Avidor-Reiss, T., Polyanovsky, A., Posakony, J. W.** (2009). Complex interplay of three transcription factors in controlling the tormogen differentiation program of Drosophila mechanoreceptors. *Dev Biol*, 1-14.

**Mito, T., Sarashina, I., Zhang, H., Iwahashi, A., Okamoto, H., Miyawaki, K., Shinmyo, Y., Ohuchi, H. and Noji, S.** (2005). Non-canonical functions of hunchback in segment patterning of the intermediate germ cricket Gryllus bimaculatus. *Development* **132**, 2069-2079.

**Mittmann, B., Wolff, C.** (2012). Embryonic development and staging of the cobweb spider P. tepidariorum C. L. Koch, 1841 (syn.: Achaearanea tepidariorum, Araneomorphae, Theridiidae). *Dev Genes Evol* **222**, 189-216.

**Mlodzik, M. and Gehring, W. J.** (1987). Expression of the caudal gene in the germ line of Drosophila: formation of an RNA and protein gradient during early embryogenesis. *Cell* **48**, 465-478.

**Montgomery, T.** (1909). The development of theridium, an aranead, up to the stage of reversion. *J Morphol* **20**, 297-352.

**Moore, J. K., Haber, J. E.** (1996). Cell cycle and genetic requirements of two pathways of nonhomologous end-joining repair of double-strand breaks in Saccharomyces cerevisiae. *Molecular and Cellular Biology* **16**, 2164–73.

**Müller, M., Weizsäcker, E., Campos-Ortega, J. A.** (1996). Expression domains of a zebrafish homologue of the Drosophila pair-rule gene hairy correspond to primordia of alternating somites. *Development* **122**, 2071-8.

**Murre, C., Brain, G., van Dijk, M. A., Engel, I., Beth, A., Furnari, M. E., Massari, J. R., et al.** (1994). Structure and function of helix-loop-helix proteins. Bba Gene *Struct Expr* **1218**, 129-135.

**Nakao, H.** (2018) A Bombyx homolog of ovo is a segmentation gene that acts downstream of Bm-wnt1(Bombyx wnt1 homolog). *Gene Expression Patterns* **27**, 1-7.

**Nambu, P., Nambu, J.** (1996). The Drosophila fishhook gene encodes a HMG domain protein essential for segmentation and CNS development. *Development* **122**, 3467-3475.

**Nanda, S., Defalco, T., Hui, Yong., Loh, S., Phochanukul, N., Camara, N., Van Doren, M., Russell, S.** (2009) Sox100B, a Drosophila Group E Sox-domain Gene, Is Required for Somatic Testis Differentiation. *Sex Dev* **3**, 26-37.

**Newman, S. A., Forgacs, G.** (2007). Complexity and Self-Organization in Biological Development and Evolution. in Complexity in Chemistry, Biology, and Ecology **1**, 1.

**Ng, S. Y., Gunning, P., Liu, S. H., Leavitt, J., Kedes, L.** (1989). Regulation of the human beta-actin promoter by upstream and intron domains. *Nucleic Acids Res* **17**, 601–615.

**Nishimura, M., Isaka, F., Ishibashi, M., Tomita, K., Tsuda, H., Nakanishi, S. and Kageyama, R.** (1998). Structure, chromosomal locus and promoter of mouse Hes2 gene, a homologue of Drosophila hairy and Enhancer of split. *Genomics* **49**, 69-75.

**Niwa, H., Nakamura, A., Urata, M., Shirae-Kurabayashi, M., Kuraku, S., Russell, S., Ohtsuka, S.** (2016). The evolutionally-conserved function of group B1 Sox family members confers the unique role of Sox2 in mouse ES cells. *BMC Evolutionary Biology* **16**, e173.

**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.

**Overton, P., Meadows, L., Urban, J., Russell, S.** (2002). Evidence for differential and redundant function of the Sox genes Dichaete and SoxN during CNS development in Drosophila. *Development* **129**, 4219-4228.

**Paese, C. L. B., Leite, D. J., Schönauer, A., McGregor, A. P., Russell, S.** (2017). Duplication and divergence of Sox genes in spiders. *bioRxiv*.

**Paese, C. L. B., Schönauer, A., Leite, D. J., Russell, S.**, **McGregor, A. P.** (2018). A SoxB gene acts as an anterior gap gene and regulates posterior segment addition in the spider *P. tepidariorum*. *bioRxiv*.

**Palmeirim, I., Henrique, D., Ish-Horowicz, D., Pourquie, O.** (1997). Avian hairy gene expression identifies a molecular clock linked to vertebrate segmentation and somitogenesis. *Cell* **28**, 639-48.

**Patel, N. H.** (2003). The ancestry of segmentation. *Dev Cell* **5**, 2-4.

**Patel, N. H., Condron, B. G. and Zinn, K.** (1994). Pair-rule expression patterns of evenskipped are found in both short- and long-germ beetles. *Nature* **367**, 429-434.

**Patel, N. H., Hayward, D. C., Lall, S., Pirk, N. R., DiPietro, D., Ball, E. E.** (2001). Grasshopper hunchback expression reveals conserved and novel aspects of axis formation and segmentation. *Development* **128**, 3459- 3472.

**Patten, W.** (1896). Variations in the development of Limulus polyphemus. *J Morphol* **12**, 17-148.

**Pavlopoulos, A., Kontarakis, Z., Liubicich, D. M., Serano, J. M., Akam, M., Patel, N. H., Averof, M.** (2009). Probing the evolution of appendage specialization by Hox gene misexpression in an emerging model crustacean*. Proc Natl Acad Sci U S A* **106**, 13897–13902.

**Pechmann, M.** (2016). Formation of the germ-disc in spider embryos by a condensationlike mechanism. *Front in Zool* **13**, 35.

**Pechmann, M., et al.** (2011) Novel function of Distal-less as a gap gene during spider segmentation. *PLoS Genet* **7**, e1002342.

**Pechmann, M., et al.** (2017) A novel role for Ets4 in axis specification and cell migration in the spider P. tepidariorum. *eLife* **6**, 1-20.

**Pechmann, M., McGregor, A. P., Schwager, E. E., Feitosa, N. M., Damen, W. G.** (2009). Dynamic gene expression is required for anterior regionalization in a spider. *Proc Natl Acad Sci U S A* **106**, 1468-72.

**Peel, A. D., Chipman, A. D., Akam, M.** (2005). Arthropod segmentation: beyond the Drosophila paradigm. *Nat Rev Genet* **6**, 905-16.

**Peel, A., Akam, M.** (2003). Evolution of Segmentation: Rolling Back the Clock. *Cell* **5**, 24.

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

**Pinkerton, A. C., Michel, K., O'Brochta, D. A., Atkinson, P. W.** (2000). Green fluorescent protein as a genetic marker in transgenic Aedes aegypti*. Insect Mol Biol* **9**, 1-10.

**Posnien, N., Zeng, V., Schwager, E. E., Pechmann, M., Hilbrant, M., Keefe, J. D., Damen, W. G., Prpic, N. M., McGregor, A. P., Extavour, C. G.** (2014). A comprehensive reference transcriptome resource for the common house spider P. tepidariorum. *PLoS One* **9**, e104885.

**Pourquié, O.** (2000). Segmentation of the paraxial mesoderm and vertebrate somitogenesis. *Curr Top Dev Biol* **47**, 81-105.

**Pourquié, O.** (2003). The segmentation clock: converting embryonic time into spatial pattern. *Science* **18**, 328-30.

**Pourquié, O.** (2003). The segmentation clock: converting embryonic time into spatial pattern. *Science* **301**, 328-30.

**Pourquie, O.** (2007). Building the spine: the vertebrate segmentation clock. *Cold Spring Harb Symp Quant Biol* **72**, 445-9.

**Prpic, N. M., Janssen, R., Wigand, B., Klinger, M., Damen, W. G.** (2003). Gene expression in spider appendages reveals reversal of exd/hth spatial specificity, altered leg gap gene dynamics, and suggests divergent distal morphogen signaling. *Dev Biol* **264**, 119-40.

**Pueyo, J. I., Lanfear, R., Couso, J. P.** (2008). Ancestral Notch-mediated segmentation revealed in the cockroach Periplaneta americana. *Proc Natl Acad Sci U S A* **105**, 16614-9.

**Quitschke, W. W., Lin, Z. Y., DePonti-Zilli, L., Paterson, B. M.** (1990). The beta actin promoter. High levels of transcription depend upon a CCAAT binding factor. *The Journal of Biological Chemistry* **264**, 9539-9546.

**Reiprich, S., Wegner, M.** (2015). From CNS stem cells to neurons and glia: Sox for everyone. *Cell and tissue research* **359**, 111-124.

**Ren, X., Sun, J., Housden, B. E., Hu, Y., Roesel, C., Lin, S., Liu, L. P., Yang, Z., Mao, D., Sun, L., Wu, Q., Ji, J. Y., Xi, J., Mohr, S. E., Xu, J., Perrimon, N., Ni, J. Q.** (2013). Optimized gene editing technology for Drosophila melanogaster using germ line-specific Cas9. *Proc Natl Acad Sci U S A* **110**, 19012-7.

**Rieckher, M., Kourtis, N., Pasparaki, A., Tavernarakis, N.** (2009). Transgenesis in Caenorhabditis elegans. *Methods Mol Biol* **561**, 21-39.

**Ringrose, L.** (2009). Transgenesis in Drosophila melanogaster. *Methods Mol Biol* **561**, 3-19.

**Ritter, A. R., Beckstead, R. B.** (2010). Sox14 is required for transcriptional and developmental responses to 20-hydroxyecdysone at the onset of drosophila metamorphosis. *Dev Dyn* **239**, 2685-2694.

**Roth, S., Stein, D., Nüsslein-Volhard, C.** (1989). A gradient of nuclear localization of the dorsal protein determines dorsoventral pattern in the Drosophila embryo. *Cell* **22**, 1189-202.

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

**Sagner, A., Gaber, Z. B., Delile, J., Kong, J. H., Rousso, D. L., Pearson, C. A. et al.** (2018). Olig2 and Hes regulatory dynamics during motor neuron differentiation revealed by single cell transcriptomics. *PLoS Biol* **16**, e2003127.

**Sanchez-Soriano, N., Russell, S.** (1998) The Drosophila Sox-domain protein Dichaete is required for the development of the central nervous system midline. *Development* **125**, 3989-3996.

**Sanggaard, K, W., Bechsgaard, J. S., Fang, X., Duan, J., Dyrlund, T. F., Gupta, V., Jiang, X., Cheng, L., Fan, D., Feng, Y. *et al*** (2014). Spider genomes provide insight into composition and evolution of venom and silk. *Nat Commun* **5**, 3765.

**Sanson, B.** (2001). Generating patterns from fields of cells. Examples from Drosophila segmentation. *EMBO Rep* **2**, 1083-8.

**Sarkar, A., Hochedlinger, K.** (2013). The sox family of transcription factors: versatile regulators of stem and progenitor cell fate. *Cell stem cell* **12**, 15-30.

**Schindelin J, et al.** (2012) Fiji: an open-source platform for biological-image analysis.

*Nature Methods* **9**, 676–682.

**Schinko, J. B., Hillebrand, K., Bucher, G.** (2012). Heat shock-mediated misexpression of genes in the beetle Tribolium castaneum. *Dev Genes Evol* **222**, 287-98.

**Schnitzler, C. E., Simmons, D. K., Pang, K., Martindale, M. Q., Baxevanis, A. D.** (2014). Expression of multiple Sox genes through embryonic development in the ctenophore Mnemiopsis leidyi is spatially restricted to zones of cell proliferation. *Evodevo* **5**, 15.

**Scholtz, G.** (2002). The Articulata hypothesis – or what is a segment? *Org Divers Evol* **2**, 197-215.

**Schönauer, A., Paese, C. L., Hilbrant, M., Leite, D., Schwager, E. E., Feitosa, N. M., Eibner, C., Damen, W. G., 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 P. tepidariorum. *Development* **143**, 2455-63.

**Schoppmeier, M., Damen, W. G.** (2001). Double-stranded RNA interference in the spider Cupiennius salei: the role of Distal-less is evolutionarily conserved in arthropod appendage formation. *Dev Genes Evol* **211**, 76-82.

**Schoppmeier, M., Damen, W. G.** (2005). Suppressor of Hairless and Presenilin phenotypes imply involvement of canonical Notch-signalling in segmentation of the spider Cupiennius salei. *Dev Biol* **280**, 211-24.

**Schwager EE, Meng Y, Extavour CG.** (2015). vasa and piwi are required for mitotic integrity in early embryogenesis in the spider P. tepidariorum. *Dev Biol* **402**, 276290.

**Schwager EE, Sharma PP, Clarke T, Leite DJ, Wierschin T, Pechmann M, AkiyamaOda 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.

**Schwager, E. E., Pechmann, M., Feitosa, N. M., McGregor, A. P., Damen, W. G.** (2009) hunchback functions as a segmentation gene in the spider Achaearanea tepidariorum. *Curr Biol* **19**, 1333-1340.

**Scotting, P. J., Rex, M.** (1996). Transcription factors in early development of the central nervous system. *Neuropathol Appl Neurobiol* **22**, 469-81.

**Shimizu, T., Nakamoto, A.** (2001). Segmentation in Annelids: Cellular and Molecular Basis for Metameric Body Plan. *Zool Sci* **18**, 285-298.

**Shoji, W., Sato-Maeda, M.** (2008). Application of heat shock promoter in transgenic zebrafish. *Dev Growth Differ* **50**, 401-6.

**Siebert, M., Banovic, D., Goellner, B., Aberle, H.** (2009). Drosophila motor axons recognize and follow a Sidestep-labeled substrate pathway to reach their target fields. *Genes Dev* **23**, 1052-62.

**Simões-Costa, M., Bronner, M. E.** (2015). Establishing neural crest identity: a gene regulatory recipe. *Development* **142**, 242-57.

**Simonnet, F., Deutsch, J., Queinnec, E.** (2004). Hedgehog is a segment polarity gene in a crustacean and a chelicerate. *Dev Genes Evol* 214, 537–545.

**Sinclair, A., Berta, P., Palmer, M., Hawkins, J., Griffiths, B., Smith, M., Foster, J., Frischauf, A., Lovell-Badge, R., Goodfellow, P.** (1990). A gene from the human sex determining region encodes a protein with homology to a conserved DNA binding motif. *Nature* **346**, 240-244.

**Sommer, R. J.** (2009). The future of evo-devo: model systems and evolutionary theory. *Nat Rev Genet* **10**, 416-22.

**Sommer, R., Tautz, D.** (1991). Segmentation gene expression in the housefly Musca domestica. *Development* **113**, 419-30.

**Spemann, H., Mangold, H.** (2001). Induction of embryonic primordia by implantation of organizers from a different species. *Int J Dev Biol* **45**, 13–38.

**Stamatakis, et al.** (2008) A rapid bootstrap algorithm for the RAxML Web servers. *Syst Biol* **57**, 758-771.

**Stollewerk, A.** (2002). Recruitment of cell groups through Delta/Notch signalling during spider neurogenesis. *Development* **129**, 5339-5348.

**Stollewerk, A.** (2016). A flexible genetic toolkit for arthropod neurogenesis. *Philos Trans R Soc Lond B Biol Sci* **371**, 20150044.

**Stollewerk, A., Chipman, A. D.** (2006). Neurogenesis in myriapods and chelicerates and its importance for understanding arthropod relationships. *Integr Comp Biol* **46**, 195-206.

**Stollewerk, A., Schoppmeier, M., Damen, W. G.** (2003). Involvement of Notch and Delta genes in spider segmentation. *Nature* **423**, 863-5.

**Stollewerk, A., Weller, M., Tautz, D.** (2001). Neurogenesis in the spider Cupiennius salei. *Development* **128**, 2673-88.

**Stolt, C. C., Wegner, M.** (2010). SoxE function in vertebrate nervous system development. *Int J Biochem Cell Biol* **42**, 437–440.

**Sun, D., Zhaojiang, G., Liu, Y., Zhang, Y.** (2017). Progress and Prospects of CRISPR/Cas Systems in Insects and Other Arthropods. *Front Physiol* **8**: 608.

**Suzuki, H., Kondo, A.** (1995). Early embryonic development, including germ-disk stage, in the Theridiid spider *Achaearanea japonica* (Bös. et Str.). *J Morphol* **224**, 147–157.

**Szostak, J. W., Orr-Weaver, T. L., Rothstein, R. J., Stahl, F. W.** (1983). The doublestrand-break repair model for recombination. *Cell* **33**, 25–35.

**Takemoto, T., Uchikawa, M., Yoshida, M., Bell, D. M., Lovell-Badge, R., Papaioannou, V. E., Kondoh, H.** (2011). Tbx6-dependent Sox2 regulation determines neural or mesodermal fate in axial stem cells. *Nature* **470**, 394-8.

**Tanaka, S., Kamachi, Y., Tanouchi, A., Hamada, H., Jing, N., Kondoh, H.** (2004). Interplay of SOX and POU Factors in Regulation of the Nestin Gene in Neural Primordial Cells. *Mol Cell Biol* **24**, 8834-8846.

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

**Teschendorf, C., Warrington, K. H., Siemann, D. W., Muzyczka, N.** (2002). Comparison of the EF-1 alpha and the CMV promoter for engineering stable tumor cell lines using recombinant adeno-associated virus. *Anticancer Res* **22**, 3325-30.

**Thamm, K., Seaver, E. C.** (2008). Notch signaling during larval and juvenile development in the polychaete annelid Capitella sp. I. *Dev Biol* **320**, 304-318.

**Thomas, J. L., Da Rocha, M., Besse, A., Mauchamp, B., Chavancy, G.** (2002). 3xP3EGFP marker facilitates screening for transgenic silkworm Bombyx mori L. from the embryonic stage onwards. *Insect Biochem Mol Biol* **32**, 247-53.

**Thümecke, S., Schröder, R.** (2018). UTR-specific knockdown of Distal-less and Sp8 leads to new phenotypic variants in the flour beetle Tribolium. *Dev Genes Evol*.

**Tilgner, K., Atkinson, S. P., Yung, S., Golebiewska, A., Stojkovic, M., Moreno, R., Lako, M., Armstrong, L.** (2010). Expression of GFP under the control of the RNA helicase VASA permits fluorescence-activated cell sorting isolation of human primordial germ cells. *Stem Cells* **28**, 84-92.

**Tokushige, K., Moradpour, D., Wakita, T., Geissler, M., Hayashi, N., Wands, J. R.** (1997). Comparison between cytomegalovirus promoter and elongation factor-1 alpha promoter-driven constructs in the establishment of cell lines expressing hepatitis C virus core protein. *J Virol Methods* **64**, 73-80.

**Trainor, P. A., Tan, S.S., Tam, P. P.** (1994). Cranial paraxial mesoderm: regionalisation of cell fate and impact on craniofacial development in mouse embryos. *Development* **120**, 2397-2408.

**Vidal, V. P. I., Charboissier, M-C., deRooij, D. G., Schedl, A.** (2001). Sox9 induces testis development in XX transgenic mice. *Nat Genet* **28**, 216-217.

**Wang, J., et al.** (2018). ATAC-Seq analysis reveals a widespread decrease of chromatin accessibility in age-related macular degeneration. *Nat Comm* **9**, 1364.

**Wanninger, A., Kristof, A. and Brinkmann, N.** (2009) Sipunculans and segmentation. *Commun Integr Biol* **2**, 56-59.

**Wedeen, C. J., Shankland, M.** (1997). Mesoderm is required for the formation of a segmented endodermal cell layer in the leech Helobdella. *Dev Biol* **191**, 202–214

**Wegner, M.** (2011). SOX after SOX: SOXession regulates neurogenesis. *Genes Dev* **25**, 2423-8.

**Wei, L., Cheng D, Li D, Meng M, Peng L, Tang L, Pan M, Xiang Z, Xia Q, Lu C.** (2010). Identification and characterization of Sox genes in the silkworm, Bombyx mori. *Mol Biol Rep* **1**, 1-12.

**Weiss, B., Davidkova, G., Zhou, L. W.** (1999). Antisense RNA gene therapy for studying and modulating biological processes. *Cell Mol Life Sci* **55**, 334-58.

**Weller, M., Tautz, D.** (2003). Prospero and Snail expression during spider neurogenesis. *Dev Genes Evol* **213**, 554-66.

**Weninger, A., Hatzl, A. M., Schmid, C., Vogl, T., Glieder, A.** (2016). Combinatorial optimization of CRISPR/Cas9 expression enables precision genome engineering in the methylotrophic yeast Pichia pastoris. *J Biotechnol* **235**, 139-49.

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

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

**Yamazaki, K., Akiyama-Oda, Y., Oda, H.** (2005). Expression patterns of a twist-related gene in embryos of the spider Achaearanea tepidariorum reveal divergent aspects of mesoderm development in the fly and spider. *Zoolog Sci* **22**, 177-185.

**Zhong, L., Wang, D., Gan, X., Yang, T., He, S.** (2011). Parallel Expansions of Sox Transcription Factor Group B Predating the Diversifications of the Arthropods and Jawed Vertebrates. *PLoS ONE* **6**, e16570.

# Vita

## CHRISTIAN LOUIS BONATTO PAESE

### Education:

- B.Sc. Biological Sciences – Western Parana State University, Cascavel, Brazil – 2013.
- M.Sc. Cell and Developmental Biology – Federal University of Santa Catarina, Florianopolis, Brazil – 2015.

### Courses Attended During the Ph.D.:

- Embryology: Concepts & Techniques in Modern Developmental Biology (576h). 2016. Marine Biological Laboratory, Woods Hole, USA.
- Eco Evo Devo Summer School (40h). 2016. Oxford Brookes University, Oxford, England.
- Gene Expression Workshop (32h). 2016. University of Liverpool, Liverpool, England.
- Field Identification of Spiders and Harvestmen (8h). 2017. Field Studies Council, FSC, England.

### Honors and Awards:

- 2018 - Best poster Prize - BSDB Annual Spring Meeting - Travel Grant for the 77 SDB Meeting in Portland - USA, Society for Developmental Biology.

- 2018 - Travel Grant for the Conference - Spring Meeting of the BSDB, British Society for Developmental Biology.
- 2016 - Scholarship to attend the course Embryology: Concepts & Techniques in Modern Developmental Biology, Marine Biological Laboratory.
- 2016 - Post-Course Research Fellowship - Visitor Scientist at Nipam H. Patel's Lab (University of California - Berkeley), Marine Biological Laboratory.

**Publications During the Ph.D.:**

- **Paese, C. L. B., Schönauer, A., Leite, D. J., Russell, S.**, **McGregor, A. P.** (2018). A SoxB gene acts as an anterior gap gene and regulates posterior segment addition in the spider *P. tepidariorum*. *bioRxiv*. Submitted in May/2018.
- **Paese, C. L. B., Leite, D. J., Schönauer, A., McGregor, A. P., Russell, S.** (2018). Duplication and divergence of Sox genes in spiders. *bioRxiv*. Submitted in November/2017.
- **Schwager EE, Sharma PP, Clarke T, Leite DJ, Wierschin T, Pechmann M, Akiyama-Oda Y, Esposito L, Bechsgaard J, Bilde T et al.** (2017). The house spider genome reveals an ancient whole-genome duplication during arachnid evolution. *BMC Biol*, **15,** 62.
- **Jaramillo, M. L., Guzman, F., Paese, C. L., Margis, R., Nazari, E. M., Ammar, D., Müller, Y. M.** (2016). Exploring developmental gene toolkit and associated pathways in a potential new model crustacean using transcriptomic analysis. *Dev Genes Evol* **226**, 325-37.
- **Schönauer, A., Paese, C. L., Hilbrant, M., Leite, D., Schwager, E. E., Feitosa, N. M., Eibner, C., Damen, W. G., McGregor, A. P.** (2016). The Wnt and DeltaNotch signalling pathways interact to direct pair-rule gene expression via caudal during segment addition in the spider P. tepidariorum. *Development* **143**, 2455-63.

**Attendance in Conferences/Meetings during the Ph.D.:**

- Spring Meeting of the British Society for Developmental Biology. 2018. Warwick, England. (Poster Presentation).
- 30th European Congress of Arachnology. 2017. Nottingham, England. (Oral Presentation).
- 9th Latin American Society for Developmental Biology Meeting. 2017. Medellin, Colombia. (Oral Presentation).
- BSCB/BSDB/Genetics Society Joint Spring Meeting. 2017. Warwick, England. (Poster Presentation).
- Symposium on Animal Development and its Evolutionary variations. 2017. Cambridge, England.
- UK Evo Devo Meeting. 2017. London, England.
- BSCB/BSDB Joint Spring Meeting. 2016. (Oral Presentation).