Evolution of Sp Transcription Factors in Metazoans

Dissertation submitted in partial fulfillment of the requirements for the degree of
„doctor rerum naturalium“
of the Georg-August-University Göttingen

from

Nina Schäper

Weinheim,
Germany

Göttingen, 2009

Referent: Prof. Dr. Ernst A. Wimmer

Korreferent: Jun. Prof. Dr. Gregor Bucher

Tag der mündlichen Prüfung: 15.01.2010

Table of Contents

1 Zusammenfassung

Die Mechanismen der Kopfentwicklung bei Arthropoden sind weit weniger gut verstanden als die Mechanismen der Entwicklung des Thorax und des Abdomens, wie Daten aus der Fruchtfliege *Drosophila melanogaster* gezeigt haben. Die posterioren, gnathocephalen Kopfsegmente (Mandibular-, Maxillar-, Labialsegment) werden wie die thorakalen und abdominalen Segmente gemustert, wohingegen die Segmentierungsmechanismen der anterioren procephalen Kopfregion (Labrum, Okular-, Antennen- und Interkalarsegment) offensichtlich andere sind. Obwohl die Kopfsegmentierungsmechanismen noch nicht genau bekannt sind, konnte eine Beteiligung der so genannten Kopflückengene *orthodenticle* (*otd*), *empty spiracles* (*ems*) and *buttonhead* (*btd*) sowie von so genannten ‚second level regulators', wie *collier* (*col*), gezeigt werden.

Das aus *D. melanogaster* bekannte Kopflückengen *btd* gehört zu der Familie von Sp Transkriptionsfaktoren. Diese Zink-Finger Proteine sind evolutionär stark konserviert und in vielen unterschiedlichen Arten zu finden. Die Orthologie dieser Sp Gene unterschiedlicher Arten war jedoch unklar und ihre evolutionäre Geschichte wurde daher kontrovers diskutiert, was insbesondere für *Dm btd* gilt. Aufgrund eines ähnlichen postblastodermalen mRNA Expressionsmusters, partiell redundanter Funktion sowie einer chromosomalen Lokalisation in derselben cytogenetischen Bande des X-Chromosoms wurde postuliert, dass *btd* und *D-Sp1* aus einer rezenten Genduplikation hervorgegangen sind. Zudem war ein direktes *btd* Ortholog in Vertebraten nicht bekannt. Um den Ursprung von *btd* aufzuklären, wurden unterschiedliche Sp Gene aus verschiedenen Arthropoden isoliert. Zudem wurden bereits sequenzierte Genome verschiedener Vertreter der Metazoa nach Sp Genen durchsucht. Phylogenetische Analysen dieser Daten, sowie Proteindomänenanalysen, chromosomale Lokalisation und mRNA Expressionsanalysen zeigten, dass *btd* Orthologe schon in so basalen Metazoa wie *Nematostella vectensis* und *Trichoplax adhearens* vorhanden sind. Es hat sich gezeigt, dass ein Satz von drei Sp Genen anzestral in den Metazoa ist und der Ursprung von *btd* somit bis zum gemeinsamen Vorfahren der Metazoa zurückverfolgt werden kann. Zudem wurde der Einfluss des *D. melanogaster btd*-Zielgens *col* auf die Kopfmusterung in verschiedenen Arthropoden untersucht. In *D. melanogaster* ist *col* früh in der Entwicklung im Parasegment 0 exprimiert, welches zum posterioren Anteil des interkalaren (tritocerebralen) Segmentes sowie zum anterioren Anteil des Mandibularsegmentes beiträgt. Zudem ist eine späte Expression im Nervensystem zu verzeichnen. Das tritocerebrale Segment bei Insekten trägt keine Anhänge, wohingegen das entsprechende Segment bei Crustaceen und Cheliceraten paarige Anhänge trägt. Unsere Daten zeigen eine frühe Funktion von *col* im tritocerebralen Segment der Insekten *Tribolium castaneum* und *Oncopeltus fasciatus*. Diese frühe Kopffunktion fehlt in dem Krebs *Parhyale hawaiensis* und der Spinne *Achaearanea tepidariorum*, wo *col* nur spät im Nervensystem exprimiert ist. Dies wirft die Frage nach einer Funktion von *col* in Hinblick auf den anhanglosen Zustand des tritocerebralen Segmentes auf.

1 Summary

In contrast to the well-known trunk segmentation mechanisms in the fruit fly *Drosophila melanogaster* the mechanisms of arthropod head segmentation are less well understood. While the posterior, gnathocephalic head segments (mandibular-, maxillary-, labial segment) are metamerized like the trunk, the anterior cephalic region, the procephalon (labrum, ocular-, antennal-, intercalary segment) is patterned in a different way. However, the mechanisms for patterning the anterior head are poorly understood, but a role in cephalic formation of the head gap-like genes *orthodenticle* (*otd*), *empty spiracles* (*ems*) and *buttonhead* (*btd*) as well as the influence of so called second level regulators, such as *collier* (*col*), have been shown.

The *D. melanogaster* head gap-like gene *btd* belongs to the Sp family of transcription factors. Sp zinc finger proteins are evolutionary conserved and are present in many animal species. The orthology of the Sp genes in different species was unclear and their evolutionary history was therefore controversially discussed. Especially the origin of *Dm btd* is discussed. Due to a similar postblastodermal expression pattern, partially redundant function and chromosomal location within the same cytogenetic band, a recent gene duplication scenario of *btd* and *D*-Sp1 was suggested. In addition, a direct *btd* ortholog was not known for vertebrates. To elucidate the origin of *btd*, different Sp genes were isolated from different arthropod species. Fully sequenced genomes of various metazoan representatives were also searched for Sp genes. Phylogenetic sequence analysis of these data as well as protein domain structure, chromosomal location and mRNA expression analyses revealed *btd* orthologs even in basal metazoan representatives as *Nematostella vectensis* and *Trichoplax adhearens*. Thus, a set of three Sp genes is ancestral in the metazoans and the origin of *btd* can be traced back to a basal metazoan ancestor.

Furthermore, the contribution of the *D. melanogaster btd* target gene *col* to head metamerization was analyzed in various arthropod species. In *D. melanogaster*, *col* is expressed early in parasegment 0, which contributes to the posterior part of the intercalary (tritocerebral) and the anterior part of the mandibular segment, and later during embryonic development in the nervous system. The tritocerebral segment of insects lacks appendages, whereas the same segment in crustaceans and chelicerates still possesses appendages. Our data revealed an early head function for *col* in the tritocerebral segment of the insect representatives *Tribolium castaneum* and *Oncopeltus fasciatus*. Intriguingly, this early head function is missing in the crustacean *Parhyale hawaiensis* and the chelicerate *Achaearanea tepidariorum* where *col* is only expressed late during embryonic development in the nervous system. This suggests a contribution of *col* to the appendage-less state of the tritocerebral segment.

2 Introduction

2.1 Mechanisms of arthropod head development – views and open questions

For a long time, embryos of different arthropod species have been used for pattern formation studies. Most molecular data for these studies are available from insects. Especially the fruit fly *Drosophila melanogaster* is a very well established model organism for developmental studies. Various studies in *D. melanogaster* showed that the trunk is patterned by a hierarchical segmentation cascade, with the help of maternal effect genes, gap-genes, pair-rule genes and segment-polarity genes (Ingham and Martinez Arias, 1992; St Johnston and Nüsslein-Volhard, 1992). Homeotic selector (Hox) genes further specify these segments. Most of these molecular studies focussed primarily on the trunk region of the embryo and much less is known about patterning processes of the embryonic head. Nevertheless, some studies from *D. melanogaster* revealed different mechanisms for patterning the head versus the trunk.
The insect embryonic head can be subdivided into the anterior procephalon and the posterior gnathocephalon. Expression patterns of the segment polarity genes *engrailed* (*en*) and *wingless* (*wg*) as well as internal head structures argue for seven head domains, four pregnathal (labral-, ocular-, antennal- and intercalary segment) and three gnathal (mandibular-, maxillary- and labial segment) (Schmitt-Ott and Technau, 1992). The segmental or non-segmental nature and further subdivision of the pre-antennal procephalic parts are still debated and remain unclear (reviewed in Finkelstein and Perrimon, 1991; Posnien and Bucher, 2009a).

2.1.1 Head gap genes and other anterior patterning genes in *D. melanogaster*

In *D. melanogaster*, the activity of pair-rule genes is completely missing in the procephalon and also Hox gene activity is absent in the very anterior part of the procephalon (see Fig.1). However, three mutations were discovered in a mutant screen which show gap-like deficiencies in the anterior head region and are therefore called gap-like head genes. The three head gap-like genes *orthodenticle* (*otd*), *empty-spiracle* (*ems*) and *buttonhead* (*btd*) are expressed early in the blastoderm in an anterior stripe that in each case corresponds to the anlagen of the larval head structures affected in the mutant. The early expression of *otd*, *ems* and *btd* is dependent on the activity of the anterior maternal effect genes *bicoid* (*bcd*) and *hunchback* (*hb*) (reviewed in Rogers and Kaufman, 1997). Analyses of mutant larval head cuticles and mutant embryos stained with segmental marker genes showed that *otd*, *ems* and *btd* are required for overlapping but not identical head structures. *Otd* is crucial for antennal- and pre-antennal (ocular- and labral

segment) structures, *ems* for the intercalary, the antennal and also pre-antennal structures (ocular segment) and *btd* is involved in the formation of the antennal, the intercalary and the mandibular segment (see also Fig. 1).

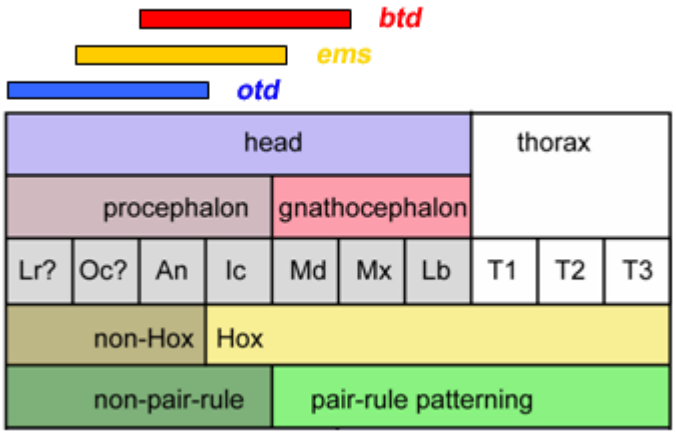


Figure 1. Schematic view of head subdivision and head gap gene expression domains in *Drosophila melanogaster* (modified after Cohen and Jürgens, 1990; Bucher and Wimmer, 2005).
Since the segmental or non-segmental identities and subdivisions of the procephalon are still unclear (reviewed in Finkelsten and Perrimon, 1991), the pre-antennal subdivisions of the ocular and labral region are indicated with a question mark. (Abbrevations: Lr- labral segment, Oc- ocular segment, An- antennal segment, Ic- intercalary segment, Md- mandibular segment, Mx- maxillary segment, Lb- labial segment, T- thoracic segments).

Due to the combinatorial overlapping expression domains and function in head metamerization, *otd*, *ems* and *btd* were thought to specify segmental identity (Cohen and Jürgens, 1990; reviewed in Finkelstein and Perrimon, 1991). Later on, Grossniklaus et al. (1992 and 1994) included a fourth gene into this combinatorial model. They showed that *sloppy paired* (*slp*) is also expressed early in the blastoderm in an anterior stripe and it functions in the development of the ocular-, the antennal- and mandibular segment. However, this combinatorial model seems to be fragmentary because misexpression experiments for *Dm btd* and *Dm otd* did not alter head segmentation (Wimmer et al., 1997; Gallitano-Mendel and Finkelstein, 1998). Further ideas for an anterior head patterning mode distinct from trunk patterning mechanisms include the activity and cross-regulatory interactions of the segment polarity genes *wg* and *hedgehog* (*hh*) (Gallitano-Mendel and Finkelstein, 1999). Also a requirement for so called second-level regulators, which act in the anterior head region like the helix loop helix transcription factor *collier* (*col*), has been proposed (Crozatier et al., 1996 and 1999) (see also 2.3).

2.1.2 Anterior patterning mechanisms in other arthropods

The genetic interactions by which the insect head is patterned are still not completely understood. This is mainly due to the high diversity of gene functions during head patterning in the insects. In *D. melanogaster*, the maternal morphogen Bcd is located anteriorly in the embryo and is the main anterior organizing factor. In weak *bcd* mutants, the head is missing whereas in strong *bcd* mutants also thoracic and abdominal structures are missing (Frohnhöfer and

Nüsslein-Volhard, 1986; Berleth et al., 1988). Since *bcd* is only found in higher dipterans (e.g. McGregor, 2005) anterior patterning gene networks seem different in diverse insects.

Analyses in the red flour beetle *Tribolium castaneum* demonstrated that *otd* is expressed maternally and *otd* knockdown via RNA interference (RNAi) leads to headless embryos (Schröder, 2003). Double knockdowns of *otd* and *hb* enhance the single *otd* phenotype and drastically reduce the embryo to very few abdominal segments. It seems therefore that *otd* and *hb* substitute for *bcd* in *T. castaneum*.

Studies on *otd* function are also available in a non-insect arthropod, namely the common house spider *Achaearanea tepidariorum*. Pechmann et al. (2008) could show that anterior regionalization in the spider is mediated by dynamic spatio-temporal expression of *hh* and *hairy* (*h*), which is controlled by *otd*. RNAi with *otd* disturbs this dynamic expression and affected embryos show a lack of all structures anterior to the pedipalpal (tritocerebral) segment.

So far, the function and regulatory mechanisms of the *D. melanogaster* head gap-like gene *btd* in other insects (and arthropods) are unclear. Recent results in *T. castaneum* suggest that *btd* has no role in head development (Schinko et al., 2008), and to date *btd* homologs from other arthropods had not been isolated.

2.2 The Sp family of transcription factors

The *D. melanogaster* head gap gene *btd* is a zinc finger transcription factor and belongs to the Sp family of transcription factors. Members of the Sp family have been described in several species (e.g. Suske et al., 2005; Zhao and Meng, 2005).

The general human transcription factor Sp1 (named after the original purification method through sephacryl and phosphocellulose columns) was the first cloned and described binding site specific transcription factor (Dynan and Tijan, 1983a; Dynan and Tijan, 1983b; Kadonaga et al., 1987). Members of this transcription factor family share some highly conserved features. All Sp factors possess three C2H2-type zinc fingers near the C-terminus, which bind to G-rich DNA elements, such as GC- and GT-boxes (Kadonaga et al., 1987). Those elements are crucial for proper activation of TATA-less promotors. It has been shown that these binding sites are present in many control regions of both tissue specific and ubiquitously expressed genes (Philipsen and Suske, 1999; Suske, 1999). This indicates that Sp factors potentially regulate a large number of diverse target genes. Previous studies in vertebrate systems could show that members of the Sp-family are involved in such diverse functions as the development of several organ systems, control of morphogenetic pathways, cell cycle regulation and have also been linked to cancer development (Marin et al., 1997; Black et al., 1999; Black et al., 2001; Nakamura et al., 2004; Safe and Abdelrahim, 2005; Chen et al., 2006).

Directly in front of the zinc-finger region is a motif located called Btd-box which seems to function as a transactivation domain (Zhao and Meng, 2005). Further structural motifs, especially at the N-terminus of the proteins, are specific for individual subgroups of the Sp-family. Based on these structural motifs, Bouwman and Philipsen (2002) subdivided all human Sp factors into two major groups, Sp1-4 and Sp5-8. At the time, only eight Sp genes were known in the human genome. Further annotations revealed a set of nine Sp genes in vertebrates like human, mouse, and chick, whereas some fish genomes posses more Sp genes (e.g. 13 in zebrafish *Danio rerio* and 11 in the pufferfish *Fugu rubripes*) due to further partial genome duplications (e.g. Zhao and Meng, 2005).

2.2.1 The vertebrate Sp1-4 group

Sp1 and *Sp3* are expressed ubiquitously during mouse embryogenesis (Saffer et al., 1991; Bouwman et al., 2000). *Sp1* mutant mice are severely retarded in development, show a broad range of abnormalities and die around day 11 of gestation (Marin et al., 1997). *Sp3* mutant mice also show growth retardation, show defects in late tooth development and die at birth due to respiratory failure (Bouwman et al., 2000). Moreover, mouse *Sp3* is involved in ossification mechanisms (Göllner et al., 2001a). *Sp2* is only known to be expressed in several mammalian cell lines (Moorefield et al., 2004), but further insights into its biological function during vertebrate development are still missing. Mouse *Sp4* is also expressed ubiquitously, with slightly enhanced expression in the CNS (Supp et al., 1996). Two-thirds of *Sp4* mutant mice die within a few days after birth. The others are retarded in growth and show deficiencies in mating behaviour (Göllner et al., 2001b). These data reveal the importance of at least *Sp1*, *Sp3* and *Sp4* during embryonic development.

2.2.2 The vertebrate Sp5-9 group

In contrast to the ubiquitous expression of the vertebrate Sp1-4 group, genes from the vertebrate Sp5-9 group show tissue-specific expression patterns. Data for *Sp5* in vertebrates are available from mouse and zebrafish. Mouse *Sp5* expression is first detected in the primitive streak during und throughout gastrulation. Afterwards, it is expressed in the tail bud, limb buds, otic vesicles, the developing central nervous system, the pharyngeal region and at the mid-hindbrain boundary (MHB) (Harrison et al., 2000; Treichel et al., 2001). Interestingly, *Sp5* null mice did not show any phenotype but *Sp5* enhances the mouse *Brachyury* phenotype (Harrison et al., 2000). This points to a complex role of *Sp5* for proper expression of several target genes. Three *Sp5* related genes are present in the genome of the zebrafish *Danio rerio*, namely *Sp5* (also known as *bts1*) (Tallafuss et al., 2001), *Sp5-like* (also known as *spr2*) (Zhao et al., 2003) and *similar-to-*

Sp5. *Dr Sp5* is expressed early in the ventral and lateral margins of the blastoderm, later in the MHB, the tail bud, somites and otic vesicles (Tallafuss et al., 2001). *Dr Sp5-like* is expressed similarly to *Dr Sp5* with early expression domains additional to the epiblast also in hypoblast cells and during later development in the tail bud, somites and in the brain (Zhao et al., 2003). Loss- and gain-of-function experiments revealed that *Dr Sp5-like* mediates *no tail* (*ntl*) expression. *ntl* is a downstream target of Fgf signalling and the zebrafish homolog of mouse *Brachyury* (Zhao et al., 2003). Therefore, *Sp5* homologs in mouse and fish seem to modify Fgf signalling.

Mouse *Sp6* is specifically expressed in the apical ectodermal ridge (AER) of the developing limbs and in hair follicle cells (Scohy et al., 2000; Nakamura et al., 2004; Hertveldt et al., 2008). *Sp6* deficient mice are nude and show defects in skin, limbs (syndactyly and oligodactyly), teeth and lung alveols. Mouse *Sp7* (also known as *osterix*) is expressed in osteoblasts and consequently, *Sp7* null mice show defects in bone formation due to impaired osteoblast formation (Nakashima et al., 2002; Milona et al., 2003; Kaback et al., 2008).

Data for *Sp8* and *Sp9* are available from mouse, zebrafish and chicken. Both genes are very similar in expression and function. They are expressed in the AER and are essential for limb and fin outgrowth in mouse, zebrafish and chicken (Bell et al., 2003; Treichel et al., 2003; Kawakami et al., 2004; Griesel et al., 2006). Additional to the function in limb development, *Sp8* is also expressed in the MHB and is required for normal development of this region (Griesel et al., 2006; Kawakami et al., 2004).

2.2.3 Insect Sp factors

As already mentioned, Sp factors are also present outside the vertebrates. However, there were few data available regarding the function of Sp transcription factors in insects. In the genome of *D. melanogaste*r, there are three Sp genes: *btd*, *D-Sp1* and *CG5669*.

Dm Btd was the first insect Sp transcription factor described and analyzed (Wimmer et al., 1993). *btd* is expressed early in an anterior stripe and later in the central and peripheral nervous system (Wimmer et al., 1993, Wimmer et al., 1996). Larval cuticles from *btd* mutant flies show defects in the antennal, intercalary and mandibular segments (Cohen and Jürgens, 1990) (see also 2.1).

A second Sp factor gene was isolated from *D. melanogaste*r and due to its sequence similarity to the (to that date the only other known vertebrate representative of the Sp family) human transcription factor Sp1 it was termed *D-Sp1* (Wimmer et al., 1996). Nowadays, we have access to completely sequenced genomes and it turned out that *D-Sp1* is more similar to the vertebrate Sp5-9 group than to the Sp1-4 group (Schaeper et al., 2009). The postblastodermal expression pattern of *D-Sp1* is similar to *btd*.

It was shown that *btd* and *D-Sp1* possess partially redundant functions in the peripheral nervous system (Schöck et al., 1999), but only *btd* is essential for early head development (Wimmer et al., 1996). Estella et al. (2003) could show a redundant function of *btd* and *D-Sp1* in leg- and antennal imaginal disc development. The knockdown via RNAi of only one of the genes does not severly affect antennal- and leg growth, whereas a reduction of both transcripts leads to strong reduction of size of legs and antennae. Chromosomal analysis showed that *btd* and *D-Sp1* are located on the same cytogenetic band on the *Drosophila* X-chromosome. This fact and the partially redundant function of both genes led the authors to propose that this gene pair arose from a recent gene duplication (Wimmer et al., 1996; Schöck et al., 1999). Supporting this, no direct *btd* homologue could be found in vertebrates. Interestingly, the mouse *Sp8* phenotype is very similar to the *Dm btd* head phenotype. Therefore, mouse *Sp8* was originally termed *mBtd* (Treichel et al., 2003).

Beerman et al. (2004) isolated and described a *D-Sp1* homologue in the red flour beetle *T. castaneum*, called *Tc Sp8*. *Tc Sp8* is expressed in the appendages and the nervous system, and is involved in allometric limb growth. A *btd* gene was also isolated from *T. castaneum*. It is expressed similar to *Dm btd* in an early stripe and late in the nervous system. RNAi experiments however did not show any phenotype (Schinko et al., 2008).

Nevertheless, the origin and orthology of insect *btd* remained obscure and had to be elucidated. Furthermore, no expression and functional data from other arthropod *Sp1-4* genes were available.

2.3 Development of the arthropod tritocerebral head segment

In *D. melanogaster* the Sp family member BTD is required for activation of another head gene, the so called second level regulator *col* (see Fig. 2) (Crozatier et al., 1996 and 1999). This requirement of *btd* for *col* activation is another interesting point for investigating *col* function in various arthropods. The segmental composition of the arthropod head is highly conserved (Damen et al., 1998; Telford and Thomas, 1998) but the specific morphology of the head segments itself can be quite diverse. The arthropod tritocerebral head segment, which gives rise to the tritocerebrum, shows divergence in external appendages among this phylum (see Fig. 3). Originally, this segment possessed a pair of appendages. In chelicerates and crustaceans, the tritocerebral segment still carries appendages and is termed pedipalp-segment or second antennal segment, respectively. However, in insects and myriapods, this segment lacks appendages and is called intercalary segment or pre-mandibular segment, respectively. The genetic mechanisms for absence or existence of appendages on this segment are not known.

Despite differences in appearance of the tritocerebral segment, the anterior-most expressed Hox gene, *labial* (*lab*), is expressed only in the tritocerebral segment in all four arthropod phyla (reviewed in Hughes and Kaufmann, 2002). More molecular data have to be collected to reveal the evolution of the appendage less state of this arthropod head segment. The *D. melanogaster* gene *collier* (*col*) is expressed early in the blastoderm in the region of parasegment 0 (PS0), which gives rise to the posterior part of the intercalary and the anterior part of the mandibular segment. Crozatier et al. (1999) suggested that *Dm col* functions as an intermediate regulator downstream of the head gap-like genes to pattern the intercalary segment (see Fig. 2). Therefore, *collier* should be a good candidate for further investigation of this particular part of the arthropod head.

2.3.1 The COE family

Dm col (also known as *knot*) can be first detected at mitotic cycle 14 in PS0 and is crucial for head structure formation in this area (Crozatier et al., 1996; Crozatier et al., 1999; Seecoomar et al., 2000). Later in development, *col* is expressed in the central and peripheral nervous system (Crozatier et al., 1996; Baumgardt et al., 2007; Crozatier and Vincent, 2008), is crucial for somatic muscle formation (Crozatier and Vincent, 1999; Dubois et al., 2007; Crozatier and Vincent, 2008) and is involved in wing patterning (Nestoras et al., 1997; Vervoort et al., 1999; Mohler et al., 2000; Crozatier et al., 2002; Hersh and Caroll, 2005).

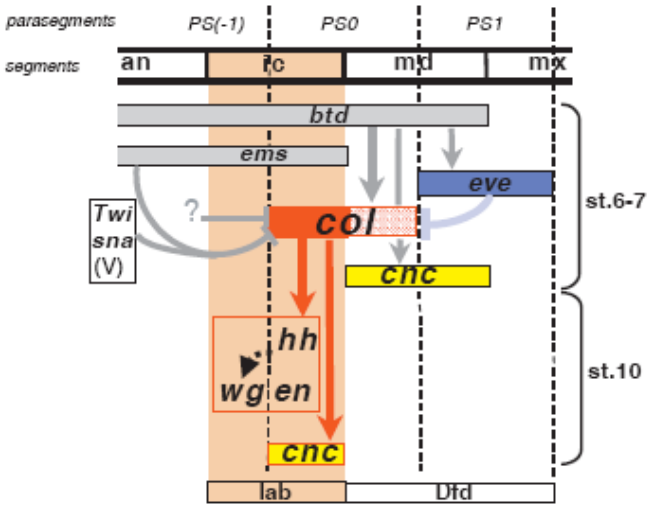


Figure 2. Schematic view of regulatory mechanisms controlling the formation of the intercalary segment in *Drosophila melanogaster* (taken from Crozatier et al., 1999).
btd activates *col* in PS0 in early blastodermal stages, however *col* is restricted by *eve* in PS1 and *twist*, *snail* in PS(-1). Later on, *col* activates *cnc* and *hh* expression in the intercalary segment, *hh* regulates *wg* activation. Therefore, *col* is required for the establishment of segment polarity gene expression in the intercalary segment.

Members of the COE family (named after the factors Collier/Olf/EBF) can be found throughout the metazoan taxon and are involved in a variety of processes during embryonic development. In vertebrates, several paralogous of the COE family are present. Fish, frog and chicken

possess three members of the COE family, four COE members are present in mammals (mice and humans). In invertebrates and also in the cnidarian *Nematostella vectensis*, only one single orthologue of the COE family can be found (reviewed in Dubois and Vincent, 2001; Liberg et al., 2002). Mouse Olf-1/EBF-1 is necessary for B-cell differentiation (reviewed in Liberg et al., 2002). Moreover, Olf/EBF in mouse is expressed in immature olfactory neuronal precursors, in the developing nervous system and mature neurons of the adult olfactory epithelium (reviewed in Dubois and Vincent, 2001). It was also shown that mouse and chicken *ebf1*, *ebf2* and *ebf3* are involved in limb development (Mella et al., 2004).

The *Caenorhabditis elegans* COE representative *unc-3* plays a role in ventral nerve cord development, axonal guidance and chemosensory neuron differentiation (Prasad et al., 1998). Similarly, the *N. vectensis coe* homolog seems to be also involved in chemoreception (Pang et al., 2003). Taken together, these data indicate an ancestral role of COE factors in chemoreception.

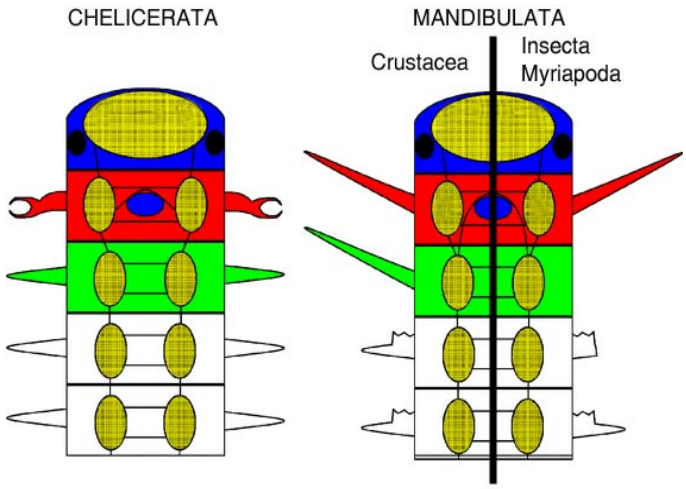


Figure 3. Schematic view of arthropod head structures (modified after Scholtz and Edgecombe, 2006). The protocerebral/ocular region (the segmental nature of which is disputed) is marked in blue, the deutocerebral segment is marked in red and the tritocerebral segment in green. Structures of the central nervous system are shaded in yellow.
The deutocerebral segment corresponds to the cheliceral segment in chelicerates, to the first antennal segment in crustaceans and to the antennal segment in insects and myriapods. In chelicerates and crustaceans, the tritocerebral segment carries a pair of appendages and is termed pedipalpal segment and second antennal segment, respectivley. In insects and myriapods, the tritocerebral segment lacks appendages and is called intercalary and pre-manibular segment, respectively.

2.4 Research objectives

Despite the conserved segmental composition of the head among the arthropods (Damen et al., 1998; Telford et al., 1998), the anterior patterning mechanisms are poorly understood. In this study, homologs of the *D. melanogaster* head gap-like gene *btd* and further Sp family members were isolated from different arthropods. The aim was to reveal a putative conserved role in head development for Sp factors. Moreover, the evolution of Sp transcription factor genes in the Metazoa had to be reconstructed in order to investigate the debated origin and homology of insect *btd* to vertebrate Sp factors.

Hence, a broad screen for Sp factors in already sequenced genomes of different metazoan species was carried out. Additionally, Sp factor genes were isolated from different arthropod species such as representatives of holometabolous insects like the fruitfly *D. melanogaster* and the red flour beetle *T. castaneum*, from the hemimetabolous insect *Oncopeltus fasciatus* (milkweed bug), the basal Hexapods *Thermobia domestica* (firebrat) and *Folsomia candida* (white springtail), and the crustacean *Parhyale hawaiensis* (amphipod). For all isolated Sp genes, the spatio-temporal expression pattern had to be determined (see 3.1) and where it was possible, loss of function experiments via RNA interference were performed (see 3.2 and 3.3). Therefore, for *O. fasciatus* and *T. domestica*, animal husbandry, embryo fixation and in situ hybridisation procedures as well as RNAi techniques for *O. fasciatus* had to be improved to our lab conditions. For the springtail *F. candida*, embryo fixation and in situ hybridisation procedure had to be newly established. To gain insight into the possible ancestral function of Sp factors, these genes had to be isolated from the cnidarian *N. vectensis* and stained by in situ hybridization (see 3.4). As another criterion to reveal the homology among various Sp factors, the Sp protein domains and the genomic location had to be compared.

To further analyze arthropod head segmentation and the origin of the appendage-less state of the tritocerebral segment, the *D. melanogaster btd*-controlled second level regulatory gene *collier* was chosen. Therefore, *col* homologues had to be isolated from different arthropod classes. We chose *T. castaneum*, the hemimetabolous insect *O. fasciatus*, *P. hawaiensis* as a crustacean representative and from the chelicerate group the common house spider *Archaearanea tepidariorum*. Data from the spatio-temporal expression pattern as well as RNAi data from the insect representatives had to be gained (see 3.5).

3 Results

Each chapter within the results section starts with a one-page description of:

- The aim of the particular manuscript in the context of the thesis as a whole.
- The status of the manuscript
- The author’s contribution to the practical work

3.1 A clustered set of three Sp-family genes is ancestral in the Metazoa: evidence from sequence analysis, protein domain structure, chromosomal location and developmental expression patterns.

The purpose of this work was to reconstruct evolution and origin of the Sp transcription factor family using evidence from phylogenetic sequence analysis and protein domain structure. This also included the isolation of Sp genes from different arthropod species and the analysis of their spatio-temporal expression pattern. In order to be able to also include expression data from Sp genes of the basic hexapod *Folsomia candida*, the whole-mount in situ hybridization technique had to be newly established for this springtail species.
Furthermore, this analysis also included the broad search for Sp factors in already sequenced and available genomes of different metazoan species and the comparison of their chromosomal location.

Nina D. Schäper, Nikola-Michael Prpic, Ernst A. Wimmer

Status: manuscript in preparation

Author contribution to the practical work:
Nina Schäper did all the experiments.
Nikola-Michael Prpic performed the computer based phylogenetic analysis of Sp factors.

A clustered set of three Sp-family genes is ancestral in the Metazoa: evidence from sequence analysis, protein domain structure, chromosomal location and developmental expression patterns.

Nina D. Schaeper[1], Nikola-Michael Prpic[1], Ernst A. Wimmer[1]*

1 Georg-August-Universität, Johann-Friedrich-Blumenbach-Institut für Zoologie und Anthropologie, Abteilung Entwicklungsbiologie, GZMB, Ernst-Caspari-Haus, Justus-von-Liebig-Weg 11, 37077 Göttingen, Germany

(*) Author for correspondence:
Ernst A. Wimmer
Georg-August-Universität, Johann-Friedrich-Blumenbach-Institut für Zoologie und Anthropologie, Abteilung Entwicklungsbiologie, GZMB, Ernst-Caspari-Haus, Justus-von-Liebig-Weg 11, 37077 Göttingen, Germany

Tel.: +49 551 39 22889
FAX: +49 551 39 5416
e-mail: ewimmer@gwdg.de

Number of words: 12134 (in total)
Number of figures: 10
Number of tables: 1

Abstract

Background

The Sp-family of transcription factors are evolutionarily conserved zinc finger proteins present in many animal species. The orthology of the Sp genes in different animals is unclear and their evolutionary history is therefore controversially discussed. This is especially the case for the Sp gene *buttonhead* (*btd*) which plays a key role in head development in *Drosophila melanogaster*, and has been proposed to have originated by a recent gene duplication in the dipteran lineage. The purpose of this study is to trace orthologs of *btd* in other insects and reconstruct the evolutionary history of the Sp genes within the metazoa.

Results

We isolated Sp genes from representatives of an holometabolous insect (*Tribolium castaneum*), an hemimetabolous insect (*Oncopeltus fasciatus*), primitively wingless hexapods (*Folsomia candida* and *Thermobia domestica*), and an amphipod crustacean (*Parhyale hawaienis*). We supplemented this data set with data from fully sequenced animal genomes. We performed phylogenetic sequence analysis with the result that all Sp factors fall into three monophyletic clades. These clades are also supported by protein domain structure, chromosomal location, and gene expression. We show that clear orthologs of the *D. melanogaster btd* gene are present even in the basal insects, and that the *Sp5*-related genes in the genome sequence of several deuterostomes and the basal metazoans *Trichoplax adhaerens* and *Nematostella vectensis* are also orthologs of *btd*.

Conclusions

All available data provide strong evidence for an ancestral cluster of three Sp-family genes as well as syntheny of this Sp cluster and the Hox cluster. The ancestral Sp gene cluster already contained a *Sp5/btd* ortholog, which strongly suggests that *btd* is notthe result of a recent geen duplication, but directly traces from an ancestral gene already present in the metazoan ancestor.

Background

Zinc finger transcription factors are a large and widespread family of DNA binding proteins and play an important role in transcriptional regulation [e.g. Suske et al., 2005]. The general transcription factor Sp1 (named after the original purification method through sephacryl and phosphocellulose columns) was the first identified and cloned binding specific human transcription factor [Dynan and Tijan, 1983a; Dynan and Tijan, 1983b; Kadonaga et al., 1987]. In the meantime a number of additional genes related to *Sp1* have been identified in the human genome, and homologous genes have been isolated from several other animal species as well (e.g. [Suske et al., 2005; Zhao and Meng, 2005]). The members of this Sp-family of transcription

factors share three highly conserved Cys2His2-type zinc fingers, which bind to G-rich DNA elements, such as GC-boxes (GGGGCGGGG) and GT/CACC-boxes (GGTGTGGGG) [3 Kadonaga et al., 1987]. These binding sites are present in many control regions of both tissue-specific and ubiquitously expressed genes [Philipsen and Suske, 1999; Suske, 1999] indicating that Sp-family transcription factors potentially regulate a large number of target genes. Indeed, it was shown that Sp-family transcription factors have diverse functions throughout the embryonic development of humans and other animals. For instance, in vertebrates they are involved in cell cycle regulation, the control of morphogenetic pathways, the development of several organ systems, and they also have been linked to the development of cancer [e.g. Wimmeret al., 1996; Marin et al., 1997; Black et al., 1999; Black et al., 2001, Treichel et al., 2001; Treichel et al., 2003; Kawakami et al., 2004; Nakamura et al., 2004; Safe and Abdelrahim, 2005; Zhao and Meng, 2005; Chen et al., 2006]. In the fly *Drosophila melanogaster,* the gene *buttonhead* (*btd*) codes for a member of the Sp-family, which represents an important factor for the formation of several head segments and is also involved in the development of the central and peripheral nervous system [Cohen and Jürgens, 1990; Wimmer et al., 1993; Wimmer et al., 1996; Schöck et al., 1999].

The number of Sp-family genes present in the genome varies in the Metazoa. Humans and mice, for example, have nine Sp-family genes [Zhao and Meng, 2005], and some teleost fishes have even more (11 in the pufferfish *Fugu rubripes* [Aparicio et al., 2002], 13 in the zebrafish *Danio rerio* [zebrafish sequencing project]). From *D. melanogaster* two Sp-family genes have been reported, *btd* and *D-Sp1* [Wimmer et al., 1996], but a third one is present in the fully sequenced genome sequence. This variable complement of Sp-family genes and their evolutionary diversification make it difficult to assign orthology between the genes of different species. Therefore, the ancestral number of Sp-family genes and the evolution and orthology of the hitherto identified Sp-family genes was unclear. This situation also led to a considerable confusion in the nomenclature of the Sp-family genes and to several unfortunate designations of not directly homologous Sp-family members with homonymous names thus misleadingly suggesting orthology. *D. melanogaster D-Sp1* is not mostly related to human *Sp1* but *Sp8* [Beerman et al., 2004] and the originally termed mouse *mBtd* is *Sp8* [Treichel et al., 2003].

Especially the origin and orthology of the *D. melanogaster* head gap gene *btd* is debated. Previous studies discovered functional similarities between *btd* and some vertebrate Sp genes, but could not confidently identify a genuine *btd* orthologue in vertebrates [Tallafuss et al., 2001; Treichel et al., 2003; Nakamura et al., 2004], and it had been proposed that the *btd* gene might be the result of a recent gene duplication when another Sp-family gene, *D-Sp1*, in the vicinity of *btd* was discovered [Wimmer et al., 1996; Schöck et al., 1999]. This gene is not only located directly next to *btd*, but the two genes also have similar postblastodermal expression patterns

and partially overlapping developmental functions [Wimmer et al., 1996; Schöck et al., 1999]. All this suggested that *btd* evolved by a tandem duplication in the phylogenetic lineage leading to *D. melanogaster*.

In order to reconstruct the evolution of the Sp-family genes, we have first tried to trace homologs of *btd* in other insects. We have surveyed not only additional dipterans and other holometabolous insects, but we have also searched for Sp-family genes in representatives of hemimetabolous insects (the heteropteran *Oncopeltus fasciatus*) and the primitively wingless ectognathous and entognathous hexapods (the zygopteran *Thermobia domestica* and the collembolan *Folsomia candida*, respectively). We could identify clear orthologs of the *D. melanogaster btd* gene in these basal haxapods, indicating that the proposed gene duplication did not take place recently within the insects. We have therefore performed a comprehensive study of Sp-family gene evolution based on phylogenetic sequence analysis, protein domain structure characteristics, genomic localisation, as well as spatio-temporal mRNA expression analysis. Our phylogenetic analysis shows that the available Sp-family factors fall into three large clades and that a true *btd* ortholog is already present in the basal metazoans *Trichoplax adhaerens* and *Nematostella vectensis*. The proteins in each clade also display similar structural characteristics and often form a cluster of three genes in the genome. Intriguingly, the available data suggest that this Sp gene cluster has been ancestrally linked to the Hox gene cluster and in the vertebrates appears to have been affected by the multiple duplications of this cluster. This syntheny and co-evolution of the Hox and the Sp clusters in the vertebrates also explains the high number of Sp-family genes in this animal group. Our analysis suggests that a clustered set of three Sp-family genes is ancestral in the Metazoa, and this Sp gene cluster has been originally linked to the Hox gene cluster.

Results and Discussion

A search for Sp-family genes in insects and crustaceans

As mentioned in the introduction, previous work had suggested that *D. melanogaster* possesses two closely related Sp genes, *btd* and *D-Sp1* [Wimmer et al., 1993; Wimmer et al., 1996]. However, a search in the fully sequenced *D. melanogaster* genome revealed the presence of an additional gene, *CG5669*, with high similarity to *btd* and *D-Sp1*. This complement of three Sp-family genes could be the result of a recent gene duplication [Wimmer et al., 1996; Schöck et al., 1999]. In order to identify when such a gene duplication event might have occured, we sought to identify the number of Sp-family genes in additional insect species.

We searched the genome sequence of selected insect species with fully sequenced genomes. In addition we performed PCR-based surveys in specially selected additional species. In the Diptera, a complement of three Sp-family genes seems to be the rule: in the genome sequences of *Drosophila pseudobscura* and the mosquito *Anopheles gambiae* we found three different Sp-family genes each. We then searched in the genomes of species outside the Diptera. In the lepidopteran *Bombyx mori* (silk moth), the hymenopterans *Apis mellifera* (honeybee) and *Nasonia vitripennis* (jewel wasp), and the coleopteran *Tribolium castaneum* (flour beetle) we also detected three Sp-family genes each. This taxon sampling included only holometabolous insects and we have therefore also isolated cDNA fragments of Sp-family genes from representatives of the hemimetabolous and the primitively wingless hexapods. In the higher hemimetabolous heteropteran *O. fasciatus* (milkweed bug), we were able to isolate two different Sp-family gene fragments. The Zygentoma represent the youngest branch of the primitively wingless insects. We have used the zygentoman *T. domestica* (firebrat), from which we could isolate three different Sp-family gene fragments. The Collembola are members of the most basal branch of the primitively hexapods (Entognatha). In the collembolan *F. candida* (white springtail) we were also able to detect three different fragments of Sp-family genes.

These results show that a complement of three Sp-family genes is present in all studied hexapod species, except for *O. fasciatus* for which the genome sequence is not available and a third Sp-family member may have been missed in our PCR-based search. We have then tried to establish the number of Sp-family genes in the Crustacea, which phylogenetically is the sister group of the insects according to recent analyses (e.g. [Friedrich and Tautz, 1995; Dohle, 1997; Dohle, 2001; Budd and Telford, 2009]). The waterflea *Daphnia pulex* is a member of the Branchiopoda, a group of crustaceans with a primarily limnic lifestyle. In the fully sequenced genome of *D. pulex* we detected the presence of three different Sp-family genes. The Malacostraca (higher crustaceans) are a group of primitively marine species. We have used PCR to isolate Sp-family gene fragments from the malacostracan *Parhyale hawaiensis* (beachhopper), which yielded two different fragments. However, as with the results for *O. fasciatus* the PCR survey may have missed an additional Sp-family gene in *P. hawaiensis*.

Taken together, these results strongly suggest that a complement of three different Sp-family genes is ancestral in the arthropods. Interestingly, three different Sp-family genes are also present in the fully sequenced genomes of the basal chordate *Branchiostoma floridae*, and the echinoderm *Strongylocentrotus purpuratus*. Three different Sp-family genes are also present in the fully sequenced genomes of the cnidarian *N. vectensis*, and the placozoan *T. adhaerens* - both representing basal branches in the metazoan phylogenetic tree. This could be taken as evidence that the possession of three Sp-family genes is ancestral in the Metazoa. On the other hand, the high number of Sp-family genes in the genomes of vertebrates (e.g. nine Sp-family

genes in humans and mice, 7 in the chicken, and more than 10 in fish), indicates that the Sp-genes can be subject to frequent duplications. Thus, the "triplets" in insects, cnidarians, placozoans, echinoderms, and basal chordates might as well have originated independently.

Phylogenetic analysis of Sp-family genes supports three large clades

In order to distinguish between a possible ancestral set of three Sp-family genes and the alternative possibility of several independent duplication events, we reconstructed the evolutionary history of selected Sp-family factors and assigned orthology by phylogenetic sequence analysis. We used the amino acid sequence of the region including the Btd box, the three zinc fingers and the sequence in between these two domains of all available Sp-family factors of *Homo sapiens* (human), *Mus musculus* (mouse), *Gallus gallus* (chicken), *D. rerio* (zebrafish), *F. rubripes* (pufferfish), *B. floridae* (lancelet), *S. purpuratus* (sea urchin), *T. adhaerens* (placozoan), *N. vectensis* (sea anemone), and the insect and crustacean species mentioned above in a maximum likelihood analysis with the Tree Puzzle program package. The resulting unrooted tree is shown in Fig. 1, used protein sequences in Fig. S1. The tree comprises three large monophyletic groups. One clade contains Sp1, Sp2, Sp3 and Sp4 of the vertebrate species and a single Sp representative of each of the invertebrate species. We term this clade the Sp1-4 clade. The second clade contains Sp5 of the vertebrate species and again a single Sp representative of each of the invertebrate species, except for *O. fasciatus* and *P. hawaiensis* for which we failed to obtain three different Sp-family genes in our PCR survey. Because this clade also contains the well-known Btd from *D. melanogaster*, we call this clade the Sp5/Btd clade. The third clade contains Sp6, Sp7, Sp8, and Sp9 of all vertrebrate species and a single Sp representative of each of the invertebrates. We call this clade the Sp6-9 clade. In order to facilitate the unique identification of the genes, we refer to all genes (except those that already have an official name) using the clade name to which they belong in our phylogenetic analysis. The distribution of a single Sp factor of each invertebrate species to each of the three clades strongly suggests that a set of three Sp-family genes, namely one *Sp1-4*, one *Sp5/btd* and one *Sp6-9* gene, is the ancestral state in the Metazoa and that the higher number in vertebrates resulted from independent duplications in the vertebrate lineage.

Protein structure supports the existence of two large groups of Sp factors

We next tried to gather additional evidence for a grouping of all Sp-family factors into three large clades. It had been noted previously that the Sp proteins contain additional structural domains besides the zinc fingers and Btd box [e.g. Bouwman and Philipsen, 2002]. A large portion of the

N-terminal end of the proteins is enriched for certain amino acid residues. We have therefore compared the composition of Sp proteins from human, sea anemone, and selected arthropods (Fig. 2). The proteins of the Sp1-4 clade are longer proteins characterized by a (mostly) bipartite glutamine-rich region divided by a region enriched mostly for serine and threonine. These proteins form a well recognizable grouping that we call Sp1-4 group. The structure of the Sp1-4 group is clearly different from the Sp proteins of the Sp5/Btd and Sp6-9 clades (Fig. 2). These two clades contain shorter proteins (on average), and are more similar to each other than each is to the Sp1-4 group and we therefore group the two clades together in a grouping that we call Sp5-9/Btd group. The N-terminal end of these proteins contains only a single long region enriched for serine and/or proline. However, we note a trend in the Sp5/Btd clade towards the accumulation of more proline, whereas in the Sp6-9 clade there is a clear trend towards accumulating serine and threonine in the N-terminal portion. Thus, the protein structure data also support the existence of three different groups of Sp-factors, but suggest that the Sp5/Btd clade and the Sp6-9 clade are more closely related.

Chromosomal location of Sp genes suggest an ancestral triplet

We have also established the location of the Sp-family genes in the genomes of fully sequenced and sufficiently annotated metazoan species; a schematic overview is shown in Fig. 3 and the exact locations are given in Table 1. Intriguingly, in the basal metazoan *N. vectensis* all three Sp-family genes are located next to each other on a single scaffold (scaffold 53). This situation is fully compatible with the notion that a triplet consisting of one Sp1-4, one Sp5/Btd, and one Sp6-9 gene is ancestral in the Metazoa. The close proximity of the genes on a single scaffold suggests that the Sp-family genes form a gene cluster of closely related genes evolved by tandem gene duplication similar to the genes in the Hox gene cluster. Ryan et al. [2007] and Putnam et al. [2007] have used the scaffold data of *N. vectensis* to reconstruct ancestral metazoan linkage groups (a kind of "ur-chromosomes"). Interestingly, the Sp cluster of *N. vectensis* is located next to the majority of the *N. vectensis* Hox genes on the hypothetical ancestral linkage group PAL A (Fig. 3, top) [Putnam et al., 2007]. Only the two Hox genes on scaffold 4 are not included in the PAL A. This suggests that the Sp gene cluster and the Hox gene cluster were ancestrally located next to each other and might have kept their syntheny and co-evolved. The Sp genes are located close to the Hox gene cluster in other animals as well (see also [Bouwman and Philipsen, 2002; Abbasi and Grzeschik, 2007]. Intriguingly, in humans, a triplet of one *Sp1-4*, one *Sp5/btd*, and one *Sp6-9* gene, namely *Sp3*, *Sp5*, and *Sp9*, is linked to the Hox D cluster and the remaining human Sp genes are arranged in duplets of one *Sp1-4* and one *Sp6-9* gene, which are linked to the remaining 3 Hox clusters respectively (Fig. 3, center). In *D. melanogaster* and *A. gambiae* only the *Sp6-9* clade gene is linked to the Hox gene cluster,

while the remaining two genes are located close to each other on the X chromosome (Fig. 3, bottom). These two genes are also located close to each other on another chromosome than the Hox gene cluster in *A. mellifera*, *T. castaneum* and the crustacean *D. pulex*. In addition, the *Sp1-4* gene representative is also not linked to the Hox cluster, although this is not fully established for *A. mellifera* and *T. castaneum*, because the *Sp1-4* gene is annotated within unassembled reads not placed in the assembled chromosome. The genomes of *S. purpuratus*, *B. floridae* and *T. adhaerens* are not yet fully assembled, but preliminary analysis provided additional evidence for Sp-family gene clustering in these species as well. In *S. purpuratus* the *Sp1-4* and *Sp5/btd* genes are located on the same scaffold. In both *B. floridae* (see also [Shimeld, 2008]) and *T. adhaerens* the *Sp5/Btd* and *Sp6-9* genes are located on the same scaffold (see also [Materna et al., 2006]). Whether the Sp-family genes are also linked to the Hox genes in *S. purpuratus* (see [Howard-Ashby et al., 2006]), *B. floridae* (see [Takatori et al., 2008; Holland et al., 2008]), or *T. adhaerens* (see [Schierwater et al., 2008]) has to await the full assembly of the scaffolds.

Embryonic expression patterns of insect and crustacean *Sp* genes

All available data collectively and consistently suggest that a small Sp gene cluster comprising three Sp genes is ancestral in the Metazoa and that the triplets present in the insects derive from these ancestral three genes, i.e. the genes in the respective clades are orthologous. This argues against the alternative hypothesis that the sets of three Sp genes in the different insect species originated by independent duplication events. As a final test of the orthologous nature of the three Sp genes in the different insect species we compared their expression patterns during embryogenesis by in situ hybridization. We reasoned that the genes of the same clade should show similar expression patterns in all species if they were true orthologos, but show different patterns if they originated through unrelated duplication events. In the following we compare the expression data from insects, the crustacean *P. hawaiensis* and published data from vertebrates arranged according to the three Sp-gene clades.

The genes of the Sp1-4 clade: *CG5669*, which is the *D. melanogaster* representative of this clade, is maternally contributed (Fig. 4A) and then expressed ubiquitously throughout development (Fig. 4B, C). In *T. castaneum* the *Sp1-4* gene is expressed ubiquitously throughout development as well (Fig. 5A-C). The same is true for the *Sp1-4* gene of *O. fasciatus* (Fig. 6A-C), *T. domestica* (Fig. 7A-C) and *F. candida* (Fig. 8A, B). In the crustacean *P. hawaiensis* the *Sp1-4 gene* is also expressed ubiquitously throughout all studied developmental stages (Fig. 9A-C). The members of this clade from the mouse have not all been studied as to their embryonic expression pattern, but data are available for murine *Sp1*, *Sp3* and *Sp4* [Saffer et al., 1991; Supp et al., 1996; Bouwman et al., 2000; Göllner et al., 2001a; Göllner et al., 2001b]. All three

genes are expressed ubiquitously during development. Taken together, these data show that all analyzed members of this clade are expressed in a similar ubiquitous fashion, strongly supporting the orthology of the genes.

The genes of the *Sp5/btd* clade: The expression of *btd* (the *D. melanogaster* representative of the *Sp5/btd* clade) has been reported previously [Wimmer et al., 1993, Wimmer et al., 1996]. The gene is first expressed in an anterior head stripe (Fig. 4D) and a dorsal spot appears slightly later (Fig. 4E). The head stripe is roughly located in the area of the intercalary and mandibulary segment and later abuts the cephalic furrow (Fig. 4F). Later a metameric (segmentally repeated) pattern emerges that might be correlated with segment formation and peripheral nervous system development (Fig. 4G-I) [Wimmer et al., 1996; Schöck et al., 1999]. Furthermore, *Dm btd* is expressed in the imaginal discs of legs and antennae [Wimmer et al., 1996; Estella et al., 2003]The expression of the *T. castaneum btd* gene has been published before [Schinko et al., 2008] and is very similar to the *D. melanogaster btd* pattern: *Tc-btd* is expressed in an early head stripe in the area of the intercalary and mandibulary segment (Fig. 5D) and later a metameric pattern emerges (Fig. 5E). In older stages the gene is also expressed in the appendages and in the nervous system (Fig. 5F). The expression pattern of *Sp5/btd* in *T. domestica* is very similar to the *T. castaneum btd* pattern. In the early blastoderm the gene is expressed in an anterior stripe (Fig. 7D), that lies in the intercalary/mandibulary area in slightly more advanced germ band stage embryos (Fig. 7E). Later a metameric pattern emerges (Fig. 7F, G) and in older stages expression in the nervous system and, weakly, in the appendages is detected (Fig. 7H). In *F. candida* we were not able to detect an early head stripe for *Sp5/btd*, because our fixation protocol did not allow us to fix blastoderm stages of this species. The later expression pattern of *Sp5/btd* in *F. candida* is very similar to the other insects: there is a metameric expression (Fig. 8C, D), a weak expression in the appendages (Fig. 8D), and expression in the nervous system (Fig. 8E).

There are 3 genes related to *Sp5* in the zebrafish genome. *Sp5* (also known as *bts1*) [Tallafuss et al., 2001], *Sp5-like* (also known as *spr2*) [Zhao et al., 2003] and *similar-to-Sp5*. *Sp5* in zebrafish is expressed in a head stripe along the midbrain-hindbrain boundary, in the otic vesicles, diencephalon, tail bud, and in the somites [Tallafuss et al., 2001]. Zebrafish *Sp5-like* expression is partially overlapping the *Sp5* expression in ectodermal and mesodermal tissue, the brain, trunk neural crest cells, and somites [Zhao et al., 2003]. Mouse *Sp5* is also expressed in a head stripe at the midbrain-hindbrain boundary, in the primitive streak, and later in the tail bud, otic vesicles, limb buds, the developing central nervous system, somites and pharyngeal region [Harrison et al., 2000; Treichel et al., 2001]. In summary, the expression of the genes in this clade are highly similar in the insects and clear similarities also exist to the expression in the vertebrates. This again supports the orthology of the genes in this clade.

The genes of the *Sp6-8* clade: The expression of *D-Sp1* (the *D. melanogaster* representative of the *Sp6-9* clade) has been published previously [Wimmer et al., 1996; Schöck et al., 1999]. The gene is maternally contributed (Fig. 4J), and earliest embryonic expression is seen in the brain (Fig. 4K, L). Later, strong expression is seen in the limb primordia of the antennae and legs (Fig. 4M, N) and in a punctate pattern in the ventral nerve cord (Fig. 4O). The expression of the *T. castaneum Sp8* gene has been reported earlier [Beermann et al., 2004]. Like the *D. melanogaster D-Sp1* gene, the *T. castaneum Sp8* gene is expressed in the brain, ventral nerve cord, and the limb buds (Fig. 5G, H). In the growing legs the gene is expressed in a pattern comprising several rings (Fig. 5H) [Beermann et al., 2004]. The gene *Sp8/9* from *O. fas*ciatus has been published recently [Schaeper et al., 2009]. *Sp8/9* is expressed in the brain, in a punctate pattern in the ventral nerve cord and in the limbs (Fig. 6D). Similar to the legs in older *T. castaneum* embryos, the *O. fasciatus Sp8/9* gene is expressed in several rings in the legs (Fig. 6E). The *Sp6-9* gene from *T. domestica* is expressed in the limb buds (Fig. 7I, J) and later in at least two rings in the legs (Fig. 7K, L). In young segments that have just separated from the growth zone there is a stripe of *Sp6-9* expression and in older segments the gene is expressed in a punctate pattern in the ventral nerve cord. There is also an expression domain in the brain. In the springtail *F. candida* the *Sp6-9* gene is expressed in the brain and in a punctate pattern in the ventral nervous system (Fig. 8F-H). The gene is also expressed in the limb buds (Fig. 8F-H) and at later stages in two separate rings in the legs (Fig. 8I). These data show that the embryonic expression pattern of the *Sp6-9* representatives is very similar in all studied insect species. These similarities extend to the crustaceans as shown by *Sp6-9* expression in *P. hawaiensis*. In this species the gene is expressed in the limb buds (Fig. 9D, E) and at later stages in the peraeopods and in the two branches of the pleopods and the first two pairs of uropods (Fig. 9F). In addition, there is a punctate expression pattern in the ventral nerve cord (Fig. 9F).

Expression data for the members of this clade are also available from vertebrates. Intitial RT-PCR analysis of mouse *Sp6* expression suggested expression in all tissues studied [Scohy et al., 2000], but later studies showed a specific expression pattern in hair follicles and the apical ectodermal ridge (AER) of the developing limbs [Nakamura et al., 2004; Hertveldt et al., 2008]. Consequently, *Sp6* null mice are nude and show defects in skin, teeth, limbs (syndactyly and oligodactyly), and lung alveols. *Sp7* (also known as *osterix*) is so far only documented to be expressed in the osteoblasts. Bone formation fails in *Sp7* deficient mice due to impaired osteoblast differentiation [Nakashima et al., 2002; Milona et al., 2003; Kaback et el., 2008]. Apart from expression domains in the nervous system (brain) both *Sp8* and *Sp9* are predominantly expressed in the AER of the limbs in mouse, chick and zebrafish and are essential for limb and fin outgrowth [Bell et al., 2003; Treichel et al., 2003; Kawakami et al., 2004; Griesel et al., 2006]. In summary, the expression patterns of the genes in this clade are strikingly similar in the insects

and crustaceans and very similar expression patterns also exist in some vertebrate representatives of this clade, again supporting the orthology of the genes in this clade.

Summarizing the available gene expression data it is evident that the gene expression profiles of the arthropod and vertebrate members within each clade are very similar. This lends further support to our notion that the Sp-family genes in the Metazoa fall into three monophyletic clades that each derives from a single ancestral gene from a cluster comprising three genes. The ubiquitous pattern of the *Sp1-4* genes separates them from the *Sp5/btd* and *Sp6-9* genes that display more complex expression patterns frequently comprising at least domains in the nervous system, limbs and segments. This observation fully agrees with our analysis of protein structure that also suggests that the *Sp5/btd* clade and the *Sp6-9* clade form a larger grouping of closely related genes (the *Sp5-9* group), which this suggests that the *Sp5/btd* and *Sp6-9* genes derive from a common ancestral gene.

Conclusions

All available data suggest that a set of three Sp-family genes comprising one *Sp1-4* gene, one *Sp5/btd*, and one *Sp6-9* gene, is ancestral in the Metazoa (Fig. 10). No data are yet available from the most basal metazoan group, the Porifera (sponges), but at least two Sp-family genes are linked in the basal metazoan *T. adhaerens.* This can serve as evidence that the Sp-family triplet formed a small gene cluster already in the basal metazoan (Fig. 10, "metazoan grade"), but it is unclear whether this Sp gene cluster was initially linked to the Hox gene cluster. It is still debated whether *T. adhaerens* has any true Hox genes and, if yes, how many, but at least the *Trox-2* gene appears to be a genuine Hox gene homolog [Srivastava et al. 2008]. Thus, *T. adhaerens* appears to represent an evolutionary grade before the evolution of the Hox gene cluster, but it is yet unclear whether the single *T. adhaerens* Hox gene is physically close to the Sp-family genes.

The Eumetazoa ancestor already possessed a triplet cluster of Sp-family genes (Fig. 10, "eumetazoa grade") as evidenced by the three closely linked Sp genes in the genome of the sea anemone *N. vectensis*. This cnidarian species has eight Hox genes of which only four appear to be clustered. It is debated whether the Cnidaria represent a grade before or after the formation of a true Hox gene cluster, but recent analyses strongly suggest that the ancestral Cnidarian had indeed a genuine Hox gene cluster comprising at least one anterior and one posterior Hox gene [Ryan et al., 2007; Hejnol and Martindale, 2009]. These seem to have been independently duplicated and partially removed from the original cluster during cnidarian evolution leading to the dispersed set of 8 Hox genes in *N. vectensis* [Ryan et al., 2007]. None of these Hox genes in

N. vectensis is on the scaffold that contains the Sp genes, but comparative genomics studies suggest that the four clustered Hox genes and the Sp gene cluster are located next to each other (on the so called "PAL A") [Putnam et al., 2007]. Thus, the Eumetazoa ancestor likely possessed a Sp gene cluster linked to the primordial Hox gene cluster (Fig. 10, "eumetazoan grade").

In the Bilateria the Hox cluster underwent further elaboration by gene duplications, whereas the nearby Sp gene cluster preserved the ancestral number of three genes. Nevertheless, the evolution of the Hox cluster influenced the evolution of the Sp cluster in different ways in different bilaterian lineages. In the insects for example, the Sp gene cluster became partially independent from the Hox gene cluster by the relocation of the *Sp5/btd* and the *Sp6-9* gene (Fig. 10, top right). In the dipterans the *Sp1-4* gene is still linked to the Hox gene cluster, but in other insects (and in the crustacean *D. pulex*) the *Sp1-4* gene appears to have become detached from the Hox gene cluster as well. In the vertebrates, the Hox gene cluster was duplicated several times leading to a total set of four Hox gene clusters [Kappen et al., 1989], and the nearby Sp gene cluster evidently was duplicated along with the Hox gene cluster (Fig. 10, top left). Additional partial genome duplications have occurred in the teleost fishes [Taylor et al., 2001; reviewed in Venkatesh, 2003] likely accounting for the additional Sp genes (e.g. in *D. rerio* and *F. rubripes*). In summary, our results show that the *btd* gene did not originate from a recent duplication in the arthropods (or even in the insects only), but traces back to an ancient *Sp5/Btd* gene already present in basal metazoans.

Methods

Arthropod husbandry, embryo collection and fixation

The *O. fasciatus* (milkweed bug) culture was kept as described in Hughes and Kaufman [2000]. Embryos of all stages were fixed as reported previously [Liu and Kaufman, 2004a]. Dissections of milkweed bug embryos were performed under a fluorescence stereomicroscope using SYTOX Green nucleic acid stain (Invitrogen) before in situ staining [Liu and Kaufman, 2004b]. *T. domestica* (firebrat) were cultured as described in Rogers et al. [1997] with some modifications. Firebrats were kept in plastic containers in an incubator at 36°C and fed with oatmeal. For better handling especially of very young embryos during the dissection procedure, firebrat eggs were first boiled for 1 min in a waterbath and cooled on ice for 1 min. Afterwards, embryos were fixed for 1 h in fixative (4% formaldehyde in phosphate buffered saline and 0,1% Tween-20). Embryos were stained with SYTOX Green nucleic acid stain and dissected as described for *O. fasciatus* [Liu and Kaufman, 2004b]. *F. candida* (white springtail) were raised at room temperature in

plastic containers with a thin layer of plaster mixed with charcoal. Springtail embryos from 0-5 days were collected with a fine brush and put into a 1,5 ml reaction tube filled with 500 µl water. Embryos were boiled for 1 min in a waterbath, cooled on ice for 1 min, then put into a 50 µm mesh net and treated with 50% bleach for 6 min. Afterwards, embroys were washed with water and put into 100% Methanol. These embryos were then sonicated for 45 sec in Methanol, vortexed several times and stored at -20°C until use. *P. hawaiensis* (amphipod beachhopper) were cultured in shallow plastic boxes at 26°C filled with a thin layer of crushed coral substrate and artificial seawater (30g/l of synthetic sea salt) and fed with dry fish flakes twice a week. Membrane pumps ventilated the water. Gravid amphipod females were anaesthesized with clove oil (10µl per 50ml seawater) and embryos were collected out of the brood prouch with forceps. Dissection and fixation was performed as described in Browne et al. [2006].

Gene cloning and sequence analysis

D. melanogaster embryos from 0-20 h, *T. castaneum* embryos from 0-72 h, *O. fasciatus* embryos from 0-96 h, *T. domestica* and *F. candida* embryos from 0-5 days, and *P. hawaiensis* embryos of all described stages [Browne et al., 2005], were used for mRNA isolation using the MicroPoly(A)Purist kit (Ambion). Double-stranded (ds) cDNA and RACE template synthesis was performed using the SMART PCR cDNA Synthesis kit and SMART RACE cDNA Amplification Kit (Clontech). Degenerate primers were designed based on alignments of differerent Sp factor sequences (e.g. *D. melanogaster*, *T. castaneum*, mouse). Sp factors of the different arthropod species used in this study were isolated with different combinations of the following degenerate primers: Fw_GRATCDCPNC (GGC MGG GCI ACI TGY GAY TGY CCI AAY TG), Fw_RCRCPNC (MGI TGY MGI TGY CCI AAY TG), Fw_CHV/IPGCGK (TGY CAY RTI CCI GGI TGY GGI AA), Rev_RSDELQRH (TGI CKY TGI ARY TCR TCI SWI C), Rev_KRFMRSDHL (ARR TGR TCI SWI CKC ATR AAI CKY AA). RACE PCR was performed with specific primers designed on the basis of the results of the degenerate primers PCR. All newly isolated sequences have been submitted to the EMBL Nucleotide Database with the following accession numbers: *Of_Sp1-4* [EMBL: FN562984]*, Td_Sp1-4* [EMBL: FN562988], *Td_Sp5/btd* [EMBL: FN562989], *Td_Sp6-9* [EMBL: FN562990], *Fc_Sp1-4* [EMBL: FN562985], *Fc_Sp5/btd* [EMBL: FN562986], *Fc_Sp6-9* [EMBL: FN562987], *Ph_Sp1-4* [EMBL: FN562991]*, Ph_Sp6-9* [EMBL: FN562992]. BLAST analysis was used to identify the *Sp1-4* homologue of *D. melanogaster* and *T. castaneum*. Gene specific primers were made to amplify *Tc_btd* [GenBank: NM_001114320.1], *Tc_Sp8* [GenBank: NM_001039420] and *Tc_Sp1-4* [GenBank: XM_967159] from *T. castaneum* cDNA, as well as *Dm_btd* [GenBank: NM_078545], *Dm_D-Sp1* [GenBank: NM_132351] and *Dm_CG5669* [GenBank: NM_142975] from *D. melanogaster* cDNA. Primer sequences are available upon request. We have used the publicly available genome sequencing

data for a selection of metazoan species: *H. sapiens* [International Human Genome Sequencing Consortium, 2001; Venter et al., 2001], *M. musculus* [Mouse Sequencing Consortium, 2002], *N. vitripennis* [Nasonia Genome Project Web site], *D. melanogaster* [Adams et al., 2000], *D. pseudobscura* [Richards et al., 2005], *A. mellifera* [Honeybee Genome Sequencing Consortium, 2006], *A. gambiae* [Holt et al., 2002], *T. castaneum* [Tribolium Genome Sequencing Consortium, 2008], *B. mori* [International Silkworm Genome Consortium, 2008], *D. pulex* [Daphnia Genomics Consortium Web site], *S. purpuratus* [Sea Urchin Genome Sequencing Consortium, 2006], *N. vectensis* [Putnam et al., 2007], *G. gallus* [International Chicken Genome Sequencing Consortium, 2004], *F. rubripes* [Aparicio et al., 2002], *D. rerio* [Zebrafish Sequencing Project Web site], *B. floridae* [Putnam et al. 2008], and *T. adhaerens* [Srivastava et al., 2008]. Phylogenetic analysis of different Sp transcription factor sequences was performed as described in Prpic et al. [2005]. The accession numbers of previously published sequences used in the phylogenetic analysis are as follows: *Dm*_CG5669 [GenBank: NP_651232], *Dm*_Btd [GenBank: NP_511100], *Dm*_D-Sp1 [GenBank: NP_572579], *Dps*_GA19045 [GenBank: XP_001358829], *Dps*_GA22354 [GenBank: XP_002134535], *Dps*_GA12282 [GenBank: XP_001354397], *Ag*_Sp1-4 [GenBank: NZ_AAAB02008898], *Ag*_Sp5/Btd [GenBank: NZ_AAAB02008847], *Ag* Sp6-9 [GenBank: NZ_AAAB01008847]; *Nav*_Sp1-4 [GenBank: XP_001599101], *Nav*_Sp5/Btd [GenBank: AAZX01008599], *Nav*_Sp6-9 [GenBank: XP_001606079], *Am*_Sp1-4 [GenBank: XP_624316.2], *Am*_Sp5/Btd [GenBank: XP_001119912], *Am*_Sp6-9 [GenBank: XP_624528], *Bm*_Sp1-4 [GenBank: BABH01010251], *Bm*_Sp5/Btd [GenBank: BABH01024462], *Bm*_Sp6-9 [GenBank: AADK01002198], *Tc*_Sp1-4 [GenBank: XP_972252], *Tc*_Btd [GenBank: NP_001107792], *Tc*_Sp8 [GenBank: NP_001034509], *Of*_Sp8/9 [EMBL: FN396612], *Nv*_Sp1-4 [GenBank: XP_001635004], *Nv*_Sp5/Btd [GenBank: XP_001635002], *Nv*_Sp6-9 [GenBank: XP_001634948], *Sp*_Sp1-4 [GenBank: XR_025838], *Sp*_Sp5/Btd [GenBank: XP_789110.1], *Sp*_Sp6-9 [GenBank: XP_793203.2], *Hs*_Sp1 [GenBank: NP_612482], *Hs*_Sp2 [GenBank :NP_003101], *Hs*_Sp3 [GenBank: NP_003102], *Hs*_Sp4 [GenBank: NP_003103], *Hs*_Sp5 [GenBank: NP_001003845], *Hs*_Sp6 [GenBank: NP_954871], *Hs*_Sp7 [GenBank: NP_690599], *Hs*_Sp8 [GenBank :NP_874359], *Hs*_Sp9 [GenBank: NP_001138722], *Mm*_Sp1 [GenBank: NP_038700], *Mm*_Sp2 [GenBank: NP_084496], *Mm*_Sp3 [GenBank: NP_035580], *Mm*_Sp4 [GenBank: NP_033265], *Mm*_Sp5 [GenBank: NP_071880], *Mm*_Sp6 [GenBank: NP_112460], *Mm*_Sp7 [GenBank: NP_569725], *Mm*_Sp8 [GenBank: NP_796056], *Mm*_Sp9 [GenBank: NP_001005343], *Dr*_Sp1 [GenBank: NP_997827], *Dr*_Sp2 [GenBank: NP_001093452], *Dr*_Sp3 [GenBank: NP_001082967], *Dr*_Sp3-like [GenBank: XP_691096], *Dr*_Sp4 [GenBank: NP_956418], *Dr*_Sp5 [GenBank: NP_851304], *Dr*_Sp5-like [GenBank: NP_919352], *Dr*_Similar_to_Sp5 [GenBank: XP_001335730], *Dr*_Sp6 [GenBank: NP_991195], *Dr*_Sp7 [GenBank: NP_998028], *Dr*_Sp8 [GenBank: NP_998406], *Dr*_Sp8-like [GenBank: NP_991113], *Dr*_Sp9 [GenBank: NP_998125], *Gg*_Sp1 [GenBank: NP_989935], *Gg*_Sp2 [GenBank: XP_423405], *Gg*_Sp3 [GenBank: NP_989934], *Gg*_Sp4 [GenBank: XP_418708], *Gg*_Sp5

[GenBank: NP_001038149], *Gg*_Sp8 [GenBank: AAU04515.1], *Gg*_Sp9 [GenBank: AAU04516.1], Fr_Sp1 [GenBank: CAAB01000453.1], Fr_Sp2 [GenBank: CAAB01001586.1], Fr_Sp3 [GenBank: CAAB01000508.1], Fr_Sp3-like [GenBank: CAAB01000254.1], Fr_Sp4 [GenBank: CAAB01001019.1], Fr_Sp5 [GenBank: CAAB01001064.1], Fr_Sp5-like [GenBank: CAAB01000006.1], Fr_Sp6 [GenBank: CAAB01004244.1], Fr_Sp7 [GenBank: CAAB01000453.1], Fr_Sp8 [GenBank: CAAB01001019.1], Fr_Sp9 [GenBank: CAAB01000508.1]. In addition, we have provisionally annotated the Sp-family genes of *D. pulex, T. adhaerens* and *B. floridae* using the following genomic regions: *Dp_Sp1-4* [NCBI_GNO_320154, scaffold_15:792601, 795915], *Dp_Sp5/btd* [NCBI_GNO_60744, scaffold_130:263041, 265220], *Dp_Sp6-9* [NCBI_GNO_424374, scaffold_42:102959, 162646], *Ta_Sp1-4* [scaffold_3:4169974, 4284735], *Ta_Sp5/btd* [scaffold_15:1197089, 1197409], *Ta_Sp6-9* [scaffold_15:1120368, 1120718], *Bf_Sp1-4* [Bf_V2_288:2820, 5436], *Bf_Sp5/btd* [Bf_V2_149:860371, 860057], *Bf_Sp6-9* [Bf_V2_149:758897, 759229].

In situ hybridization

The length of the templates used for digoxygenin labeled RNA probe synthesis is given in the following listing. *Dm_Sp1-4*: 872bp, *Dm_btd:* 869bp, *Dm_D-Sp1:* 718bp, *Tc_S*p1-4: 748bp, *Tc_btd:* 509bp, *Tc_Sp8:* 868bp, *Of_Sp1-4:* 594bp, *Of_Sp8/9:* 1078bp, *Td_Sp1-4:* 647bp, *Td_Sp5/btd:* 774bp, *Td_Sp6-9:* 772bp, *Fc_Sp1-4:* 737bp, *Fc_Sp5/btd:* 954bp, *Fc_Sp6-9:* 736bp, *Ph_Sp1-4:* 1799bp, *Ph_Sp6-9:* 1502bp. *D. melanogaster* and *T. castaneum in situ* was performed essentially as described in Tautz and Pfeifle [1989], *O. fasciatus in situ* hybridization was done according to Liu and Kaufman [2004a], *P. hawaiensis in situ* was performed as reported in Browne et al. [2006], and *in situ* hybridizations for *F. candida* and *T. domestica* were done essentially as described in Hughes et al. [2004].

List of abbreviations

The abbreviations used to denominate the different species are as follows: *Ag*, *Anopheles gambiae*; *Am*, *Apis mellifera*; *Bf*, *Branchiostoma floridae*; *Bm*, *Bombyx mori*; *Dm*, *Drosophila melanogaster*; *Dp*, *Daphnia pulex*; *Dps*, *Drosophila pseudobscura*; *Dr*, *Danio rerio*; *Fc*, *Folsomia candida*; *Fr*, *Fugu rubripes*; *Gg*, *Gallus gallus*; *Hs*, *Homo sapiens*; *Mm*, *Mus musculus*; *Nav*, *Nasonia vitripennis*; *Nv*, *Nematostella vectensis*; *Of*, *Oncopeltus fasciatus*; *Ph*, *Parhyale hawaiensis*; *Sp*, *Strongylocentrotus purpuratus*; *Ta*, *Trichoplax adhaerens*; *Tc*, *Tribolium castaneum*; *Td*, *Thermobia domestica*.

Authors´ contributions

NDS carried out the molecular genetic studies, genome location and protein domain analyses, and embryological work, and drafted the manuscript. NMP performed the phylogenetic sequence analysis and helped to draft the manuscript. EAW and NMP helped in the analysis of the data and participated in the design and coordination of the study. EAW conceived of the study. All authors read and approved the final manuscript.

Acknowledgements

This work has been funded by the European Community's Marie Curie Research Training Network ZOONET under contract MRTN-CT-2004-005624 (to EAW), and a grant from the Deutsche Forschungsgemeinschaft (PR 1109/1-1 to NMP).

References

Abbasi AA, Grzeschik KH: **An insight into the phylogenetic history of HOX linked gene families in vertebrates.** *BMC Evol Biol* 2007, **30;7:**239.

Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PG, Scherer SE, Li PW, Hoskins RA, Galle RF, George RA, Lewis SE, Richards S, Ashburner M, Henderson SN, Sutton GG, Wortman JR, Yandell MD, Zhang Q, Chen LX, Brandon RC, Rogers YH, Blazej RG, Champe M, Pfeiffer BD, Wan KH, Doyle C, Baxter EG, Helt G, Nelson CR, Gabor GL, Abril JF, Agbayani A, An HJ, Andrews-Pfannkoch C, Baldwin D, Ballew RM, Basu A, Baxendale J, Bayraktaroglu L, Beasley EM, Beeson KY, Benos PV, Berman BP, Bhandari D, Bolshakov S, Borkova D, Botchan MR, Bouck J, Brokstein P, Brottier P, Burtis KC, Busam DA, Butler H, Cadieu E, Center A, Chandra I, Cherry JM, Cawley S, Dahlke C, Davenport LB, Davies P, de Pablos B, Delcher A, Deng Z, Mays AD, Dew I, Dietz SM, Dodson K, Doup LE, Downes M, Dugan-Rocha S, Dunkov BC, Dunn P, Durbin KJ, Evangelista CC, Ferraz C, Ferriera S, Fleischmann W, Fosler C, Gabrielian AE, Garg NS, Gelbart WM, Glasser K, Glodek A, Gong F, Gorrell JH, Gu Z, Guan P, Harris M, Harris NL, Harvey D, Heiman TJ, Hernandez JR, Houck J, Hostin D, Houston KA, Howland TJ, Wei MH, Ibegwam C, Jalali M, Kalush F, Karpen GH, Ke Z, Kennison JA, Ketchum KA, Kimmel BE, Kodira CD, Kraft C, Kravitz S, Kulp D, Lai Z, Lasko P, Lei Y, Levitsky AA, Li J, Li Z, Liang Y, Lin X, Liu X, Mattei B, McIntosh TC, McLeod MP, McPherson D, Merkulov G, Milshina NV, Mobarry C, Morris J, Moshrefi A, Mount SM, Moy M,

Murphy B, Murphy L, Muzny DM, Nelson DL, Nelson DR, Nelson KA, Nixon K, Nusskern DR, Pacleb JM, Palazzolo M, Pittman GS, Pan S, Pollard J, Puri V, Reese MG, Reinert K, Remington K, Saunders RD, Scheeler F, Shen H, Shue BC, Sidén-Kiamos I, Simpson M, Skupski MP, Smith T, Spier E, Spradling AC, Stapleton M, Strong R, Sun E, Svirskas R, Tector C, Turner R, Venter E, Wang AH, Wang X, Wang ZY, Wassarman DA, Weinstock GM, Weissenbach J, Williams SM, WoodageT, Worley KC, Wu D, Yang S, Yao QA, Ye J, Yeh RF, Zaveri JS, Zhan M, Zhang G, Zhao Q, Zheng L, Zheng XH, Zhong FN, Zhong W, Zhou X, Zhu S, Zhu X, Smith HO, Gibbs RA, Myers EW, Rubin GM, Venter JC: **The genome sequence of *Drosophila melanogaster*.** *Science* 2000, **287:**2185-2195.

Aparicio S, Chapman J, Stupka E, Putnam N, Chia JM, Dehal P, Christoffels A, Rash S, Hoon S, Smit A, Gelpke MD, Roach J, Oh T, Ho IY, Wong M, Detter C, Verhoef F, Predki P, Tay A, Lucas S, Richardson P, Smith SF, Clark MS, Edwards YJ, Doggett N, Zharkikh A, Tavtigian SV, Pruss D, Barnstead M, Evans C, Baden H, Powell J, Glusman G, Rowen L, Hood L, Tan YH, Elgar G, Hawkins T, Venkatesh B, Rokhsar D, Brenner S: **Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*.** *Science* 2002, **297:**1301-1310.

Bailey WJ, Kim J, Wagner GP, Ruddle FH: **Phylogenetic reconstruction of vertebrate hox cluster duplications.** *Mol Biol Evol* 1997, **14:**843-853.

Beermann A, Aranda M, Schröder R: **The *Sp8* zinc-finger transcription factor is involved in allometric growth of the limbs in the beetle *Tribolium castaneum*.** *Development* 2004, **131:**733-742.

Bell SM, Schreiner CM, Waclaw RR, Campbell K, Potter SS, Scott WJ: ***Sp8* is crucial for limb outgrowth and neuropore closure.** *Proc Natl Acad Sci USA* 2003, **100:**12195-12200.

Black AR, Jensen D, Lin SY, Azizkhan-Clifford J: **Growth/Cell Cycle Regulation of Sp1 Phosphorylation.** *Journ of Biol Chem* 1999, **274:**1207–1215.

Black AR, Black JD, Azizkhan-Clifford J: **Sp1 and krüppel-like factor family of transcription factors in cell growth regulation and cancer.** *J Cell Physiol* 2001, **188:**143-160.

Bouwman P, Göllner H, Elsässer HP, Eckhoff G, Karis A, Grosveld F, Philipsen S, Suske G: **Transcription factor Sp3 is essential for post-natal survival and late tooth development.** *EMBO J* 2000, **19:**655-661.

Bouwman P, Philipsen S: **Regulation of the activity of Sp1-related transcription factors.** *Mol Cell Endocrinol* 2002, **95:**27-38.

Browne WE, Schmid BGM, Wimmer EA, Martindale MQ: **Expression of *otd* orthologs in the amphipod crustacean, *Parhyale hawaiensis.*** *Dev Genes Evol* 2006, **216:**581–595.

Browne WE, Price AL, Gerberding M, Patel NH: **Stages of embryonic development in the amphipod crustacean, *Parhyale hawaiensis*.** *Genesis* 2005, **42:**124-149.

Budd GE, Telford MJ: **The origin and evolution of arthropods.** *Nature* 2009, **457:**812-7.

Chen Y, Guo Y, Ge X, Itoh H, Watanabe A, Fujiwara T, Kodama T, Aburatani H: **Elevated expression and potential roles of human Sp5, a member of Sp transcription factor family, in human cancers.** *Biochem Biophys Res Commun* 2006, **340:**758-766.

Cohen SM, Jürgens G: **Mediation of *Drosophila* head development by gap-like segmentation genes.** *Nature* 1990, **346:**482-485.

Daphnia Genomics Consortium [http://daphnia.cgb.indiana.edu/]

Dohle W.: **Myriapod-insect relationships as opposed to an insect crustacean sister group relationship**. *Arthropod Relationships* (Eds. Fortey RA, Thomas RH) 1997, pp 305-315. Chapman & Hall, London.

Dohle W: **Are the insects terrestrial crustaceans? A discussion of some new facts and arguments and the proposal of the proper name 'Tetraconata' for the monophyletic unit Crustacea plus Hexapoda.** *Ann Soc Entomol France NS* 2001, **37:**85-103.

Dynan WS, Tjian R: **The promoter-specific transcription factor Sp1 binds to upstream sequences in the SV40 early promoter.** *Cell* 1983a, **35:**79-87.

Dynan WS, Tjian R: **Isolation of transcription factors that discriminate between different promoters recognized by RNA polymerase II.** *Cell* 1983b, **32:**669-680.

Friedrich M, Tautz D: **Ribosomal DNA phylogeny of the major extant arthropod classes and the evolution of myriapods.** *Nature* 1995, **376:**165-7.

Göllner H, Dani C, Phillips B, Philipsen S, Suske G: **Impaired ossification in mice lacking the transcription factor Sp3.** *Mech Dev* 2001a, **106:**77-83.

Göllner H, Bouwman P, Mangold M, Karis A, Braun H, Rohner I, Del Rey A, Besedovsky HO, Meinhardt A, van den Broek M, Cutforth T, Grosveld F, Philipsen S, Suske G: **Complex phenotype of mice homozygous for a null mutation in the Sp4 transcription factor gene.** *Genes Cells* 200 b, **6:**689-697.

Griesel G, Treichel D, Collombat P, Krull J, Zembrzycki A, van den Akker WM, Gruss P, Simeone A, Mansouri A. ***Sp8*** **controls the anteroposterior patterning at the midbrain-hindbrain border.** *Development* 2006, **133:**1779-1787.

Harrison SM, Houzelstein D, Dunwoodie SL, Beddington RS: ***Sp5*****, a new member of the *Sp1* family, is dynamically expressed during development and genetically interacts with *Brachyury*.** *Dev Biol* 2000, **227:**358-372.

Hejnol A, Martindale MQ: **Coordinated spatial and temporal expression of Hox genes during embryogenesis in the acoel *Convolutriloba longifissura*.** *BMC Biol* 2009, **7:**65.

Hertveldt V, Louryan S, van Reeth T, Drèze P, van Vooren P, Szpirer J, Szpirer C: **The development of several organs and appendages is impaired in mice lacking Sp6.** *Dev Dyn* 2008, **237:**883-892.

Holland LZ, Albalat R, Azumi K, Benito-Gutiérrez E, Blow MJ, Bronner-Fraser M, Brunet F, Butts T, Candiani S, Dishaw LJ, Ferrier DE, Garcia-Fernàndez J, Gibson-Brown JJ, Gissi C, Godzik A, Hallböök F, Hirose D, Hosomichi K, Ikuta T, Inoko H, Kasahara M, Kasamatsu J, Kawashima T, Kimura A, Kobayashi M, Kozmik Z, Kubokawa K, Laudet V, Litman GW, McHardy AC, Meulemans D, Nonaka M, Olinski RP, Pancer Z, Pennacchio LA, Pestarino M, Rast JP, Rigoutsos I, Robinson-Rechavi M, Roch G, Saiga H, Sasakura Y, Satake M, Satou Y, Schubert M, Sherwood N, Shiina T, Takatori N, Tello J, Vopalensky P, Wada S, Xu A, Ye Y, Yoshida K, Yoshizaki F, Yu JK, Zhang Q, Zmasek CM, de Jong PJ, Osoegawa K, Putnam NH, Rokhsar DS, Satoh N, Holland PW: **The amphioxus genome illuminates vertebrate origins and cephalochordate biology.** *Genome Res* 2008, **18:**1100-1111.

Holt RA, Subramanian GM, Halpern A, Sutton GG, Charlab R, Nusskern DR, Wincker P, Clark AG, Ribeiro JM, Wides R, Salzberg SL, Loftus B, Yandell M, Majoros WH, Rusch DB, Lai Z, Kraft CL, Abril JF, Anthouard V, Arensburger P, Atkinson PW, Baden H, de Berardinis V,

Baldwin D, Benes V, Biedler J, Blass C, Bolanos R, Boscus D, Barnstead M, Cai S, Center A, Chaturverdi K, Christophides GK, Chrystal MA, Clamp M, Cravchik A, Curwen V, Dana A, Delcher A, Dew I, Evans CA, Flanigan M, Grundschober-Freimoser A, Friedli L, Gu Z, Guan P, Guigo R, Hillenmeyer ME, Hladun SL, Hogan JR, Hong YS, Hoover J, Jaillon O, Ke Z, Kodira C, Kokoza E, Koutsos A, Letunic I, Levitsky A, Liang Y, Lin JJ, Lobo NF, Lopez JR, Malek JA, McIntosh TC, Meister S, Miller J, Mobarry C, Mongin E, Murphy SD, O'Brochta DA, Pfannkoch C, Qi R, Regier MA, Remington K, Shao H, Sharakhova MV, Sitter CD, Shetty J, Smith TJ, Strong R, Sun J, Thomasova D, Ton LQ, Topalis P, Tu Z, Unger MF, Walenz B, Wang A, Wang J, Wang M, Wang X, Woodford KJ, Wortman JR, Wu M, Yao A, Zdobnov EM, Zhang H, Zhao Q, Zhao S, Zhu SC, Zhimulev I, Coluzzi M, della Torre A, Roth CW, Louis C, Kalush F, Mural RJ, Myers EW, Adams MD, Smith HO, Broder S, Gardner MJ, Fraser CM, Birney E, Bork P, Brey PT, Venter JC, Weissenbach J, Kafatos FC, Collins FH, Hoffman SL: **The genome sequence of the malaria mosquito *Anopheles gambiae*.** *Science* 2002, **298:**129-149.

Honeybee Genome Sequencing Consortium: **Insights into social insects from the genome of the honeybee Apis mellifera.** *Nature* 2006, **443:**931-949.

Howard-Ashby M, Materna SC, Brown CT, Chen L, Cameron RA, Davidson EH: **Identification and characterization of homeobox transcription factor genes in *Strongylocentrotus purpuratus*, and their expression in embryonic development.** *Dev Biol* 2006, **300:**74-89.

Hughes CL, Kaufman TC: **RNAi analysis of Deformed, proboscipedia and Sex combs reduced in the milkweed bug Oncopeltus fasciatus: novel roles for Hox genes in the Hemipteran head.** *Development* 2000, **127:**3683-3694.

Hughes CL, Liu PZ, Kaufman TC: **Expression patterns of the rogue Hox genes *Hox3/zen* and *fushi tarazu* in the apterygote insect *Thermobia domestica*.** *Evolution and Development* 2004, **6:**393-401.

International Chicken Genome Sequencing Consortium: **Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution.** *Nature* 2004, **432:**695-716.

International Human Genome Sequencing Consortium: **Initial sequencing and analysis of the human genome.** *Nature* 2001, **409:**860-921.

International Silkworm Genome Consortium: **The genome of a lepidopteran model insect, the silkworm *Bombyx mori*.** *Insect Biochem Mol Biol* 2008, **38:**1036-1045.

Kaback LA, Soung do Y, Naik A, Smith N, Schwarz EM, O'Keefe RJ, Drissi H: **Osterix/Sp7 regulates mesenchymal stem cell mediated endochondral ossification.** *J Cell Physiol* 2008, **214:**173-182.

Kadonaga JT, Carner KR, Masiarz FR, Tjian R: **Isolation of cDNA encoding transcription factor Sp1 and functional analysis of the DNA binding domain.** *Cell* 1987, **51:**1079-1090.

Kappen C, Schughart K, Ruddle FH: **Two steps in the evolution of Antennapedia-class vertebrate homeobox genes.** *Proc Natl Acad Sci USA* 1989, **86:**5459-5463.

Kawakami Y, Esteban CR, Matsui T, Rodríguez-León J, Kato S, Izpisúa Belmonte JC: ***Sp8* and *Sp9*, two closely related *buttonhead*-like transcription factors, regulate *Fgf8* expression and limb outgrowth in vertebrate embryos.** *Development*. 2004, **131:**4763-4774.

Lemke S, Stauber M, Shaw PJ, Rafiqi AM, Prell A, Schmidt-Ott U: **Bicoid occurrence and Bicoid-dependent hunchback regulation in lower cyclorrhaphan flies.** *Evol Dev* 2008, **10:**413-420.

Liu PZ, Kaufman TC: ***hunchback* is required for suppression of abdominal identity, and for proper germband growth and segmentation in the intermediate germband insect *Oncopeltus fasciatus*.** *Development* 2004a, **131:**1515-1527.

Liu PZ, Kaufman TC: **Krüppel is a gap gene in the intermediate germband insect *Oncopeltus fasciatus* and is required for development of both blastoderm and germband-derived segments.** *Development* 2004b, **131:**4567-4579.

Materna SC, Howard-Ashby M, Gray RF, Davidson EH: **The C2H2 zinc finger genes of *Strongylocentrotus purpuratus* and their expression in embryonic development.** *Dev Biol* 2006, **300:**108-120.

Marin M, Karis A, Visser P, Grosveld F, Philipsen S: **Transcription factor Sp1 is essential for early embryonic development but dispensable for cell growth and differentiation.** *Cell* 1997, **89:**619-267.

Milona MA, Gough JE, Edgar AJ: **Expression of alternatively spliced isoforms of human Sp7 in osteoblast-like cells.** *BMC Genomics* 2003, **4:**43.

Mouse Genome Sequencing Consortium: **Initial sequencing and comparative analysis of the mouse genome.** *Nature* 2002, **420:**520-562.

Nakamura T, Unda F, de-Vega S, Vilaxa A, Fukumoto S, Yamada KM, Yamada Y: **The Krüppel-like factor epiprofin is expressed by epithelium of developing teeth, hair follicles, and limb buds and promotes cell proliferation.** *J Biol Chem* 2004, **279:**626-634.

Nakashima K, Zhou X, Kunkel G, Zhang Z, Deng JM, Behringer RR, de Crombrugghe B: **The novel zinc finger-containing transcription factor Osterix is required for osteoblast differentiation and bone formation.** *Cell* 2002, **108:**17–29.

Nasonia Genome Project [http://www.hgsc.bcm.tmc.edu/project-species-i-Nasonia%20vitripennis.hgsc?pageLocation=Nasonia%20vitripennis]

Philipsen S, Suske G: **A tale of three fingers: the family of mammalian Sp/XKLF transcription factors.** *Nucleic Acids Res* 1999, **27:**2991-3000.

Prpic NM, Janssen R, Damen WGM, Tautz D: **Evolution of dorsal–ventral axis formation in arthropod appendages: *H15* and *optomotor-blind/bifid*-type T-box genes in the millipede *Glomeris marginata* (Myriapoda: Diplopoda).** *Evol Dev* 2005, **7:**51-57.

Putnam NH, Srivastava M, Hellsten U, Dirks B, Chapman J, Salamov A, Terry A, Shapiro H, Lindquist E, Kapitonov VV, Jurka J, Genikhovich G, Grigoriev IV, Lucas SM, Steele RE, Finnerty JR, Technau U, Martindale MQ, Rokhsar DS : **Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization.** *Science* 2007, **317:**86-94.

Putnam NH, Butts T, Ferrier DE, Furlong RF, Hellsten U, Kawashima T, Robinson-Rechavi M, Shoguchi E, Terry A, Yu JK, Benito-Gutiérrez EL, Dubchak I, Garcia-Fernàndez J, Gibson-Brown JJ, Grigoriev IV, Horton AC, de Jong PJ, Jurka J, Kapitonov VV, Kohara Y, Kuroki Y, Lindquist E, Lucas S, Osoegawa K, Pennacchio LA, Salamov AA, Satou Y, Sauka-Spengler T, Schmutz J, Shin-I T, Toyoda A, Bronner-Fraser M, Fujiyama A, Holland LZ, Holland PW, Satoh N, Rokhsar DS: **The amphioxus genome and the evolution of the chordate karyotype.** *Nature* 2008, **453:**1064-1071.

Richards S, Liu Y, Bettencourt BR, Hradecky P, Letovsky S, Nielsen R, Thornton K, Hubisz MJ, Chen R, Meisel RP, Couronne O, Hua S, Smith MA, Zhang P, Liu J, Bussemaker HJ, van Batenburg MF, Howells SL, Scherer SE, Sodergren E, Matthews BB, Crosby MA, Schroeder AJ,

Ortiz-Barrientos D, Rives CM, Metzker ML, Muzny DM, Scott G, Steffen D, Wheeler DA, Worley KC, Havlak P, Durbin KJ, Egan A, Gill R, Hume J, Morgan MB, Miner G, Hamilton C, Huang Y, Waldron L, Verduzco D, Clerc-Blankenburg KP, Dubchak I, Noor MA, Anderson W, White KP, Clark AG, Schaeffer SW, Gelbart W, Weinstock GM, Gibbs RA: **Comparative genome sequencing of *Drosophila pseudoobscura*: chromosomal, gene, and cis-element evolution.** *Genome Res* 2005, **15:**1-18.

Rogers BT, Peterson MD, Kaufman TC: **Evolution of the insect body plan as revealed by the *Sex combs reduced* expression pattern.** *Development* 1997, **124:**149-157.

Ryan JF, Mazza ME, Pang K, Matus DQ, Baxevanis AD, Martindale MQ, Finnerty JR: **Pre-bilaterian origins of the hox cluster and the hox code: evidence from the sea anemone, *Nematostella vectensis*.** *PLoS One* 2007, **2:**e153.

Safe S, Abdelrahim M: **Sp transcription factor family and its role in cancer.** *Eur J Cancer* 2005, **41:**2438-2448.

Saffer JD, Jackson SP, Annarella MB: **Developmental expression of *Sp1* in the mouse.** *Mol Cell Biol* 1991, **11:**2189-2199.

Schierwater B, Kamm K, Srivastava M, Rokhsar D, Rosengarten RD, Dellaporta SL: **The early ANTP gene repertoire: insights from the placozoan genome.** *PLoS One* 2008, **3:**e2457.

Schinko JB, Kreuzer N, Offen N, Posnien N, Wimmer EA, Bucher G: **Divergent functions of *orthodenticle*, *empty spiracles* and *buttonhead* in early head patterning of the beetle *Tribolium castaneum* (Coleoptera).** *Dev Biol* 2008, **317:**600-613.

Schöck F, Purnell BA, Wimmer EA, Jäckle H: **Common and diverged functions of the *Drosophila* gene pair *D-Sp1* and *buttonhead*.** *Mech Dev* 1999, **89:**125-132.

Scohy S, Gabant P, Van Reeth T, Hertveldt V, Drèze PL, Van Vooren P, Rivière M, Szpirer J, Szpirer C: **Identification of KLF13 and KLF14 (SP6), novel members of the SP/XKLF transcription factor family.** *Genomics* 2000, **70:**93-101.

Sea Urchin Genome Sequencing Consortium: **The genome of the sea urchin *Strongylocentrotus purpuratus*.** *Science* 2006, **314:**941-952.

Shimeld SM: **C2H2 zinc finger gene of the Gli, Zic, KLF, SP, Wilms' tumor, Huckebein, Snail, ovo, Spalt, Odd, Blimp-1, Fez and related geen families from *Branchiostoma floridae*.** *Dev Genes Evol* 2008, **218:**639-649.

Srivastava M, Begovic E, Chapman J, Putnam NH, Hellsten U, Kawashima T, Kuo A, Mitros T, Salamov A, Carpenter ML, Signorovitch AY, Moreno MA, Kamm K, Grimwood J, Schmutz J, Shapiro H, Grigoriev IV, Buss LW, Schierwater B, Dellaporta SL, Rokhsar DS: **The *Trichoplax* genome and the nature of placozoans.** *Nature* 2008, **7207:**955-960.

Stauber M, Prell A, Schmidt-Ott U: **A single *Hox3* gene with composite *bicoid* and *zerknüllt* expression characteristics in non-Cyclorrhaphan flies.** *Proc Natl Acad Sci USA* 2002, **99:**274-279.

Strimmer K, von Haeseler A: **Quartet puzzling: a quartet maximum likelihood method for reconstructing tree topologies.** *Mol Biol Evol* 1996, **13:**964-969.

Supp DM, Witte DP, Branford WW, Smith EP, Potter SS: **Sp4, a member of the Sp1-family of zinc finger transcription factors, is required for normal murine growth, viability, and male fertility.** *Dev Biol* 1996, **176:**284-299.

Suske G: **The Sp-family of transcription factors.** *Gene* 1999, **238:**291-300.

Suske G, Bruford E, Philipsen S: **Mammalian SP/KLF transcription factors: bring in the family.** *Genomics* 2005, **85:**551-556.

Takatori N, Butts T, Candiani S, Pestarino M, Ferrier DE, Saiga H, Holland PW: **Comprehensive survey and classification of homeobox genes in the genome of amphioxus, *Branchiostoma floridae*.** *Dev Genes Evol* 2008, **218:**579-590.

Tallafuss A, Wilm TP, Crozatier M, Pfeffer P, Wassef M, Bally-Cuif L: **The zebrafish buttonhead-like factor Bts1 is an early regulator of *pax2.1* expression during mid-hindbrain development.** *Development* 2001, **128:**4021-4034.

Tautz D, Pfeifle C: **A non-radioactive in situ hybridization method for the localization of specific RNAs in *Drosophila* embryos reveals translational control of the segmentation gene *hunchback*.** *Chromosoma* 1989, **98:**81-85.

Taylor JS, Van der Peer Y, Braasch I, Meyer A: **Comparative genomics provides evidence for an ancient genome duplication event in fish.** *Phil Trans R Soc Lond B* 2001, **356:**1661-1679.

Treichel D, Becker MB, Gruss P: **The novel transcription factor gene *Sp5* exhibits a dynamic and highly restricted expression pattern during mouse embryogenesis.** *Mech Dev* 2001, **101:**175-179.

Treichel D, Schöck F, Jäckle H, Gruss P, Mansouri A: ***mBtd* is required to maintain signaling during murine limb development.** *Genes Dev* 2003, **17:**2630-2635.

Tribolium Genome Sequencing Consortium: **The genome of the model beetle and pest *Tribolium castaneum*.** *Nature* 2008, **452:**949-955.

Venkatesh B: **Evolution and diversity of fish genomes.** *Curr Opin Genet Dev* 2003, **13:**588-592.

Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, Ballew RM, Huson DH, Wortman JR, Zhang Q, Kodira CD, Zheng XH, Chen L, Skupski M, Subramanian G, Thomas PD, Zhang J, Gabor Miklos GL, Nelson C, Broder S, Clark AG, Nadeau J, McKusick VA, Zinder N, Levine AJ, Roberts RJ, Simon M, Slayman C, Hunkapiller M, Bolanos R, Delcher A, Dew I, Fasulo D, Flanigan M, Florea L, Halpern A, Hannenhalli S, Kravitz S, Levy S, Mobarry C, Reinert K, Remington K, Abu-Threideh J, Beasley E, Biddick K, Bonazzi V, Brandon R, Cargill M, Chandramouliswaran I, Charlab R, Chaturvedi K, Deng Z, Di Francesco V, Dunn P, Eilbeck K, Evangelista C, Gabrielian AE, Gan W, Ge W, Gong F, Gu Z, Guan P, Heiman TJ, Higgins ME, Ji RR, Ke Z, Ketchum KA, Lai Z, Lei Y, Li Z, Li J, Liang Y, Lin X, Lu F, Merkulov GV, Milshina N, Moore HM, Naik AK, Narayan VA, Neelam B, Nusskern D, Rusch DB, Salzberg S, Shao W, Shue B, Sun J, Wang Z, Wang A, Wang X, Wang J, Wei M, Wides R, Xiao C, Yan C, Yao A, Ye J, Zhan M, Zhang W, Zhang H, Zhao Q, Zheng L, Zhong F, Zhong W, Zhu S, Zhao S, Gilbert D, Baumhueter S, Spier G, Carter C, Cravchik A, Woodage T, Ali F, An H, Awe A, Baldwin D, Baden H, Barnstead M, Barrow I, Beeson K, Busam D, Carver A, Center A, Cheng ML, Curry L, Danaher S, Davenport L, Desilets R, Dietz S, Dodson K, Doup L, Ferriera S, Garg N, Gluecksmann A, Hart B, Haynes J, Haynes C, Heiner C, Hladun S, Hostin D, Houck J, Howland T, Ibegwam C, Johnson J, Kalush F, Kline L, Koduru S, Love A, Mann F, May D, McCawley S, McIntosh T, McMullen I, Moy M, Moy L, Murphy B, Nelson K, Pfannkoch C, Pratts E, Puri V, Qureshi H, Reardon M, Rodriguez R, Rogers YH, Romblad D, Ruhfel B, Scott R, Sitter C, Smallwood M, Stewart E,

Strong R, Suh E, Thomas R, Tint NN, Tse S, Vech C, Wang G, Wetter J, Williams S, Williams M, Windsor S, Winn-Deen E, Wolfe K, Zaveri J, Zaveri K, Abril JF, Guigó R, Campbell MJ, Sjolander KV, Karlak B, Kejariwal A, Mi H, Lazareva B, Hatton T, Narechania A, Diemer K, Muruganujan A, Guo N, Sato S, Bafna V, Istrail S, Lippert R, Schwartz R, Walenz B, Yooseph S, Allen D, Basu A, Baxendale J, Blick L, Caminha M, Carnes-Stine J, Caulk P, Chiang YH, Coyne M, Dahlke C, Mays A, Dombroski M, Donnelly M, Ely D, Esparham S, Fosler C, Gire H, Glanowski S, Glasser K, Glodek A, Gorokhov M, Graham K, Gropman B, Harris M, Heil J, Henderson S, Hoover J, Jennings D, Jordan C, Jordan J, Kasha J, Kagan L, Kraft C, Levitsky A, Lewis M, Liu X, Lopez J, Ma D, Majoros W, McDaniel J, Murphy S, Newman M, Nguyen T, Nguyen N, Nodell M, Pan S, Peck J, Peterson M, Rowe W, Sanders R, Scott J, Simpson M, Smith T, Sprague A, Stockwell T, Turner R, Venter E, Wang M, Wen M, Wu D, Wu M, Xia A, Zandieh A, Zhu X: **The sequence of the human genome.** *Science* 2001, **291:**1304-1351.

Wimmer EA, Frommer G, Purnell BA, Jäckle H: ***buttonhead* and *D-Sp1*: a novel *Drosophila* gene pair.** *Mech Dev* 1996, **59:**53-62.

Wimmer EA, Jäckle H, Pfeifle C, Cohen SM: **A *Drosophil*a homologue of human Sp1 is a head-specific segmentation gene.** *Nature* 1993, **366:**690-694.

Zhao J, Cao Y, Zhao C, Postlethwait J, Meng A: **An SP1-like transcription factor Spr2 acts downstream of Fgf signaling to mediate mesoderm induction.** *EMBO J* 2003, **22:**6078-6088.

Zhao C, Meng A: **Sp1-like transcription factors are regulators of embryonic development in vertebrates.** *Develop Growth Differ* 2005, **47:**201–211.

Zebrafish Sequencing Project [http://www.sanger.ac.uk/Projects/D_rerio/]

Figures

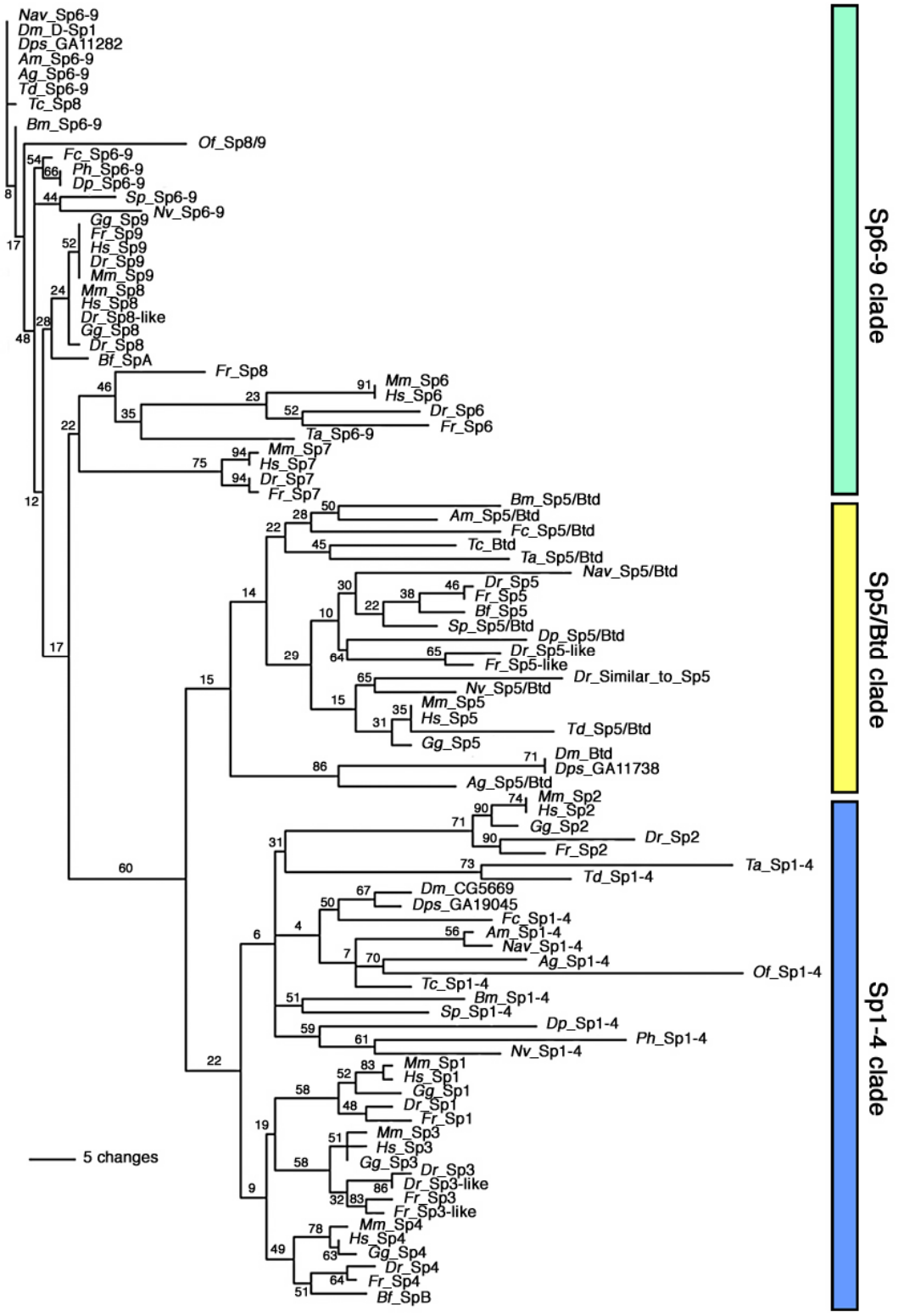


Figure 1. Phylogenetic sequence analysis of Sp factors from diverse metazoan species. The analysis reveals three large monophyletic clades (see text for details). Shown is the unrooted majority rule consensus computed from 1000 intermediate trees produced with the Quartet Puzzling method [Strimmer and von Haeseler, 1996]. The reliability values are given at the tree egdes. The abbreviations for the species names are given in the list of abbreviations.

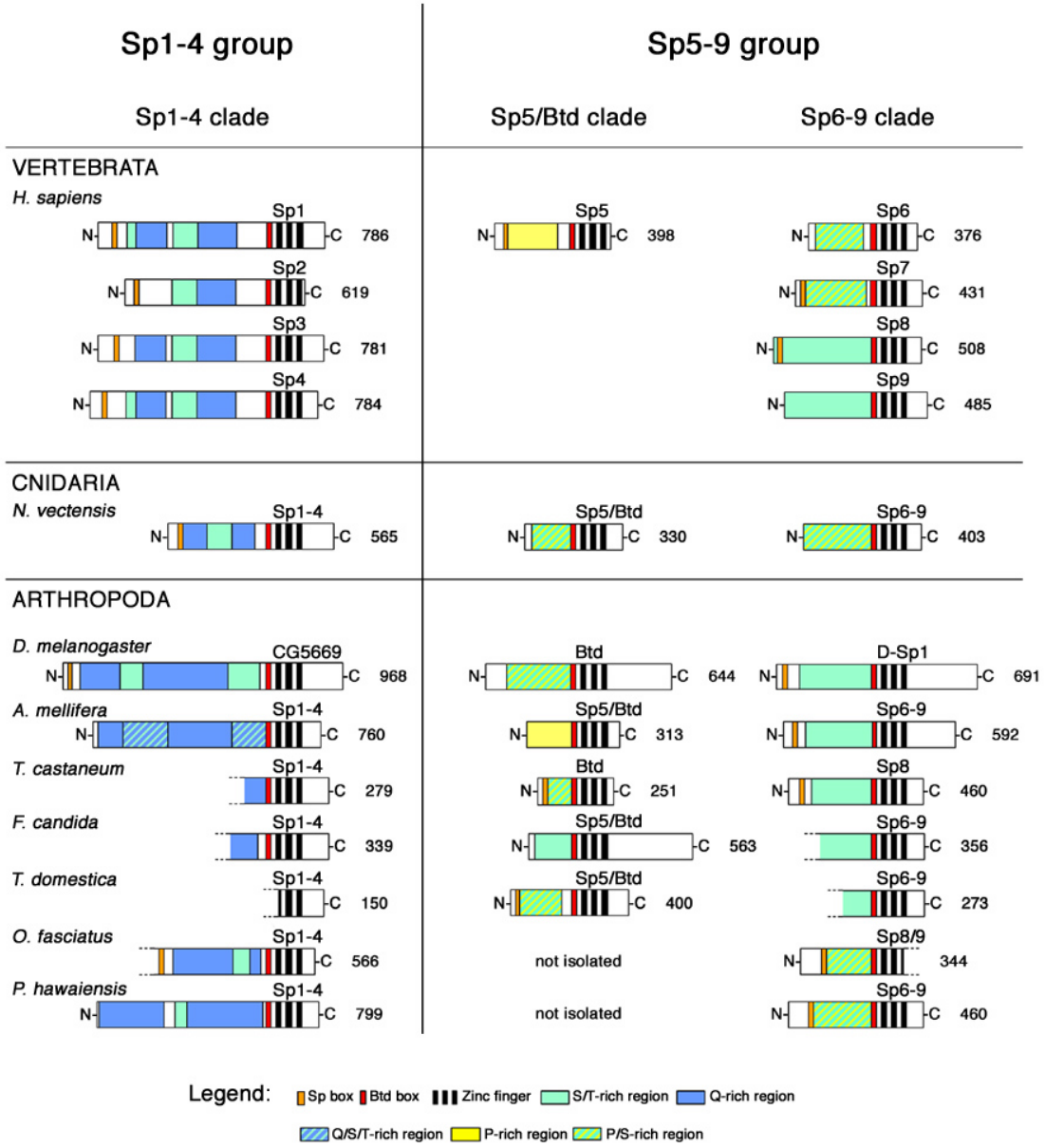


Figure 2. Protein domain structure of selected Sp-family proteins. The proteins are arranged into columns according to the clades obtained in the phylogenetic sequence analysis (Sp1-4 clade, Sp5/Btd clade, Sp6-9 clade). All proteins are oriented with their amino-terminus (N) to the left, and the carboxy-terminus (C) to the right. The length of each protein is given next to the C-terminus (number of amino acids), and the name of the protein is given above the protein. Incomplete termini of proteins are indicated by dahed lines. Structural domains are indicated by different colors and shadings explained in the legend below the proteins. Protein domain color coding after Bouwman and Philipsen [2002].

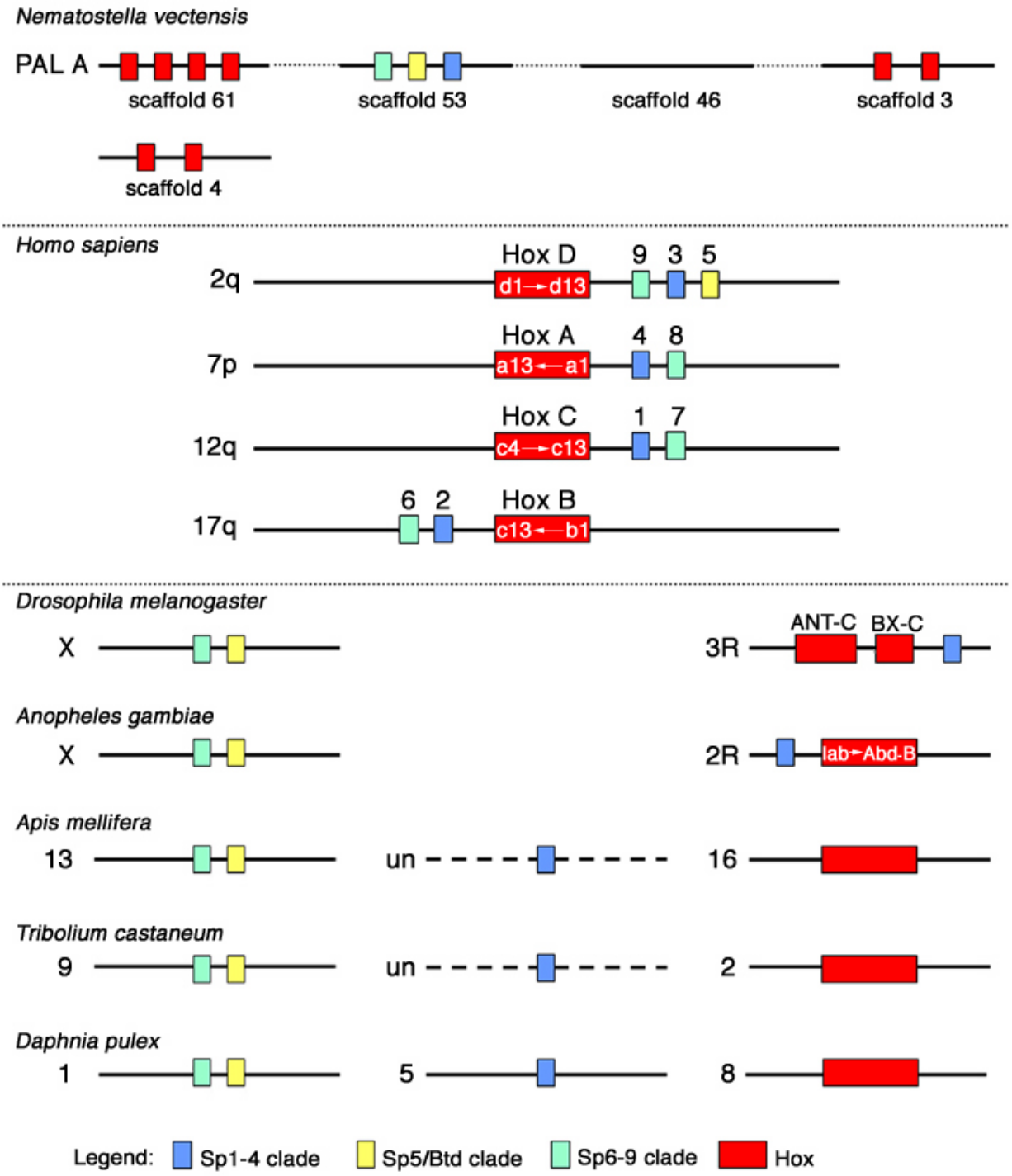

Nematostella vectensis
PAL A
scaffold 61
scaffold 53
scaffold 46
scaffold 3
scaffold 4
Homo sapiens
2q
Hox D
9 3 5
d1→d13
7p
Hox A
4 8
a13←a1
12q
Hox C
1 7
c4→c13
17q
6 2
Hox B
c13←b1
Drosophila melanogaster
X
3R
ANT-C
BX-C
Anopheles gambiae
X
2R
lab→Abd-B
Apis mellifera
13
un
16
Tribolium castaneum
9
un
2
Daphnia pulex
1
5
8
Legend:
Sp1-4 clade
Sp5/Btd clade
Sp6-9 clade
Hox

Figure 3

Figure 3. Chromosomal location of the Sp-family genes in the genomes of selected animal species. Single genes are represented by small colored boxes, Hox gene clusters are denoted by a larger red box. The Sp-family genes are color-coded according to the three clades revealed by the phylogenetic sequence analysis; the colors are explained in the legend below the drawings. The Hox cluster in *D. melanogaster* is split into two parts, the Antennapedia complex (ANT-C) and the Bithorax complex (BX-C). Continuous genomic regions are indicated by solid black lines, dashed lines indicate unclear conditions due to incomplete genome assembly. Genes, gene clusters and genomic loci are not drawn to scale. The numbers above the Sp genes in *H. sapiens* denote the gene name (*Sp1* to *Sp9*). The denominations left of the black lines indicate the chromosome (X is the X-chromosome), or the linkage group (for *A. mellifera*, *T. castaneum*); un denotes localisation of the gene in an unassembled region of the genome in species where the genome assembly is incomplete. For *N. vectensis* the scaffolds containing Hox and Sp genes are shown and are arranged into the hypothetical ancestral linkage group PAL A [Putnam et al., 2007]. The hypothetical linkage between these scaffolds is indicated by the dotted lines.

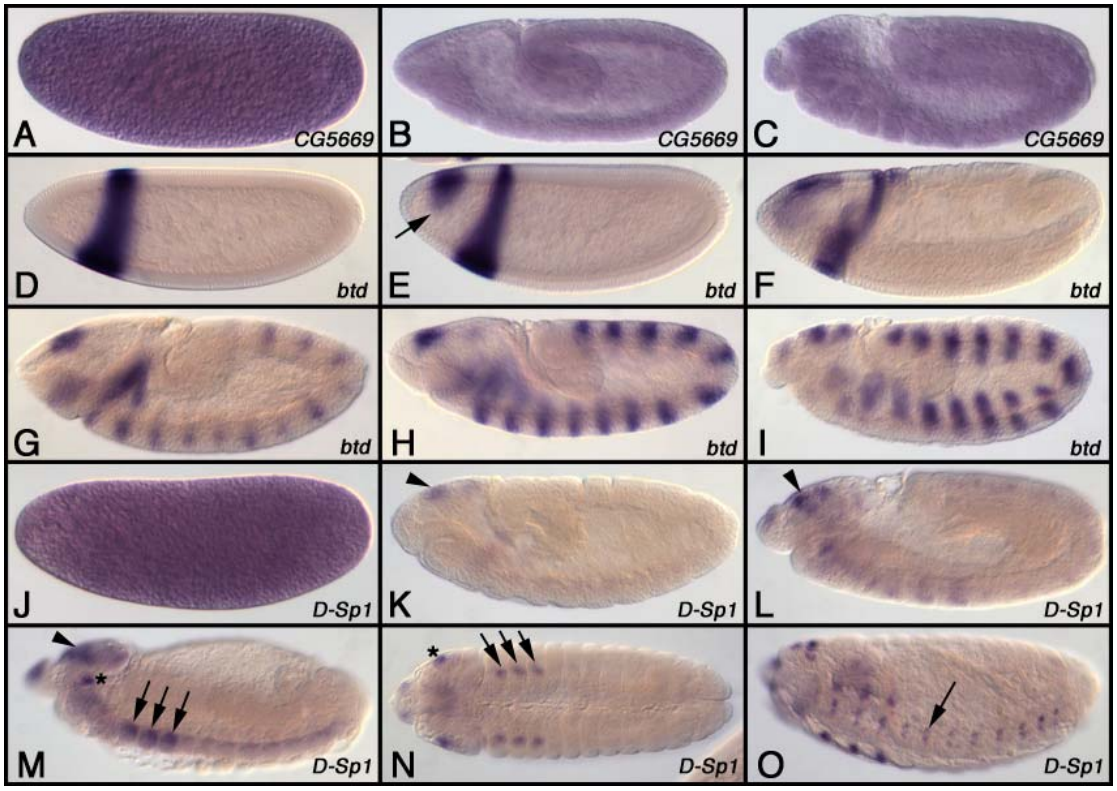


Figure 4. Embryonic expression patterns of Sp genes in *Drosophila melanogaster*. (A-C) Expression of the *Sp1-4* representative *CG5669*. (A) Stage 2 embryo. (B) Stage 9 embryo. (C) Stage 11 embryo. (D-I) Expression of the *Sp5/btd* representative *btd*. (D) Stage 4 embryo. (E) Stage 5 embryo. The arrow points to the anterior head domain that appears after the head stripe. (F) Stage 6 embryo. (G) Stage 8 embryo. (H) Stage 9 embryo. (I) Stage 11 embryo. (J-O) Expression of the *Sp6-9* representative *D-Sp1*. (J) Stage 2 embryo. (K) Stage 10 embryo. (L) Stage 11 embryo. (M) Stage 13 embryo. (N) Stage 13 embryo, ventral view. (O) Stage 15 embryo. The arrowhead in K, L, M points to expression in the developing brain. The asterisk in M, N indicates expression in the antennal primordium. The arrows in M, N point to the thoracic limb primordia. The arrow in O points to expression in the developing ventral nerve cord. All embryos are arranged with anterior to the left.

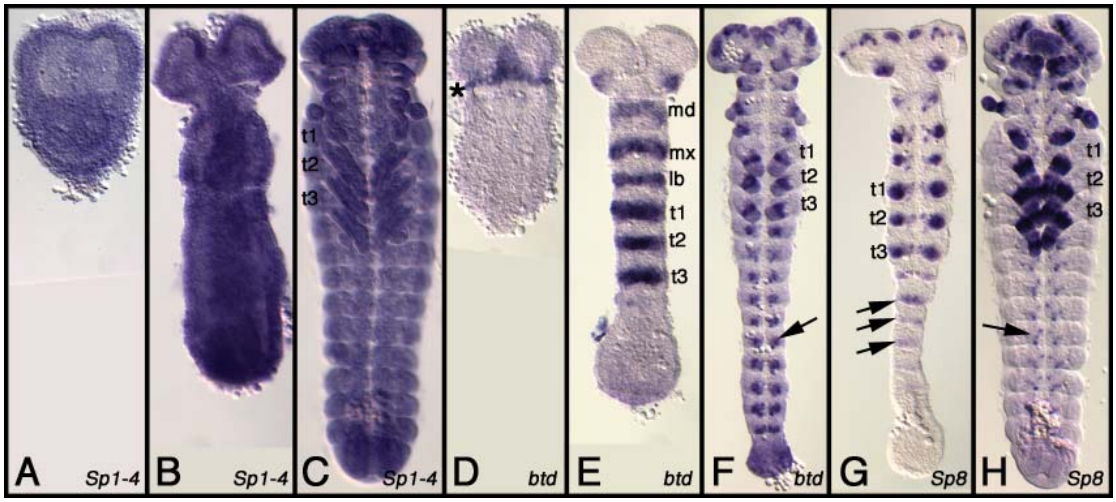


Figure 5. Embryonic expression patterns of Sp genes in *Tribolium castaneum*. (A-C) Expression of the *Sp1-4* representative at the gastrulating germband stage (A), after serosal closure (B) and at mid germband retraction (C). (D-F) Expression of the *Sp5/btd* representative *Tc-btd* shortly before serosal closure (D), at beginning germ band elongation (E), and beginning germ band retraction (F). The asterisk in D denotes the early head stripe expression domain. The arrow in F points to expression in the developing ventral nerve cord. (G, H) Expression of the *Sp6-9* representative *Tc-Sp8* at mid germband elongation (G) and at mid germband retraction (H). The arrows in G and H point to expression in the developing ventral nerve cord. All embryos are oriented with anterior to the top. Abbreviations: md, mandibulary segment; mx, maxillary segment; lb, labial segment; t1 - t3, thoracic segments 1 to 3.

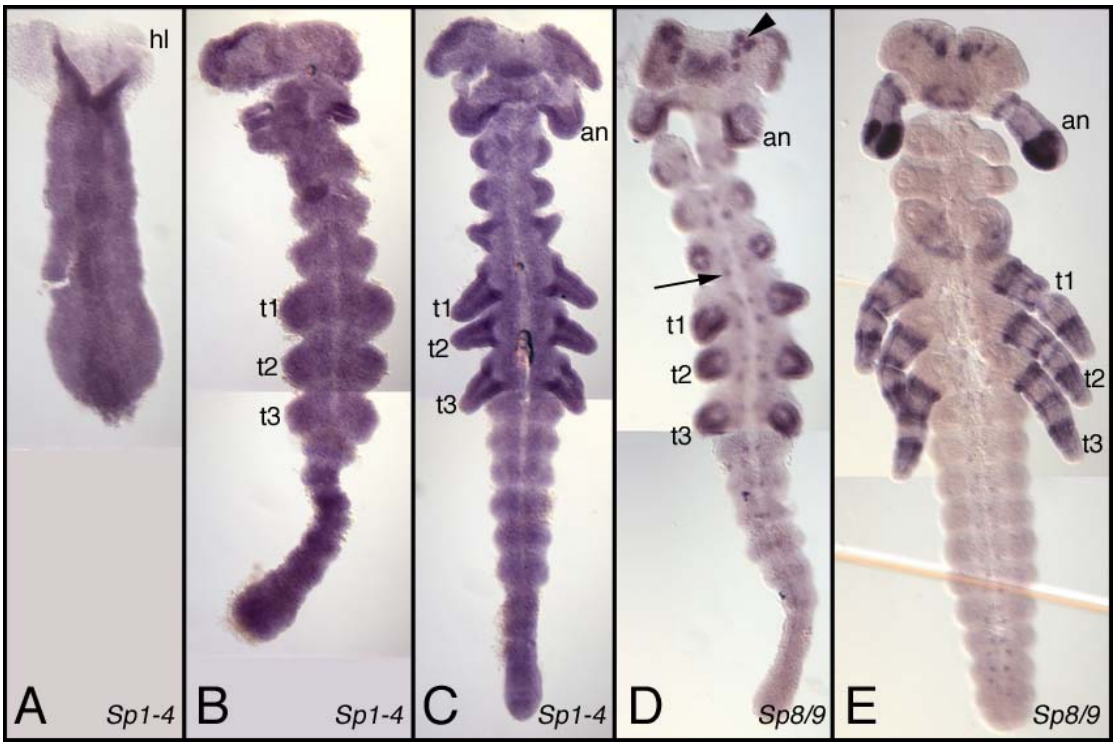


Figure 6. Embryonic expression patterns of Sp genes in *Oncopeltus fasciatus*. (A-C) Expression of the *Sp1-4* representative after serosal closure (A), at early germband elongation (B), and at full germband elongation (C). (D, E) Expression of the *Sp6-9* representative *Of-Sp8/9* at full germband elongation (D) and mid germband retraction (E). The arrowhead and arrow in D denote expression in the brain and ventral nervous system, respectively. All embryos are oriented with anterior to the top. Abbreviations see Fig. 5. Additional abbreviations: hl, head lobe; an, antennal segment/appendage.

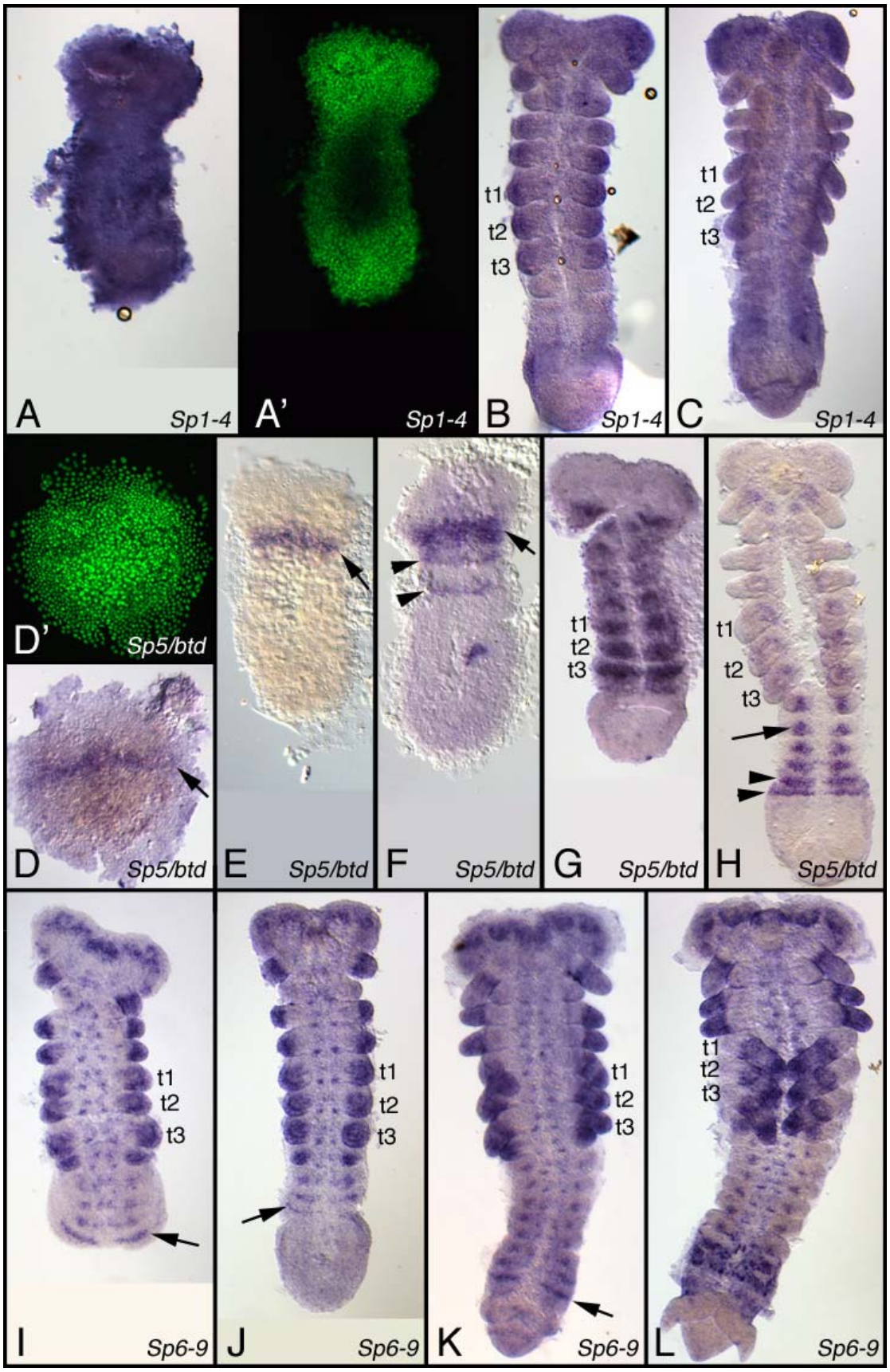

A
Sp1-4
A'
Sp1-4
B
Sp1-4
t1
t2
t3
C
Sp1-4
t1
t2
t3
D'
Sp5/btd
D
Sp5/btd
E
Sp5/btd
F
Sp5/btd
G
Sp5/btd
t1
t2
t3
H
Sp5/btd
t1
t2
t3
I
Sp6-9
t1
t2
t3
J
Sp6-9
t1
t2
t3
K
Sp6-9
t1
t2
t3
L
Sp6-9
t1
t2
t3

Figure 7

Figure 7. Embryonic expression patterns of Sp genes in *Thermobia domestica*. (A-C) Expression of the *Sp1-4* representative at the early germ band stage (A), mid germ band elongation (B), and late germ band elongation (C). (A') is the epifluorescence image (Sytox Green staining) of the embryo in A. (D-H) Expression of the *Sp5/btd* representative at the blastoderm stage (D), early germ band stage (E), starting germ band elongation (F), early germ band elongation (G), and late germ band elongation (H). The arrow in D-F points to the early head stripe. The arrowheads in F, H point to metameric stripes. The arrow in H points to expression in ventral nervous system. (D') is the epifluorescence image (Sytox Green staining) of the embryo in D. (I-L) Expression of the *Sp6-8* representative at mid germ band elongation (I, J; the embryo in J is slightly more advanced than the one in I), early germ band retraction (K), and late germ band retraction (L). The arrow in I-K points to segmental stripes in young segments that just have formed from the posterior growth zone. All embryos are oriented with anterior to the top. Abbreviations see Fig. 5.

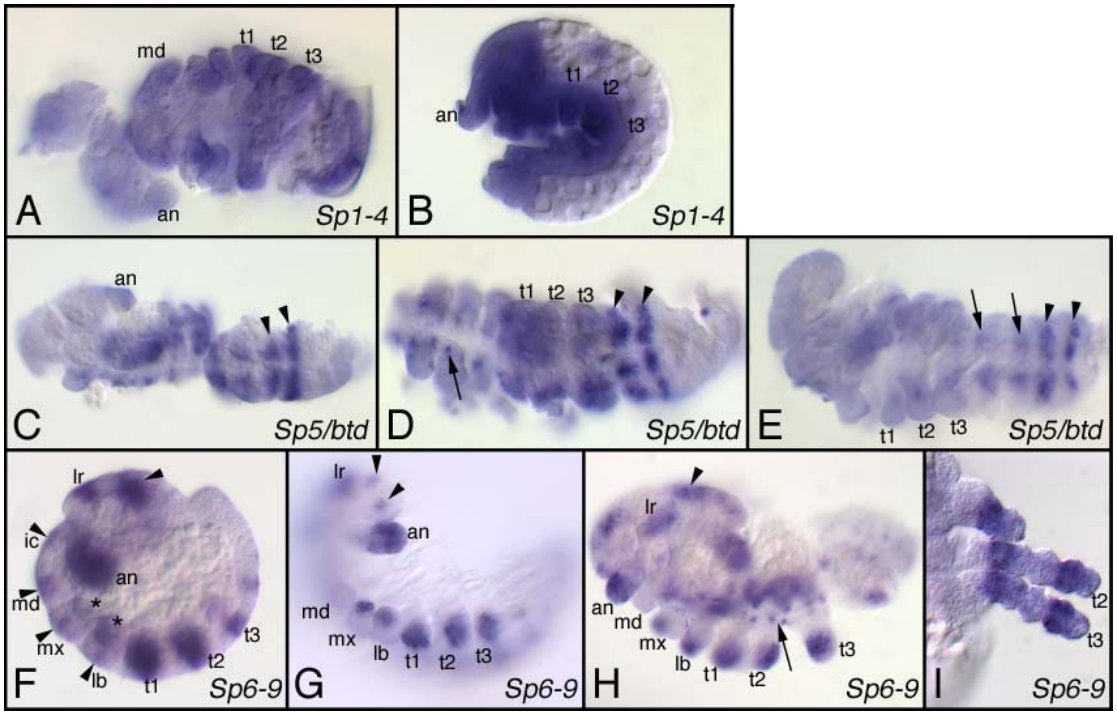


Figure 8. Embryonic expression patterns of Sp genes in *Folsomia candida*. (A, B) Expression of the *Sp1-4* representative at early germ band retraction (A) and beginning dorsal extension (B). (C-E) Expression of the *Sp5/btd* representative at mid germ band extension (C), late germ band extension (D, note that the anterior head has been lost during preparation), and early germ band retraction (E). The arrowheads in C-D point to metameric stripes. The arrows in D, E point to expression in the developing ventral nervous system. (F-I) Expression of the *Sp6-9* representative at early germ band extension (F), mid germ band extension (G), and late germ band retraction (H). The asterisks in F denote expression in the limb buds of the maxillary and labial segment. The arrowheads in F-H point to expression spots in the brain and neural tissue of the head segments. The arrow in H points to punctate expression in the developing ventral nerve cord. (I) shows dissected legs of an embryo after dorsal closure. Note that the first thoracic leg has been damaged during preparation. Anterior is to the left in panels A-H. All panels show ventral aspects except for B, F, G which are lateral views. In I distal is to the right. Abbreviations see Fig. 5. Additional abbreviations: an, antennal segment/appendage; lr, labrum; ic, intercalary segment.

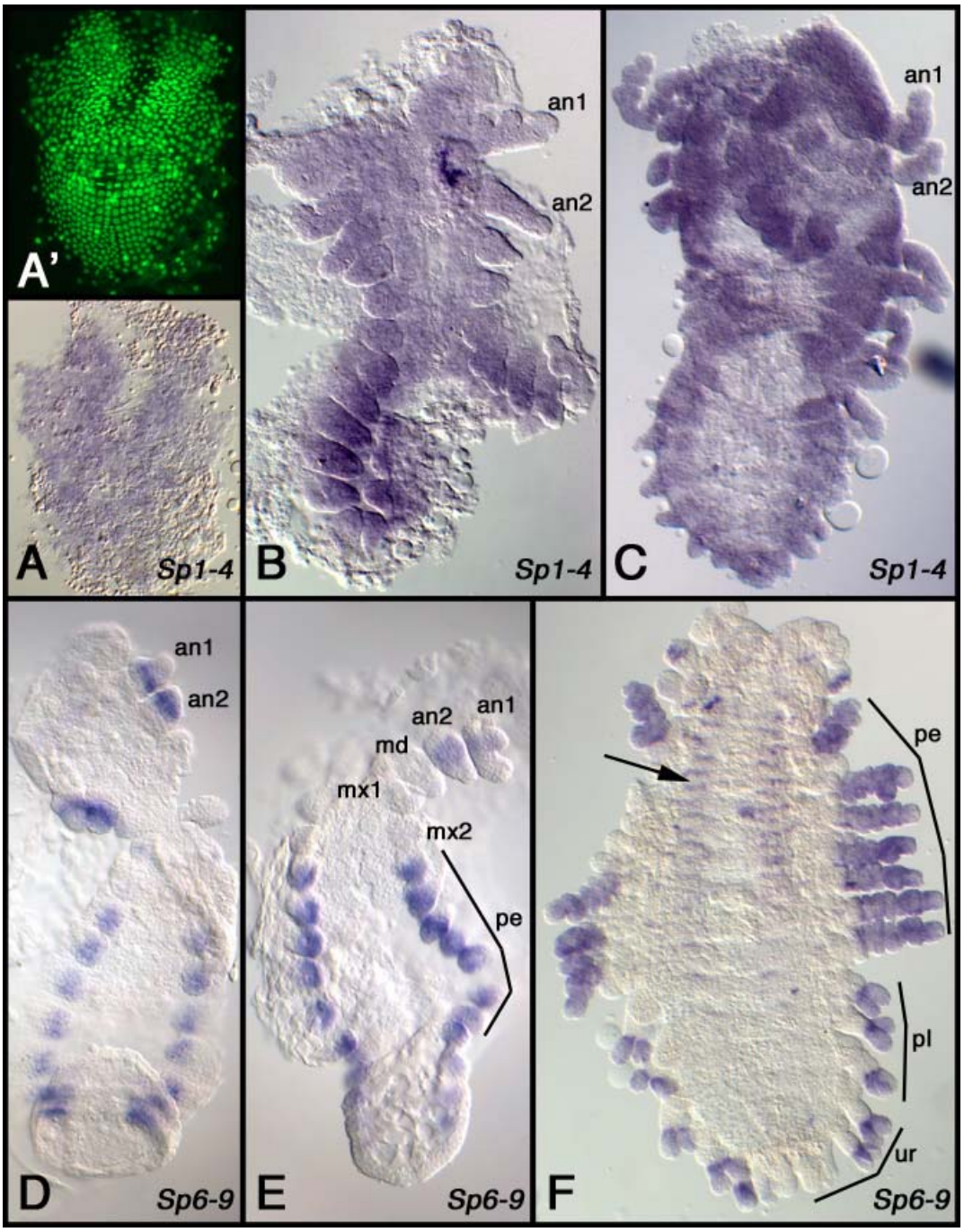


Figure 9. Embryonic expression patterns of Sp genes in *Parhyale hawaiensis*. (A-C) Expression of the *Sp1-4* representative at stage S12 (A), stage S19 (B), and stage S22 (C). Staging according to Browne et al. [2005]. (A') is the epifluorescence image (Sytox Green staining) of the embryo in A. (D-F) Expression of the *Sp6-9* representative at stage S17 (D), stage S18 (E), and stage S23 (F). The arrow in F points to expression in the ventral nervous system. Abbreviations: an1, first antenna; an2, second antenna; md, mandible; mx1, first maxilla; mx2, second maxilla; pe, peraeopods; pl, pleopods; ur, uropods.

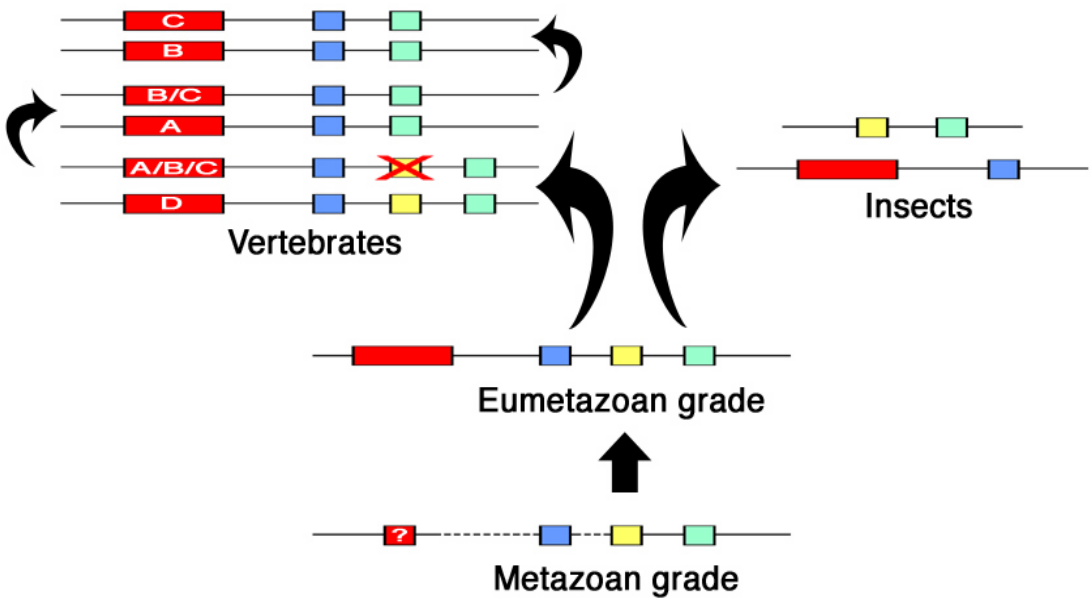


Figure 10. Evolution and orthology of the Sp genes in the Metazoa. The ancestral state in the Metazoa (metazoan grade) is hypothesized to have been a cluster of three Sp genes (one of each clade, color coded as in Fig. 3) linked to a single Hox gene (small red box; no Hox cluster is present yet). This is consistent with the data from *T. adhaerens*. Linkage data for the *Sp1-4* gene, however, are missing and the nature of *Trox2* gene is debated (indicated by the dashed lines and the question mark). In addition, data from the most basal metazoan group, the Porifera, are not yet available. The Sp gene cluster is conserved in the Eumetazoa (eumetazoan grade) and is linked to the Hox cluster. This is consistent with data from *N. vectensis*, and is further supported by comparative genomics [Putnam et al., 2007]. Further evolution in the Vertebrata lineage lead to the multiplication of the Sp gene cluster along with the Hox gene cluster. The number of Sp gene clusters, their chromosomal location and Sp gene complement is fully compatible with the sequence (D(A(C, B))) of vertebrate Hox gene cluster duplication proposed by Bailey et al. [1997]. We propose that the ancestral Sp gene cluster was duplicated, the duplicate lost the *Sp5/btd* gene, and this reduced cluster served as template for two additional Hox/Sp duplications. In the Insecta lineage the ancestral linkage of the Sp cluster with the Hox cluster was partially disbanded by the relocation of the *Sp5/btd* and *Sp6-9* genes. Please note that the sequence of Sp genes along the chromosome and their location to the right of the Hox cluster as shown in the figure is a simplification for reasons of brevity. The actual sequence of Sp genes and their relative location to the Hox gene cluster may differ in different species.

Tables and captions

Table 1. Genomic locations of Sp genes and Hox genes. This table supplements the schematic overview given in Fig. 3. The first column gives the chromosome (or linkage group/scaffold) of a given species. The second column gives the Sp genes and Hox genes present on this chromosome (linkage group/scaffold); only representative Hox genes are given for reasons of clarity. The third column gives the exact location of the genes on the chromosome (base pair numbers are given for linkage groups, scaffolds, and some chromosomes where no mapping data was available). The base pair values are based on the following genome assembly versions: *A. gambiae*: AgamP3.3, *A. mellifera*: Amel_4.0, *T. castaneum*: Tcas_3.0, *D. pulex*: JGI-2006-09, *N. vectensis*: Nematostella vectensis v1.0. The data for the *N. vectensis* Hox genes can be found in the references given in the table. Alternating shading for different species is used in the table to enhance the legibility of the table. Abbreviations: LG, linkage group; un, unassembled portions of the genome.

species/location	Sp genes/Hox genes	chromosomal location
Human	*Sp3*	2q31
Chromosome 2	*Sp5*	2q31.1
	Sp9	2q31.1
	Hox D1	2q31.1
	Hox D13	2q31.1
Human	*Sp4*	7p15.2
Chromosome 7	*Sp8*	7p15.2
	Hox A1	7p15.3
	Hox A13	7p15-14
Human	*Sp1*	12q13.1
Chromosome 12	*Sp7*	12q13.13
	Hox C4	12q13.3
	Hox C13	12q13.3
Human	*Sp2*	17q21.32
Chromosome 17	*Sp6*	17q21.32
	Hox B1	17q21.3
	Hox B13	17q21.2

D. melanogaster	*D-Sp1* (*Sp6-9*)	9A1
X-chromosome	*btd*	8F10
D. melanogaster	*CG5669* (*Sp1-4*)	95F2-95F3
Chromosome 3R	Antennapedia complex	84A5-84B2
	Bithorax complex	89D6-89E5
A. gambiae	*Sp6-9*	13.344-13.349kbp
X-chromosome	*Sp5/btd*	13.439-13.440kbp
A. gambiae	*Sp1-4*	57.814-57.831kbp
Chromosome 2R	Hox *(Scr)*	59.732-59.775kbp
	Hox *(Ubx)*	60.110k-60.173kbp
A. mellifera	*Sp6-9*	1.371-1.393kbp
LG 13	*Sp5/btd*	1.441-1.442kbp
A. mellifera	*Sp1-4*	8352-11768bp
LG un		
A. mellifera	Hox (*lab*)	3.892-3.909kbp
LG 16		
T. castaneum	*Sp8*	19.911-19.927kbp
LG 9	*btd*	19.960-19.961kbp
T. castaneum	*Sp1-4*	<618-6806bp
LG un		
T. castaneum	Hox (*labial*)	10.378-10.410kbp
LG 2	Hox (*Abd-B*)	9.691-9.700kbp
D. pulex	*Sp6-9*	102.625-163.646bp
Chromosome 1	*Sp5/btd*	263.041-281.587bp
D. pulex	*Sp1-4*	781.813-986.944bp
Chromosome 5		
D. pulex	Hox	~544kbp
Chromosome 8		
N. vectensis	*Sp1-4*	330.760-331.110bp
Scaffold 53	*Sp5/btd*	305.778-306.053bp
	Sp6-9	275.750-276.850bp
N. vectensis	Hox	[Putnam et al., 2007]
Scaffolds 3, 61, 4		[Ryan et al., 2007]

Supplementary Information

```
                1                                                          60
Dm_D_Sp1        GRATCDCPNCQEAERLGP------------------AGVHLRKKNIHSCHIPGCGKVYG
Dps_GA11282     GRATCDCPNCQEAERLGP------------------AGVHLRKKNIHSCHIPGCGKVYG
Am_Sp6-9        GRATCDCPNCQEAERLGP------------------AGVHLRKKNIHSCHIPGCGKVYG
Nav_Sp6-9       GRATCDCPNCQEAERLGP------------------AGVHLRKKNIHSCHIPGCGKVYG
Ag_Sp6-9        GRATCDCPNCQEAERLGP------------------AGVHLRKKNIHSCHIPGCGKVYG
Td_Sp6-9        GRATCDCPNCQEAERLGP------------------AGVHLRKKNIHSCHIPGCGKVYG
Tc_Sp8          GRATCDCPNCQEAERLGP------------------AGVHLRKKNIHSCHIPGCGKVYG
Of_Sp8/9        GRATCDCPNCQEAERLGP------------------AGVHLRKKNIHSCHIPGCGKVYG
Bm_Sp6-9        GRATCDCPNCQEAERLGP------------------AGAHLRKKNIHSCHIPGCGKVYG
Fc_Sp6-9        GRATCDCPNCQEAERLGP------------------AGAHLRKKNIHSCHIPGCGKVYG
Ph_SP6-9        GRATCDCPNCQEADRLGP------------------AGAHLRKRNIHSCHIPGCGKVYG
Dp_Sp6-9        GRATCDCPNCQEADRLGP------------------AGAHLRKRNIHSCHIPGCGKVYG
Gg_Sp9          GRATCDCPNCQEAERLGP------------------AGASLRRKGLHSCHIPGCGKVYG
Fr_Sp9          GRATCDCPNCQEAERLGP------------------AGASLRRKGLHSCHIPGCGKVYG
Dr_Sp9          GRATCDCPNCQEAERLGP------------------AGASLRRKGLHSCHIPGCGKVYG
Hs_Sp9          GRATCDCPNCQEAERLGP------------------AGASLRRKGLHSCHIPGCGKVYG
Mm_Sp9          GRATCDCPNCQEAERLGP------------------AGASLRRKGLHSCHIPGCGKVYG
Mm_Sp8          GRATCDCPNCQEAERLGP------------------AGASLRRKGLHSCHIPGCGKVYG
Hs_Sp8          GRATCDCPNCQEAERLGP------------------AGASLRRKGLHSCHIPGCGKVYG
Dr_Sp8-like     GRATCDCPNCQEAERLGP------------------AGASLRRKGLHSCHIPGCGKVYG
Gg_Sp8          GRATCDCPNCQEAERLGP------------------AGASLRRKGLHSCHIPGCGKVYG
Dr_Sp8          GRATCDCPNCQEAERLGP------------------AGASLRRKGLHSCHIPGCGKVYG
Fr_Sp8          GRATCDCPNCQEAESLGP------------------GGASPQRRGLHSCHIPGCGKVYG
Bf_SpA          GRATCDCPNCQEAERLGP------------------AGAAIRRKNQHSCHIPGCGKVYG
Sp_Sp6-9        GRATCDCPNCQENERLAA------------------AGQPLRKKNVHSCHIPGCGKIYG
Mm_Sp7          GRSTCDCPNCQELERLGA------------------AAAGLRKKPIHSCHIPGCGKVYG
Hs_Sp7          GRSSCDCPNCQELERLGA------------------AAAGLRKKPIHSCHIPGCGKVYG
Dr_Sp7          GRSSCDCPNCQELERLGA------------------SAASLRKKPVHSCHIPGCGKVYG
Fr_Sp7          TRSSCDCPNCQELERLGA------------------SAASLRKKPVHSCHIPGCGKVYG
Nv_Sp6-9        GRATCDCPNCQDNERMSA------------------SGAPFRKKSQHICHIPGCGKVYG
Mm_Sp6          GQTVCRCPNCLEAERLGAP-----------------CGPDGGKKKHLHNCHIPGCGKAYA
Hs_Sp6          GQTVCRCPNCLEAERLGAP-----------------CGPDGGKKKHLHNCHIPGCGKAYA
Dr_Sp6          GQASCRCPNCLSAESLGN------------------SGDASKRKHLHNCHIPGCGKAYV
Fr_Sp6          AQAVCRCPNCIHAEQLGQ------------------STDDTRRKHMHNCHIPGCGKAYA
Ta_Sp6-9        NRSSCNCPYCQENERLAAT-----------------GQPVPRRKAQHICHIADCGKVYT
Bm_Sp5/Btd      RCAKCRCPNCLTEAAGFGP-----------------NYGKDGAKREHVCHVPGCGKVYG
Tc_Btd          RCIKCQCPNCVNEEVGLK-----------------------KPSSKKVHVCHYQGCDKVYG
Fc_Sp5/Btd      KSRRCRCPNCLAGIQPTQV-----------------P-GEKKKRQHICHMIGCGKIYG
Am_Sp5/Btd      KCIRCRCPNCQTENGGSQL-----------------G---LDGKRQHVCHVPGCGKVYG
Nav_Sp5/Btd     KCTRCQCPNCMADGG---------------------NVGRDGKRQHSCHIDGCNKVYG
Dr_Sim-to-Sp5   QSRRCMCPNCQKNAD----------------------TPGR--RKQHACHIPGCAKVYG
Dr_Sp5          RCRRCRCPNCQSSS----------------------SSDEPGKKKQHICHIPGCGKVYG
Fr_Sp5          RCRRCRCPNCQSST----------------------SSDEPGKKKQHICHIPGCGKVYG
Bf_Sp5          RCRRCRCPNCQNST----------------------AGS-PNKKKQHICHIPGCGKVYG
Sp_Sp5/Btd      RCRRCRCPNCLNPS----------------------ISSEPAKKRQHICHIPGCGKVYG
Dp_Sp5/Btd      RCRRCRCPNCQDPTTGASIT---------------GHSNSQHKKKQHLCHVPGCGKVYG
Mm_Sp5          RCRRCRCPNCQAAGG---------------------APEAEPGKKKQHVCHVPGCGKVYG
Hs_Sp5          RCRRCRCPNCQAAGG---------------------APEAEPGKKKQHVCHVPGCGKVYG
Gg_Sp5          RCRRCRCPNCQAAAGS--------------------APEAEPGKKKQHVCHIPGCGKVYG
Nv_Sp5/Btd      RCRRCRCPNCQLAAT----------------------TGNTSKRKQHVCHIPGCGKVYG
Dr_Sp5-like     RCRRCKCPNCQSTGNGG-------------------AALEFGKKRLHICHIPDCGKVYK
Fr_Sp5-like     RCRRCKCPNCQANGGG--------------------LEFGKKRLHVCHIPDCGKVYK
Td_Sp5/Btd      RCSRCRCPNCQNGGN---------------------TAGQQPQRKRQHVCHVPGCSKTYG
Mm_Sp2          RRMACTCPNCKDGEK----------------------RSGEQGKKKHVCHIPDCGKTFR
Hs_Sp2          RRMACTCPNCKDGEK----------------------RSGEQGKKKHVCHIPDCGKTFR
Gg_Sp2          RRMACTCPNCKDGEK----------------------RPGDQGKKKHICHIPECGRTFR
Dr_Sp2          RHISCLCFLCY--------------------------RPGEVGKRKHICHIAGCEKTFR
Fr_Sp2          RRMACTCPNCKDADK----------------------RPGEVGKRKHICHVPGCEKTFR
Ta_Sp1-4        ------------------------------------------NHLSHL---------
Dm_CG5669       KRVACTCPNCTDGEKHS-------------------------DKKRQHICHITGCHKVYG
Dps_GA19045     KRVACTCPNCTDGEKHS-------------------------DKKRQHICHIPGCQKVYG
Fc_Sp1-4        RRVACTCPNCTDIDASSNNSRN--------------------NGEKKKQHLCHIPGCGKVYG
Am_Sp1-4        RRVACTCPNCGDGDRNRD------------------------MTRKRQHVCHIAGCNKVYG
Nav_Sp1-4       RRVACTCPNCGDGDRNRD------------------------MTRKRQHVCHIPGCSKVYG
Tc_Sp1-4        RRVACTCPNCQEGERHAD------------------------RKKQHICHIPGCNKVYG
Ag_Sp1-4        RRVACTCPNCEVKSSGPP------------------------DRKRQHICHVSGCNKVYG
Of_Sp1-4        RRVACSCPNCIEGHAGCSRLRAEEATTSLEHVRTINKATWGGIRKKQHVCHVAGCNKVYG
Bm_Sp1-4        KRVACTCPNCDQGENVFSR-----------------------VVDRKKQHVCHIAGCNKVYG
Td_Sp1-4        ------------------------------------------------------NKVYG
Mm_Sp1          RREACTCPYCKDSEGRASG-----------------------DPGKKKQHICHIQGCGKVYG
Hs_Sp1          RREACTCPYCKDSEGRGSG-----------------------DPGKKKQHICHIQGCGKVYG
Gg_Sp1          RREACTCPYCKDSEGRSSG-----------------------DPGKKKQHICHIPGCGKVYG
Dr_Sp1          RREACTCPYCKDGEGR--------------------------DPSKKKQHICHIPGCGKVYG
Fr_Sp1          RREACTCPFCKDGEGR--------------------------DPTKKKQHICHISGCGKIYG
Dp_Sp1-4        RRVACTCPNCKDGEVARKSNRPK-----------IFPSRVGENKRKIHICHHQGCGKVYG
Mm_Sp3          RRVACTCPNCKEGGGRGTN-----------------------LGKKKQHICHIPGCGKVYG
Hs_Sp3          RRVACTCPNCKEGGGRGTN-----------------------LGKKKQHICHIPGCGKVYG
Gg_Sp3          RRVACTCPNCKEGGGRGSN-----------------------LGKKKQHICHIPGCGKVYG
Dr_Sp3          RRVACTCPNC-EAGGRGSS-----------------------MGKKKQHICHIPGCGKVYG
Dr_Sp3-like     RRVACTCPNCKEAGGRGSS-----------------------MGKKKQHICHIPGCGKVYG
Fr_Sp3          RRVACTCPNCKESGGRGSS-----------------------TGKKKQHICHIAGCGKVYG
Fr_Sp3-like     RRVACTCPNCKESGGRGSG-----------------------MGKKKQHICHIAGCGKVYG
Sp_Sp4          RRLACTCPNCKDGDGRG-------------------------KKQHICHIADCGKIYG
Mm_Sp4          RRVACSCPNCREGEGRGSS-----------------------EPGKKKQHVCHIEGCGKVYG
Hs_Sp4          RRVACSCPNCREGEGRGSN-----------------------EPGKKKQHICHIEGCGKVYG
Gg_Sp4          RRVACSCPNCREGEGRSSN-----------------------EPGKKKQHICHIEGCGKVYG
```

```
Dr_Sp4            RRVACSCPNCRDGEGRNNS--------------------DPSKKKQHVCHMEGCGKVYG
Fr_Sp4            RRVACSCPNCRDGEGRNSG--------------------DPTKKKQHICHMEGCGKVYG
Bf_SpB            RRVACSCPNCREGEGRGNG--------------------D-SKKKQHICHIAGCGKVYG
Ph_Sp1-4          KRVACTCPNCREGGVNNERGEPSSSNGI-----------GGSTKRRQHICHIPGCNKTYG
Nv_Sp1-4          -RIACTCPNCRDG-------EGRTANG-----------------RKQHVCHVPGCGKVYG
Ta_Sp1-4          RRVACCCPNCRNPDYKPPQ----------------------GGKKMHVCHYQGCSKVYG
Dm_Btd            RSVRCTCPNCTNEMSGLPP-----------------IVGPDERGRKQHICHIPGCERLYG
Dps_GA11738       RSVRCTCPNCTNEMSGLPP-----------------IVGPDERGRKQHICHIPGCERLYG
Ag_Sp5/Btd        RCARCTCPNCINELSGLPP-----------------VVGPDEKGKRQHICHIPGCEKIYG


                 61                                                        120
Dm_D_Sp1          KTSHLKAHLRWHTGERPFVCNWLFCGKRFTRSDELQRHLRTHTGEKRFACPVCNKRFMRS
Dps_GA11282       KTSHLKAHLRWHTGERPFVCNWLFCGKRFTRSDELQRHLRTHTGEKRFACPVCNKRFMRS
Am_Sp6-9          KTSHLKAHLRWHTGERPFVCNWLFCGKRFTRSDELQRHLRTHTGEKRFACPVCNKRFMRS
Nav_Sp6-9         KTSHLKAHLRWHTGERPFVCNWLFCGKRFTRSDELQRHLRTHTGEKRFACPVCNKRFMRS
Ag_Sp6-9          KTSHLKAHLRWHTGERPFVCNWLFCGKRFTRSDELQRHLRTHTGEKRFACPVCNKRFMRS
Td_Sp6-9          KTSHLKAHLRWHTGERPFVCNWLFCGKRFTRSDELQRHLRTHTGEKRFACPVCNKRFMRS
Tc_Sp8            KTSHLKAHLRWHTGERPFVCNWLFCGKRFTRSDELQRHLRTHTGEKRFACPICNKRFMRS
Of_Sp8/9          KTSHLKAHLRWHTGERPFVCNWLFCGKRFTRSDELQRHLRTHTGEKRFACPVC-------
Bm_Sp6-9          KTSHLKAHLRWHTGERPFVCNWLFCGKRFTRSDELQRHLRTHTGEKRFACPVCNKRFMRS
Fc_Sp6-9          KTSHLKAHLRWHTGERPFVCNWLFCGKRFTRSDELQRHLRTHTGEKRFACPICNKRFMRS
Ph_Sp6-9          KTSHLKAHLRWHTGERPFVCNWLFCGKRFTRSDELQRHLRTHTGEKRFACPVCNKRFMRS
Dp_Sp6-9          KTSHLKAHLRWHTGERPFVCNWLFCGKRFTRSDELQRHLRTHTGEKRFACPVCNKRFMRS
Gg_Sp9            KTSHLKAHLRWHTGERPFVCNWLFCGKRFTRSDELQRHLRTHTGEKRFACPVCNKRFMRS
Fr_Sp9            KTSHLKAHLRWHTGERPFVCNWLFCGKRFTRSDELQRHLRTHTGEKRFACPVCNKRFMRS
Dr_Sp9            KTSHLKAHLRWHTGERPFVCNWLFCGKRFTRSDELQRHLRTHTGEKRFACPVCNKRFMRS
Hs_Sp9            KTSHLKAHLRWHTGERPFVCNWLFCGKRFTRSDELQRHLRTHTGEKRFACPVCNKRFMRS
Mm_Sp9            KTSHLKAHLRWHTGERPFVCNWLFCGKRFTRSDELQRHLRTHTGEKRFACPVCNKRFMRS
Mm_Sp8            KTSHLKAHLRWHTGERPFVCNWLFCGKRFTRSDELQRHLRTHTGEKRFACPVCNKRFMRS
Hs_Sp8            KTSHLKAHLRWHTGERPFVCNWLFCGKRFTRSDELQRHLRTHTGEKRFACPVCNKRFMRS
Dr_Sp8-like       KTSHLKAHLRWHTGERPFVCNWLFCGKRFTRSDELQRHLRTHTGEKRFACPVCNKRFMRS
Gg_Sp8            KTSHLKAHLRWHTGERPFVCNWLFCGKRFTRSDELQRHLRTHTGEKRFACPVCNKRFMRS
Dr_Sp8            KTSHLKAHLRWHTGERPFVCNWLFCGKRFTRSDELQRHLRTHTGEKRFSCPVCNKRFMRS
Fr_Sp8            KTSHLKAHLRWHTGERPFVCNWLFCGKRFTRSDELQRHLRTHTGEKRFECAICQKRFMRS
Bf_SpA            KTSHLKAHLRWHTGERPFVCNWLFCGKRFTRSDELQRHLRTHTGEKRFACPVCNKRFMRS
Sp_Sp6-9          KTSHLKAHLRWHTGERPFVCNWLFCGKRFTRSDELQRHLRTHTGEKRFSCPTCNKRFMRS
Mm_Sp7            KASHLKAHLRWHTGERPFVCNWLFCGKRFTRSDELERHVRTHTREKKFTCLLCSKRFTRS
Hs_SP7            KASHLKAHLRWHTGERPFVCNWLFCGKRFTRSDELERHVRTHTREKKFTCLLCSKRFTRS
Dr_Sp7            KASHLKAHLRWHTGERPFVCNWLFCGKRFTRSDELERHVRTHTREKKFTCLLCNKRFTRS
Fr_Sp7            KASHLKAHLRWHTGERPFVCNWLFCGKRFTRSDELERHVRTHTREKKFTCLLCNKRFTRS
Nv_Sp6-9          KTSHLKAHLRWHTGERPFVCNWLFCGKRFTRSDELQRHLRTHTGEKRFACPICSKRFMRS
Mm_Sp6            KTSHLKAHLRWHSGDRPFVCNWLFCGKRFTRSDELQRHLQTHTGTKKFPCAVCSRVFMRS
Hs_Sp6            KTSHLKAHLRWHSGDRPFVCNWLFCGKRFTRSDELQRHLQTHTGTKKFPCAVCSRVFMRS
Dr_Sp6            KTSHLKAHLRWHSGDRPFVCNWLFCGKRFTRSDELQRHLQTHTGAKRFGCSSCPRVFLRA
Fr_Sp6            KTSHLKAHLRWHSGDRPFVCNWLFCGKRFTRSDELQRHLQTHTGAKKFSCALCPRVFMRN
Ta_Sp6-9          KTSHLKAHLRWHSGERPFVCNWLFCGKRFTRSDELQRHLRTHTGEKRFVCKVCSKRFMRS
Bm_Sp5/Btd        KTSHLKAHLRWHTGERPFVCNWLFCGKRFTRSDELQRHLRTHTGEKRFSCQLCTKRFMRS
Tc_Btd            KTSHLQAHLRWHTGERPFVCNWLFCGKRFTRSDELQRHLRTHTGEKRFACSVCTKRFMRS
Fc_Sp5/Btd        KTSHLKAHLRGHTGERPFVCNWLFCGKRFTRSDELQRHLRTHTGEKRFICNVCQKRFMRS
Am_Sp5/Btd        KTSHLKAHLRWHTGERPFVCNWLFCGKRFTRSDELQRHLRTHTGEKRFACPTCGKRFMRS
Nav_Sp5/Btd       KTSHLKAHLRWHSGEKPYFCNWKPCGKSFTRSDELQRHMRTHTGEKRFVCPECGKRFMRS
Dr_Sim-to-Sp5     KTSHLKAHLRWHAGERPFICSWMFCGKSFTRSDELQRHLRTHTGEKRFTCPDCSKRFMRS
Dr_Sp5            KTSHLKAHLRWHSGERPFVCNWLFCGKSFTRSDELQRHLRTHTGEKRFVCPDCCKRFMRS
Fr_Sp5            KTSHLKAHLRWHSGERPFVCNWLFCGKSFTRSDELQRHLRTHTGEKRFVCPDCCKRFMRS
Bf_Sp5            KTSHLKAHLRWHTGERPFVCNWLFCGKSFTRSDELQRHLRTHTGEKRFACPDCGKRFMRS
Sp_Sp5/Btd        KTSHLKAHLRWHTGERPFVCNWLFCGKSFTRSDELQRHLRTHTGEKRFVCPECGKRFMRS
Dp_Sp5/Btd        KTSHLKAHLRWHAGERPFSCQWLFCGKSFTRSDELQRHLRTHTGEKRFACPVCNKRFMRS
Mm_Sp5            KTSHLKAHLRWHTGERPFVCNWLFCGKSFTRSDELQRHLRTHTGEKRFACPECGKRFMRS
Hs_Sp5            KTSHLKAHLRWHTGERPFVCNWLFCGKSFTRSDELQRHLRTHTGEKRFACPECGKRFMRS
Gg_Sp5            KTSHLKAHLRWHTGERPFVCNWLFCGKSFTRSDELQRHLRTHTGEKRFVCPECGKRFMRS
Nv_Sp5/Btd        KTSHLKAHLRWHTGERPFVCNWLFCGKSFTRSDELQRHLRTHTGEKRFQCDECGKRFMRS
Dr_Sp5-like       KTSHLKAHLRWHAGERPFICNWLFCGKSFTRSDELQRHLRTHTGEKRFGCQQCGKRFMRS
Fr_Sp5-like       KTSHLKAHLRWHAGERPFICNWLFCGKSFTRSDELQRHLRTHTGEKRFGCQQCGKRFMRS
Td_Sp5/Btd        KTSHLKAHLRWHTGERPFVCNWLFCGKSFTRSDELQRHLRTHTGEKRFACPVCNKRFMRS
Mm_Sp2            KTSLLRAHVRLHTGERPFVCNWFFCGKRFTRSDELQRHARTHTGDKRFECAQCQKRFMRS
Hs_Sp2            KTSLLRAHVRLHTGERPFVCNWFFCGKRFTRSDELQRHARTHTGDKRFECAQCQKRFMRS
Gg_Sp2            KTSLLRAHVRLHTGERPFVCNWVFCGKRFTRSDELQRHARTHTGDKRFECAQCQKRFMRS
Dr_Sp2            KTSLLRAHVRLHTGERPFVCNWVFCGKRFTRSDELQRHARTHTGKKRFECNKCQKRFMRS
Fr_Sp2            KTSLLRAHVRLHTGERPFVCSWVFCGKRFTRSDELQRHARTHTGDKRFECSQCQKRFMRS
Ta_Sp1-4          --SLLKAHQQAKVGERPFPCTWSECGKRFARSDELARHYRTHTGEKRFECPVCNKRFMRS
Dm_CG5669         KTSHLRAHLRWHTGERPFVCSWAFCGKRFTRSDELQRHRRTHTGEKRFQCQECNKKFMRS
Dps_GA19045       KTSHLRAHLRWHTGERPFVCSWVFCGKRFTRSDELQRHRRTHTGEKRFQCRECNKKFMRS
Fc_Sp1-4          KTSHLRAHLRWHAGERPFVCQWLFCGKRFTRSDELQRHRKTHTGEKRFQCPECLKKFMRS
Am_Sp1-4          KTSHLRAHLRWHTGERPFVCSWIFCGKKFTRSDELQRHRRTHTGEKRFQCPECTKKFMRS
Nav_Sp1-4         KTSHLRAHLRWHTGERPFVCNWIFCGKKFTRSDELQRHRRTHTGEKRFECPECTKKFMRS
Tc_Sp1-4          KTSHLRAHLRWHTGERPFVCSWLFCGKRFTRSDELQRHRRTHTGEKRFQCNECSKKFMRS
Ag_Sp1-4          KTSHLRAHLRWHTGERPFICNWGTCGKRFTRSDELQRHRRTHTGEKRFECVECNKKFMRS
Of_Sp1-4          KTSHLRAHLRWHTGERPFICSWEFCGKCFTRSDELQRHRRTHTGEKRFMCTQCHKKFMRS
Bm_Sp1-4          KTSHLRAHLRWHSGERPFLCNWLFCGKRFTRSDELQRHRRTHTGEKRFECPECSKRFMRS
Td_Sp1-4          KTSHLRAHLRWHTGERPFVCTWMFCGKRFTRSDELQRHRRTHTGEKRFQCPECSKKFMRS
Mm_Sp1            KTSHLRAHLRWHTGERPFMCNWSYCGKRFTRSDELQRHKRTHTGEKKFACPECPKRFMRS
Hs_Sp1            KTSHLRAHLRWHTGERPFMCTWSYCGKRFTRSDELQRHKRTHTGEKKFACPECPKRFMRS
Gg_Sp1            KTSHLRAHLRWHTGERPFICGWMLCGKRFTRSDELQRHKRTHTGEKKFACPECPKRFMRS
Dr_Sp1            KTSHLRAHLRWHTGERPFVCSWSFCGKRFTRSDELQRHKRTHTGEKKFSCTECPKRFMRS
Fr_Sp1            KTSHLRAHLRWHTGERPFVCSWSFCGKRFTRSDELQRHKRTHTGEKRFACPECPKRFMRS
Dp_Sp1-4          KTSHLRAHLRWHSGDRPFVCSWYYCGKRFTRSDELQRHRRTHTGEKRFQCPECGKRFMRS
Mm_Sp3            KTSHLRAHLRWHSGERPFICNWMFCGKRFTRSDELQRHRRTHTGEKKFVCPECSKRFMRS
Hs_Sp3            KTSHLRAHLRWHSGERPFVCNWMYCGKRFTRSDELQRHRRTHTGEKKFVCPECSKRFMRS
Gg_Sp3            KTSHLRAHLRWHSGERPFVCNWMFCGKRFTRSDELQRHRRTHTGEKKFVCPECSKRFMRS
```

```
Dr_Sp3            KTSHLRAHLRWHSGERPFICSWSYCGKRFTRSDELQRHRRTH-GEKKFVCPECSKRFMRS
Dr_Sp3-like       KTSHLRAHLRWHSGERPFICSWSYCGKRFTRSDELQRHRRTHTGEKKFVCPECSKRFMRS
Fr_Sp3            KTSHLRAHLRWHSGERPFVCSWMLCGKRFTRSDELQRHRRTHTGEKKFVCPECSKRFMRS
Fr_Sp3-like       KTSHLRAHLRWHSGERPFVCNWMFCGKRFTRSDELQRHRRTHTGEKKFVCMECSKRFMRS
Sp_Sp4            KTSHLRAHLRWHTGERPFVCDWLFCGKRFTRSDELQRHRRTHTGEKKFVCKSCGKKFMRS
Mm_Sp4            KTSHLRAHLRWHTGERPFICNWMFCGKRFTRSDELQRHRRTHTGEKRFECPECSKRFMRS
Hs_Sp4            KTSHLRAHLRWHTGERPFICNWMFCGKRFTRSDELQRHRRTHTGEKRFECPECSKRFMRS
Gg_Sp4            KTSHLRAHLRWHTGERPFICNWVFCGKRFTRSDELQRHRRTHTGEKRFECPECSKRFMRS
Dr_Sp4            KTSHLRAHLRWHTGERPFVCNWIFCGKRFTRSDELQRHRRTHTGEKRFECPECSKRFMRS
Fr_Sp4            KTSHLRAHLRWHTGERPFVCNWIFCGKRFTRSDELQRHRRTHTGEKRFECPECSKRFMRS
Bf_SpB            KTSHLRAHLRWHTGERPFVCNWLFCGKRFTRSDELQRHRRTHTGEKRFTCPDCSKKFMRS
Ph_Sp1-4          KTSHLRAHLRWHAGDRPFACNWMFCTKRFTRSDELQRHKRTHTGEKRFHCPDCDKKFMRS
Nv_Sp1-4          KTSHLRAHLRWHSGERPFVCNWLFCGKRFTRSDELQRHRRTHTGEKRFTCPECNKKFMRS
Ta_Sp1-4          KTSHLRAHLRWHTGERPFVCNWLFCGKRFTRSDELQRHRRTHTGEKKFACQDCGKRFMRS
Dm_Btd            KASHLKTHLRWHTGERPFLC--LTCGKRFSRSDELQRHGRTHTNYRPYACPICSKKFSRS
Dps_GA11738       KASHLKTHLRWHTGERPFLC--LTCGKRFSRSDELQRHGRTHTNYRPYACPICSKKFSRS
Ag_Sp5/Btd        KTSHLKAHLRWHTGERPFQCKWLFCGKRFTRSDELQRHFRTHTGEKRFTCTICSKKFMRS


                 121        130
Dm_D_Sp1           DHLAKHVKTH
Dps_GA11282        DHLAKHVKTH
Am_Sp6-9           DHLAKHVKTH
Nav_Sp6-9          DHLAKHVKTH
Ag_Sp6-9           DHLAKHVKTH
Td_Sp6-9           DHLAKHVKTH
Tc_Sp8             DHLAKHVKTH
Of_Sp8/9           ----------
Bm_Sp6-9           DHLAKHVKTH
Fc_Sp6-9           DHLSKHVKTH
Ph_Sp6-9           DHLSKHVKTH
Dp_Sp6-9           DHLSKHVKTH
Gg_Sp9             DHLSKHIKTH
Fr_Sp9             DHLSKHIKTH
Dr_Sp9             DHLSKHIKTH
Hs_Sp9             DHLSKHIKTH
Mm_Sp9             DHLSKHIKTH
Mm_Sp8             DHLSKHVKTH
Hs_Sp8             DHLSKHVKTH
Dr_Sp8-ike         DHLSKHVKTH
Gg_Sp8             DHLSKHVKTH
Dr_Sp8             DHLSKHVKTH
Fr_Sp8             DHLSKHIRTH
Bf_SpA             DHLAKHTKTH
Sp_Sp6-9           DHLSKHVKTH
Mm_Sp7             DHLSKHQRTH
Hs_Sp7             DHLSKHQRTH
Dr_Sp7             DHLSKHQKTH
Fr_Sp7             DHLSKHQKTH
Nv_Sp6-9           DHLSKHVKTH
Mm_Sp6             DHLAKHMKTH
Hs_Sp6             DHLAKHMKTH
Dr_Sp6             DHLAKHMRVH
Fr_Sp6             DHLAKHMRTH
Ta_Sp6-9           DHLSKHMKTH
Bm_Sp5/Btd         DHLAKHVKTH
Tc_Btd             DHLAKHVKTH
Fc_Sp5/Btd         DHLAKHVKTH
Am_Sp5/Btd         DHLTKHVKTH
Nav_Sp5/Btd        DHLNKHRKTH
Dr_Sim-to-Sp5      DHLAKHLKTH
Dr_Sp5             DHLAKHVKTH
Fr_Sp5             DHLAKHVKTH
Bf_Sp5             DHLAKHVKTH
Sp_Sp5/Btd         DHLSKHSKTH
Dp_Sp5/Btd         DHLSKHSKTH
Mm_Sp5             DHLAKHVKTH
Hs_Sp5             DHLAKHVKTH
Gg_Sp5             DHLAKHVKTH
Nv_Sp5/Btd         DHLRKHQKTH
Dr_Sp5-like        DHLSKHVKTH
Fr_Sp5-like        DHLSKHVKTH
Td_Sp5/Btd         DHLAKHVKTH
Mm_Sp2             DHLTKHYKTH
Hs_Sp2             DHLTKHYKTH
Gg_Sp2             DHLTKHYKTH
Dr_Sp2             DHLTKHYKTH
Fr_SP2             DHLTKHYKTH
Ta_Sp1-4           DHLSKHAKRH
Dm_CG5669          DHLSKHIKTH
Dps_GA19045        DHLSKHIKTH
Fc_Sp1-4           DHLSKHIKTH
Am_Sp1-4           DHLTKHIKTH
Nav_Sp1-4          DHLTKHIKTH
Tc_Sp1-4           DHLSKHLKTH
Ag_Sp1-4           DHLSKHIRTH
Of_Sp1-4           DHLSKHIKTH
Bm_Sp1-4           DHLAKHVRIH
Td_Sp1-4           DHLSKHIKTH
Mm_Sp1             DHLSKHIKTH
```

```
Hs_Sp1            DHLSKHIKTH
Gg_Sp1            DHLSKHIKTH
Dr_Sp1            DHLSKHIKTH
Fr_Sp1            DHLSKHIKTH
Dp_Sp1-4          DHLSKHIKTH
Mm_Sp3            DHLAKHIKTH
Hs_Sp3            DHLAKHIKTH
Gg_Sp3            DHLAKHIKTH
Dr_Sp3            DHLAKHIKTH
Dr_Sp3-like       DHLAKHIKTH
Fr_Sp3            DHLAKHIKTH
Fr_Sp3-like       DHLAKHIKTH
Sp_Sp4            DHLAKHQRTH
Mm_Sp4            DHLSKHVKTH
Hs_Sp4            DHLSKHVKTH
Gg_Sp4            DHLSKHVKTH
Dr_Sp4            DHLSKHIKTH
Fr_Sp4            DHLSKHIKTH
Bf_SpB            DHLSKHIKTH
Ph_Sp1-4          DHLSKHVKTH
Nv_Sp1-4          DHLSKHVKTH
Ta_Sp1-4          DHLSKHVKTH
Dm_Btd            DHLSKHKKTH
Dps_GA11738       DHLSKHKKTH
Ag_Sp5/Btd        DHLAKHVKTH
```

Figure S1. CLUSTAL X (1.81) multiple sequence alignment of different Sp factors comprising the conserved region of the Btd box and the zinc finger region.

3.2 The *buttonhead*-related gene *Sp1-4* is required for embryonic head development in *Tribolium castaneum*.

The purpose of this work was the analysis of the function and interaction of the three Sp transcription factor genes *Sp1-4*, *buttonhead* and *Sp8* in the red flour beetle *Tribolium castaneum*. Previous studies already characterized the function of *Tc Sp8* in leg development, but for *Tc btd* no RNAi phenotype could be described (see also 2.2). *Tc Sp1-4* had not been analyzed so far. Besides, potential interactions of these three genes remained to be analyzed. Therefore, dsRNA injections of *Tribolium* Sp genes were carried out. Besides single gene RNA injections also double injections were carried out to test for functional redundancy. Larval RNAi cuticles and embryos stained for several marker genes were analyzed.

Nina D. Schäper, Nikola-Michael Prpic, Ernst A. Wimmer

Status: manuscript in preparation

Author contribution to the practical work:
Nina Schäper did all the in situ hybridizations and performed most of the parental RNAi injections and embryo fixations.

The *buttonhead*-related gene *Sp1-4* is required for embryonic head development in *Tribolium castaneum*.

Nina D. Schaeper[1], Nikola-Michael Prpic[1], Ernst A. Wimmer[1]*

1 Georg-August-Universität, Johann-Friedrich-Blumenbach-Institut für Zoologie und Anthropologie, Abteilung Entwicklungsbiologie, GZMB, Ernst-Caspari-Haus, Justus-von-Liebig-Weg 11, 37077 Göttingen, Germany

(*) Author for correspondence:
Ernst A. Wimmer
Georg-August-Universität, Johann-Friedrich-Blumenbach-Institut für Zoologie und Anthropologie, Abteilung Entwicklungsbiologie, GZMB, Ernst-Caspari-Haus, Justus-von-Liebig-Weg 11, 37077 Göttingen, Germany

Tel.: +49 551 39 22889
FAX: +49 551 39 5416
e-mail: ewimmer@gwdg.de

Number of words: 3760 (in total)
Number of figures: 4
Number of tables: 1

Running head: *Tribolium* head development requires the *Sp1-4* gene

Summary

Head segmentation in *Drosophila melanogaster* involves the genes *orthodenticle* (*otd*), *empty-spiracles* (*ems*) and *buttonhead* (*btd*) [e.g. Cohen and Jürgens, 1990; Wimmer et al., 1993; Wimmer et al., 1996]. These genes act as gap-like genes defining larger areas of the head and orchestrating the further refinement of gene expression, resulting in the proper metamerisation of the head [Cohen and Jürgens, 1990; Finkelstein and Perrimon, 1991]. However, recent comparative studies of insect head development hint to a considerable diversity of head patterning modes [Lynch et al., 2006; McGregor, 2006]. Studies of head formation in the flour beetle *Tribolium castaneum* have shown that the beetle *btd* homolog *Tc btd* has no role in head development [Schinko et al., 2008]. Here we present a functional study of the *btd* paralog *Sp1-4* of *T. castaneum*. We show that this gene is required for head development in *T. castaneum*. The RNAi phenotype of *Tc Sp1-4* includes headless larvae, thus demonstrating that it is required for the formation of all head segments. We also show that the activation of the head gene *Tc collier* (*col*) requires *Tc Sp1-4* instead of *Tc btd*. This demonstrates how developmental functions can be switched between paralogs of the same gene functions.

Results and Discussion

Tc btd belongs to a larger family of genes, termed Sp gene family, and we have recently shown that *T. castaneum* has two additional members of this gene family [Schaeper et al., in preparation (a) (see 3.1)]. One of these, the gene *Tc Sp8*, has been studied previously and was shown to have a role in appendage development [Beermann et al., 2004]. However, the third Sp gene, termed *Sp1-4*, has not been studied previously, neither in *T. castaneum* nor in any other arthropod species. We were therefore interested in investigating the function of this gene in *T. castaneum* with a focus on the role of *Tc Sp1-4* in head development.

We have shown previously that *Tc Sp1-4* is expressed ubiquitously throughout the embryonic development of *T. castaneum* [Schaeper et al., in preparation (a) (see 3.1)]. We have used parental double-stranded RNA mediated interference (pRNAi) [Bucher et al., 2002] to study the function of *Tc Sp1-4*. In *Tc Sp1-4* depleted larvae we detected a spectrum of head defects comprising several degrees of head reduction (Fig. 1). The weakest degree of head reduction was represented by larvae with a recognizable, but severely reduced head capsule (Fig. 1B). In these larvae, all head appendages are present, but are malformed and abnormally crowded. Stronger degrees of head reduction were represented by larvae with no recognizable head capsule (Fig. 1C, D). In these larvae some head appendages were also lacking. In some cases the identity of the remaining appendages could still be inferred from their position and shape

(Fig. 1C), but in other cases only remnants of unrecognizably deformed head appendages were present (Fig. 1D). The strongest degree of head reduction was seen in larvae lacking a head altogether (Fig. 1E). In these larvae the body begins with the first thoracic segment and no traces of head segments were visible in the cuticles. In addition to the defective or missing head, *Tc Sp1-4* RNAi larvae show twisted thoracic appendages with split or branched claw tips (Fig. 1B-E, and data not shown). The obtained phenotypes were grouped into two classes. Larvae in which the head is reduced in size but a head capsule was still recognizable were grouped into class I, whereas class II contains all larvae with no head capsule and a varying degree of loss of head appendages, including headless larvae (see Tab. 1).

These data show that this Sp family gene has a significant role in head development in *T. castaneum* being required for the formation of all head segments. In order to exclude the possibility that a head role of the remaining two Sp family genes *Tc btd* and *Tc Sp8* has gone unnoticed in previous studies, we performed pRNAi experiments with these genes as well (Fig. 2). In agreement with previous work [Schinko et al., 2008], we did not detect any effect on larval head morphology after RNAi with either *Tc btd* or *Tc Sp8* (Fig. 2C, E). In *Tc btd* RNAi larvae the head appendages were also normal, but the thoracic legs were mildly malformed showing fusions of the femur with the tibiotarsus (Fig. 2D). The degree of the *Tc btd* phenotype is not variable and therefore we included all affected larvae in one class (class I) (see Tab. 1). In *Tc Sp8* RNAi larvae the head appendages (except for the mandible) have distal defects, most prominently seen in the antenna and the maxillary palps (Fig. 2E) [Beermann et al., 2004]. Consistent with previous results [Beermann et al., 2004] the legs of *Tc Sp8* RNAi larvae were shortened (Fig. 2F, and data not shown). Strongly affected *Tc Sp8* larvae were grouped into class II, whereas milder phenotypes were grouped into class I (see Tab. 1).

The proteins encoded by the three Sp family genes in *T. castaneum* share a very similar region containing three zinc fingers as putative DNA binding motifs. Based on these similarities all three proteins might have a similar DNA binding activity in vivo and are able to regulate similar sets of target genes. Indeed, Schöck et al. (2000) could show a complete rescue of the *D. melanogaster btd* mutant phenotype to wildtype by providing an engineered form of Btd in which the *Dm* Btd zinc finger motif had been replaced by the human Sp1 zinc finger motif. Interestingly, human Sp1 belongs to the Sp1-4 clade of Sp transcription factors and *Dm* Btd to the Sp5/Btd clade [Schaeper et al., in preparation (a) (see 3.1)].

There might thus be a certain degree of functional redundancy of the Sp genes of different clades. Indeed, partial functional redundancy of Sp family factors has been demonstrated in *D. melanogaster*, in which *D-Sp1* (the ortholog of *Tc Sp8*) and *btd* can partially substitute for each other in mediating leg growth [Estella et al., 2003]. Functional redundancies might mask a head role of *Tc btd* or *Tc Sp8*, or might mask an additional role of *Sp1-4* in the formation of the thorax or the abdomen. In order to test for possible functional redundancies in *T. castaneum*, we have

performed combinatorial double-injections. Double injections of dsRNA of *Tc Sp1-4* and *Tc btd* produced larvae with reduced heads, reduced head appendages and twisted legs (Fig. 3A) and thus virtually identical to larvae produced by *Tc Sp1-4* RNAi alone. However, the leg phenotype of *Tc btd* is anyway mild and we therefore cannot exclude that it is unrecognizable in the twisted and malformed legs of the double RNAi larvae. Double injections of dsRNA of *Tc Sp1-4* and *Tc Sp8* also produced larvae with reduced heads, reduced head appendages and twisted legs (Fig. 3B). These larvae are also indistinguishable from larvae resulting from *Tc Sp1*-4 RNAi alone. This might be the result of the reduced concentrations of each dsRNA in double injection experiments, because the total volume that is injected into the animals cannot be doubled and thus it is possible that the dsRNA of one of the two Sp genes in double injections outcompetes the other one. By contrast, the double injections of ds RNA of *Tc btd* and *Tc Sp8* produced larvae with a normal head, but shortened legs (Fig. 3C) that are more severely shortened than in larvae resulting after RNAi with *Tc Sp8* alone. Thus, in this case the effects of *Tc Sp8* RNAi and *Tc btd* RNAi seem to enhance each other, indicating that the two genes have partially redundant functions in leg development, similar to the situation in *D. melanogaster* [Estella et al., 2003].

We then attempted to better understand the molecular basis of the larval phenotypes after RNAi and have studied in RNAi embryos the expression of several genes known to be involved in head development in *T. castaneum*: *Tc otd* [Cohen and Jürgens, 1990], *Tc col* [Economou and Telford, 2009], *Tc cnc* [Economou and Telford, 2009; Schaeper et al, in preparation (see 3.5)], *Tc lab* [Nie et al., 2001], and *Tc kni* [Cerny et al., 2008]. In addition we studied the expression of the segmentation genes *Tc eve* [Brown et al., 1997], *Tc wg* [Nagy and Carroll, 1994] and *Tc hh* [Farzana and Brown, 2008] that are also expressed during head formation. We were not able to detect any obvious changes in the expression of *otd*, *cnc*, *lab*, *kni* and *eve* in either the single injections or the double injections (data not shown). *Tc wg* and *Tc hh* are reduced in the head region of *Tc Sp1-4* RNAi embryos (Fig. 4B and 4G; and data not shown). Significant changes after RNAi could be observed for *Tc col*. In the wildtype, this gene is expressed in a stripe at the boundary between the mandibular and intercalary segment (Fig. 4A). This expression domain is significantly reduced in *Tc Sp1-4* RNAi embryos (Fig. 4B), as well as in *Tc Sp1-4/Tc btd* RNAi embryos (Fig. 4C) and *Tc Sp1-4/Tc Sp8* RNAi embryos (Fig. 4D). The *Tc col* expression domain is unchanged in *Tc btd* RNAi embryos (not shown), *Tc Sp8* RNAi embryos (not shown), and in *Tc btd/Tc Sp8* RNAi embryos (Fig. 4E). Interestingly, the *col* gene is activated by *btd* in *D. melanogaster* and *col* expression is eliminated in *btd* mutants [Crozatier et al., 1999; Tallafuss et al., 2001]. This suggests that *Tc Sp1-4* replaces *Tc btd* in *Tc col* activation.

In addition, we have studied the expression of the leg gene *Tc Dll* in the RNAi embryos, because leg defects were present in all RNAi larvae. This gene is expressed in the wildtype in all appendages (including the labrum), except for the mandible (Fig. 4F) and is required for leg

formation [Beermann et al., 2001]. Although the limbs are malformed and shorter than in the wildtype, the expression level of *Tc Dll* in the segmnetal appendages is normal in *Tc Sp1-4* (Fig. 4G), *Tc btd* (not shown), *Tc Sp8* (not shown), *Tc Sp1-4/Tc btd* (Fig. 4H), and *Tc Sp1-4/Tc Sp8* RNAi embryos (Fig. 4I). However, the *Tc Dll* expression in the labrum is completely abolished in *Tc Sp1-4* RNAi embryos and in *Tc Sp1-4/Tc btd* and *Tc Sp1-4/Tc Sp8* RNAi embryos, indicating that *Tc Sp1-4* is required for *Tc Dll* expression in this non-segmental appendage-like structure. In *Tc Sp8/Tc btd* RNAi embryos the expression domain of *Tc Dll* is significantly reduced and restricted to the tips of the appendages (Fig. 4J), consistent with the strong leg size reduction seen in *Tc Sp8/Tc btd* RNAi larvae.

Recent studies in *T. castaneum* have already indicated that the head patterning mechanisms in this species differ from those in *D. melanogaster*. The global head organizing function of *bcd* is in *T. castaneum* replaced by a maternal contribution of *Tc otd* and *Tc hb* [Schröder, 2003]. Intriguingly, the reduced head phenotypes after *Tc Sp1-4* RNAi are virtually identical to the *otd* RNAi phenotypes [Schröder, 2003]. This suggests that *Tc otd* and *Tc Sp1*-4 function on similar levels in the same genetic circuit during head patterning. Since *Tc otd* is expressed normally in *Tc Sp1-4* RNAi embryos we suggest that *Tc Sp1-4* functions directly downstream of *Tc otd*, but upstream of factors determining smaller areas of the head, like e.g. *Tc col*. Additional studies are necessary to establish the roles of *Tc otd* and *Tc hb* in regulating *Tc Sp1-4* in the head.

Conclusion

In our previous study, we have shown that the Sp-family genes in the Metazoa fall into three large clades: Sp1-4, Sp5/Btd, and Sp6-9 [Scaeper et al., in preparation (see 3.1)]. The *btd* gene of *D. melanogaster* is a member of the Sp5/Btd clade and is involved in head development in the fly. Previous work, however, has demonstrated that the homologous genes from this Sp5/Btd clade in the beetle *T. castaneum* and the mouse are not involved in head development. *T. castaneum btd* RNAi larvae have a normal head and only mild leg defects [Schinko et al. 2008; this work] and mice deficient for *Sp5* do not show any obvious phenotype [Treichel et al., 2003]. Intriguingly, the role in head development is performed by another Sp-family gene from one of the remaining clades in *T. castaneum* as well as in the mouse. In mice lacking *Sp8*, a member of the Sp6-9 clade, a strong malformation of the head is observed [Treichel et al., 2001], and we could show here that the loss of *Sp1-4* function in *T. castaneum* results in headless larvae. These data therefore strongly suggest that the developmental role in head formation has been exchanged between the members of this family of paralogous genes several times during evolution.

Despite the ubiquitous expression pattern of *Tc Sp1-4* our results show that the gene has a tissue specific function. This suggests that the gene is not regulated on the level of transcription, and instead implies spatially regulated translation of the *Sp1-4* mRNA or a requirement of spatially restricted co-factors for the Sp1-4 protein. A similar mode of regulation has been suggested previously for another Sp-family gene, btd, in *D. melanogaster*. Misexpression of *btd* in the entire anterior portion of embryos mutant for btd rescues the phenotype, but without disturbing development in areas where btd is not normally expressed [Wimmer et al., 1997].

Experimental Procedures

Cloning of *T. castaneum Sp* genes and phylogenetic analyses

The sequences of the *T. castaneum* Sp family genes can be obtained from GenBank: *Tc Sp1-4* (accession number XM_967159). *Tc btd* (accession number NM_001114320.1), *Tc Sp8* (accession number NM_001039420).

The isolation and phylogenetic analysis of the *T. castaneum* Sp family genes has been reported previously [Schaeper et al., in preparation (see 3.1)]. A 842bp fragment of *Tc Sp1-4* was isolated via PCR using the primers Sp1-4_fw1 GTC TGG GTA ACG TGC AAG TA ATCG CGC and Sp1-4_re1 TCA GCT ATC AAG CGG ATC AGC AAT GAT CAG. To obtain two non-overlapping fragments of this gene, we used the primer pairs Sp1-4_fw1 + Sp1-4_re2 CTG AGC GCG TGA AGC GCT TCC CAC (444bp) and Sp1-4_fw2 AGT TGC AAA GAC ATA GGA GGA CAC + Sp1-4_re1 (395bp). The primers btd_fw1 GTG ACT ACT ATG ATG GCT TCC TCG GTC C and btd_re2 CAC CTC AAG TGG GCC TGG AGA TGC GAG G were used to amplify a 509bp fragment of *Tc btd*. Two non-overlapping fragments were amplified using btd_fw1 + btd_re5 CAC CTC CTC GTT CAC GCA GTT CGG (408bp) and btd_fw2 GGG CTC AAG AAA CCC TCC TCG AAG + btd_re4 CCC AT TCCC GCA CAC ATT CTC AC (333bp). For *Tc Sp8*, the specific primers Sp8_fw1 CGC ATG CTG TTC CGG TCC ATC TCA TCG C and Sp8_re2 GGC CAG TGT AAC GTC TCT GTG ACC TGG were used to amplify a 1064bp fragment. Two non-overlapping fragments of *Tc Sp8* were isolated using the primers Sp8_fw1 + Sp8_re3 GGC GTT CGA CGT CAC TGA ATC ACT C (766bp) as well as Sp8_fw2 GGA GTT GGC TGG ATA TGA GTG GAG G and Sp8_re1 GGG CCT CAC GCT AGC GAT GAT AGC CTG G (840bp).

In situ hybridization

The 842bp fragment of *Tc Sp1-4*, 509bp of *Tc btd* and 1064bp of *Tc Sp8* were used for digoxygenin-labeled RNA probe synthesis (Roche). Fluorescein-labeled probes were used for

double stainings using INT/BCIP (Roche). In situ hybridization was performed essentially as described in [Wohlfrom et al., 2006].

Adult RNAi

DsRNA for *T. castaneum* parental RNAi was synthesised using the MEGAscript T7 Kit (Ambion) and resuspended in 1x injection buffer (1,4mM NaCl, 0,07mM Na_2HPO_4, 0,03mM KH_2PO_4, 4mM KCl) at a concentration of 4 µg/µl. Injections of adult female beetles was performed as described in [Schaeper et al., in preparation (b)]. Injection of 1x injection buffer served as a negative control for *T. castaneum* RNAi. Injection of dsRNA transcribed from the non-overlapping gene fragments (see above for fragment lengths) served as a control for potential off-target effects and lead to identical results as with the full fragments. Efficient degradation of mRNA below the level of detection was confirmed by in situ hybridization.

Acknowledgements

We thank Gregor Bucher for maintaining the Göttingen Tribolium Facility, the members of the department for helpful discussions, as well as Inga Karch for helping with *T. castaneum* parental RNAi injections and embryonic fixations. This work has been funded by the European Community's Marie Curie Research Training Network ZOONET under contract MRTN-CT-2004-005624 (to EAW), and a grant from the Deutsche Forschungsgemeinschaft (PR 1109/1-1 to NMP).

References

Beermann, A., Jay, D.G., Beeman, R.W., Hülskamp, M., Tautz, D., Jürgens, G. (2001). The Short antennae gene of *Tribolium* is required for limb development and encodes the orthologue of the *Drosophila* Distal-less protein. Development 128, 287-297.

Beermann, A., Aranda, M., Schröder, R. (2004). The *Sp8* zinc-finger transcription factor is involved in allometric growth of the limbs in the beetle *Tribolium castaneum*. Development 131, 733-742.

Brown, S.J., Parrish, J.K., Beeman, R.W., Denell, R.E. (1997). Molecular characterization and embryonic expression of the *even-skippe*d ortholog of *Tribolium castaneum*. Mech. Dev. 61, 165-173.

Bucher, G., Scholten, J., Klingler, M. (2002). Parental RNAi in *Tribolium* (Coleoptera). Curr. Biol. 12, R85-86.

Cerny, A.C., Grossmann, D., Bucher, G., Klingler, M. (2008). The *Tribolium* ortholog of *knirps* and *knirps-related* is crucial for head segmentation but plays a minor role during abdominal patterning. Dev. Biol. 321, 284-294.

Cohen, S.M., Jürgens, G. (1990). Mediation of *Drosophila* head development by gap-like segmentation genes. Nature 346, 482-485.

Crozatier, M., Valle, D., Dubois, L., Ibnsouda, S., Vincent, A. (1999). Head versus trunk patterning in the *Drosophila* embryo; *collier* requirement for formation of the intercalary segment. Development 126, 4385-4394.

Economou, A.D., Telford, M.J., (2009). Comparative gene expression in the heads of *Drosophila melanogaster* and *Tribolium castaneum* and the segmental affinity of the *Drosophila* hypopharyngeal lobes. Evol. and Dev. 11, 88-96.

Estella, C., Rieckhof, G., Calleja, M., Morata, G. (2003). The role of *buttonhead* and *Sp1* in the development of the ventral imagical discs of *Drosophila*. Development 130, 5929-5941.

Farzana, L., Brown, S.J. (2008). Hedgehog signaling pathway function conserved in Tribolium segmentation. Dev Genes Evol 218, 181-192.

Finkelstein, R., Perrimon, N. (1991). The molecular genetics of head development in *Drosophila melanogaster*. Development 112, 899-912.

Lynch, J.A., Brent, A.E., Leaf, D.S., Pultz, M.A., Desplan, C. (2006). Localized maternal *orthodenticle* patterns anterior and posterior in th elong germ wasp *Nasonia*. Nature, 439, 728-732.

McGregor, A.P. (2005). How to get ahead: the origin, evolution and function of *bicoid*. Bioessays 27, 904-913.

McGregor, A.P. (2006). Wasps, beetles and the beginning of the ends. Bioessays 28, 683-686.

Nagy, L.M., Carroll, S. (1994).Conservation of wingless patterning functions in the short-germ embryos of *Tribolium castaneum*. Nature 367, 460-463.

Nie, W., Stronach, B., Panganiban, G., Shippy, T., Brown, S., Denell, R. (2001). Molecular characterization of *Tclabial* and the 3' end of the Tribolium homeotic complex. Dev. Genes Evol. 211, 244-251.

Prpic, N.M., Janssen, R., Damen, W.G.M., Tautz, D. (2005). Evolution of dorsal–ventral axis formation in arthropod appendages: *H15* and *optomotor-blind/bifid*-type T-box genes in the millipede *Glomeris marginata* (Myriapoda: Diplopoda). Evol. Dev. 7, 51-57.

Prpic, N.M., Damen, W.G.M., (2005). Cell Death During Germ Band Inversion, Dorsal Closure, and Nervous System Development in the Spider *Cupiennius salei.* Dev. Dyn. 234, 222-228.

Schaeper, N.D., Prpic, N.M., Wimmer, E.A. (In preparation (a)). A clustered set of three Sp-family genes is ancestral in the Metazoa: evidence from sequence analysis, protein domain structure, chromosomal location and developmental mRNA expression.

Schaeper, N.D., Pechmann, M., Damen, W.G.M., Prpic, N.M., Wimmer, E.A. (In preparation (b)). Evolutionary plasticity of *collier* function in head development of diverse arthropods.

Schinko, J.B., Kreuzer, N., Offen, N., Posnien, N., Wimmer, E.A., Bucher, G. (2008). Divergent functions of *orthodenticle, empty spiracles* and *buttonhead* in early head patterning of the beetle *Tribolium castaneum* (Coleoptera). Dev. Biol. 317, 600-613.

Schöck, F., Purnell, B.A., Wimmer, E.A., Jäckle, H. (1999). Common and diverged functions of the *Drosophila* gene pair *D-Sp1* and *buttonhead*. Mech. Dev. 89, 125-132.

Schöck, F., Reischl, J., Wimmer, E., Taubert, H., Purnell, B.A., Jäckle, H. (2000). Phenotypic suppression of *empty spiracles* is prevented by *buttonhead*. Nature 405, 351-354.

Schröder, R. (2003). The genes *orthodenticle* and *hunchback* substitute for *bicoid* in the beetle *Tribolium*. Nature 422, 621-625.

Tallafuss, A., Wilm, T.P., Crozatier, M., Pfeffer, P., Wassef, M., Bally-Cuif, L. (2001). The zebrafish buttonhead-like factor Bts1 is an early regulator of *pax2.1* expression during mid-hindbrain development. Development 128, 4021-4034.

Treichel, D., Becker, M.B., Gruss, P. (2001). The novel transcription factor gene *Sp5* exhibits a dynamic and highly restricted expression pattern during mouse embryogenesis. Mech Dev 101, 175-179.

Treichel, D., Schöck, F., Jäckle, H., Gruss, P., Mansouri, A. (2003) *mBtd* is required to maintain signaling during murine limb development. Genes Dev 17, 2630-2635.

Wimmer, E.A., Jäckle, H., Pfeifle, C., Cohen, S.M. (1993). A *Drosophil*a homologue of human Sp1 is a head-specific segmentation gene. Nature 366, 690-694.

Wimmer, E.A., Frommer, G., Purnell, B.A., Jäckle, H. (1996). *buttonhead* and *D-Sp1*: a novel *Drosophila* gene pair. Mech. Dev. 59, 53-62.

Wimmer E.A., Cohen, S.M., Jäckle, H., Desplan, C. (1997). *buttonhead* does not contribute to a combinatorial code proposed for *Drosophila* head development. Development 124, 1509-1517.

Wohlfrom, H., Schinko, J.B., Klingler, M., Bucher, G. (2006). Maintenance of segment and appendage primordia by the *Tribolium* gene *knödel*. Mech. Dev. 123, 430-439.

Figures

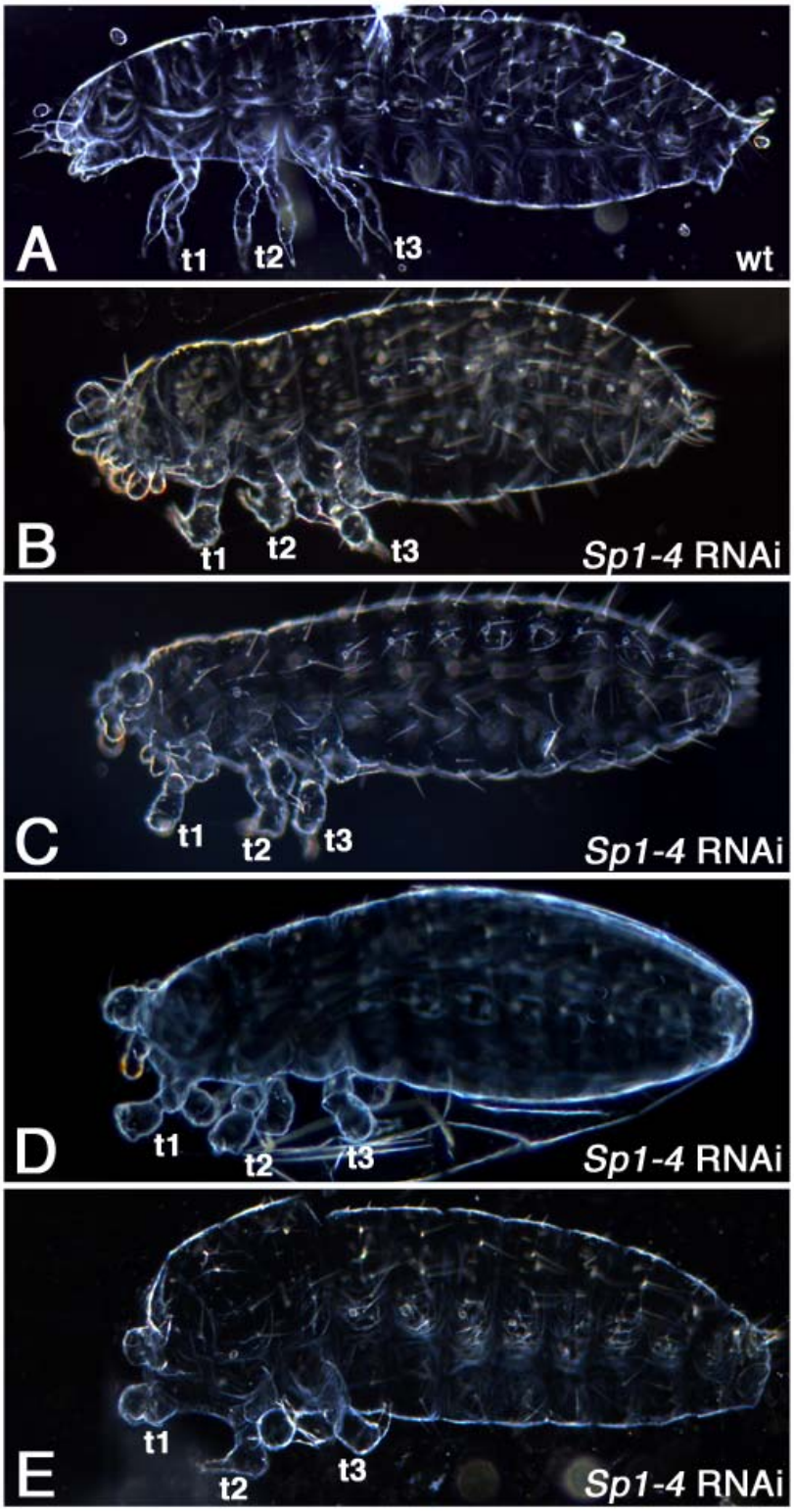


Figure 1

Figure 1. Head defects resulting from *Tc Sp1-4* RNAi. (A) Wildtype larval cuticle preparation of *T. castaneum*. (B-E) Larval cuticles after *Tc Sp1-4* RNAi. (B) The head capsule is visible but reduced in size. The head appendages are malformed (also note the prominent split labrum), but are all present and crowded in front of the reduced head capsule. (C-D) No head capsule is present and some of the head appendages are lacking. In (C) the remaining head appendages are the (split) labrum, antennae and the labium. In (D) the large protusion likely is a remnant of the labrum, the other outgrowth cannot be identified with confidence. (E) Completely headless larva; the body begins with the prothoracic segment.
Note the twisted appendages in all RNAi larvae. Anterior is to the left in all panels. Abbreviations: t1, prothoracic leg; t2, mesothoracic leg; t3 metathoracic leg.

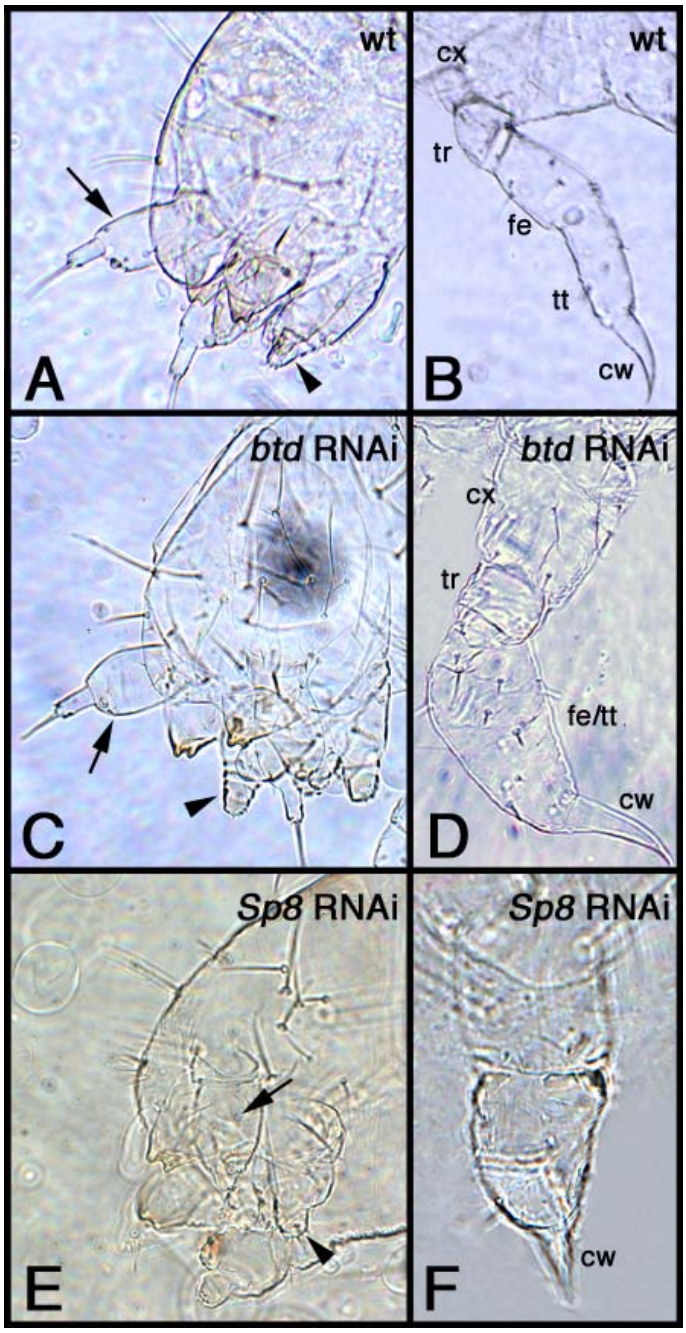


Figure 2. Appendage defects resulting from RNAi with *Tc btd* and *Tc Sp8*. (A) Head of a wildtype larva. (B) Leg of a wildtype larva. (C) Head of a *Tc btd* RNAi larva. No difference to the wildtype can be detected. (D) Leg of a *Tc btd* RNAi larva. Note the fusion of the femur and the tibiotarsus leading to a thickened fusion podomere. (E) Head of a *Tc Sp8* RNAi larva. The head capsule is normal, but the distal portion of the head appendages (except for the mandibles) are malformed. This is most obvious in the antenna (arrow) and the maxillary palp (arrowhead). (F) Leg of a *Tc Sp8* RNAi larva. The leg is significantly shorter than wildtype legs. Abbreviations: cx, coxa; tr, trochanter; fe, femur; tt, tibiotarsus; cw, pretarsal claw. The arrow and arrowhead in all panels point to the antenna and maxillary palp, respectively.

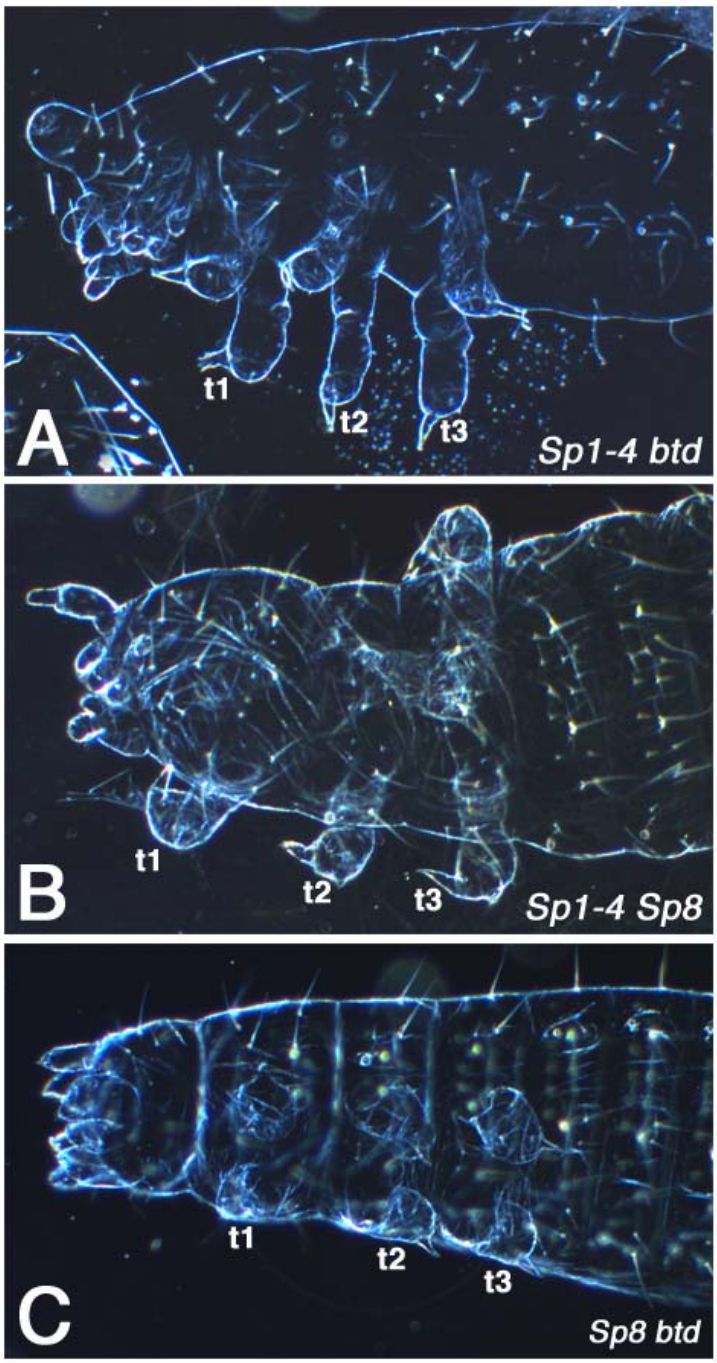


Figure 3. Head and appendage defects produced by combinatorial double RNAi. Larval cuticle after (A) *Tc Sp1-4/Tc btd* RNAi, and (B) *Tc Sp1-4/Tc Sp8* RNAi. In both cases, no head capsule is present and the head is much reduced. The head and thoracic appendages are malformed. (C) Larval cuticle after *Tc Sp8/Tc btd* RNAi. The head capsule is normal, but the head appendages are affected in their distal portion and the legs are severely shortened, being reduced to small round outgrowths with a small claw at the tip. Anterior is to the left in all panels. Abbreviations see Fig. 1.

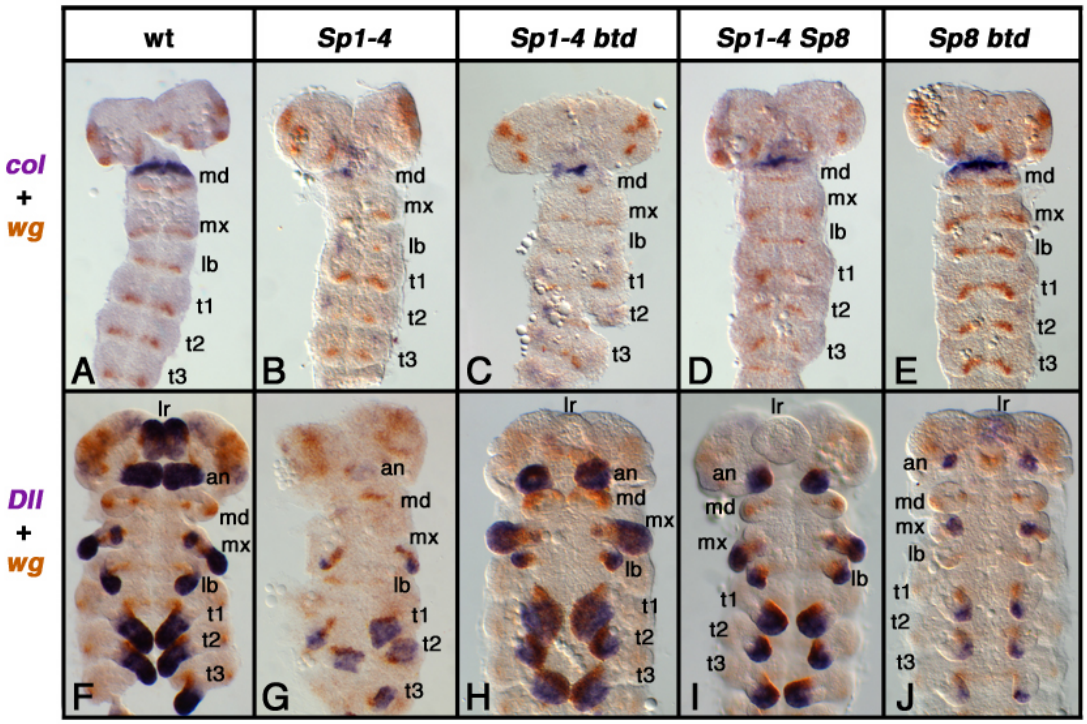


Figure 4. Molecular marker expression in RNAi embryos. (A-E) Expression of *Tc col* (blue) and *Tc wg* (orange). In wildtype embryos *Tc col* is expressed strongly in the area between mandibular and intercalary segment (A). Expression of *Tc col* is reduced in *Tc Sp1-4* RNAi (B), *Tc Sp1-4/Tc btd* RNAi (C), and *Tc Sp1-4/Sp8* RNAi embryos (D). Expression of *Tc col* is normal in *Tc Sp8/Tc btd* embryos (E). (F-J) Expression of *Tc Dll* (blue) and *Tc wg* (orange). In wildtype embryos *Tc Dll* is expressed in the appendages (except for the mandibles). Expression of *Tc Dll* is relatively normal in *Tc Sp1-4* RNAi (G), *Tc Sp1-4/Tc btd* RNAi (H), and *Tc Sp1-4/Sp8* RNAi embryos (I), despite visible defects in the appendages. Note, however, the lack of *Tc Dll* expression in the labrum of these RNAi embryos. Expression of *Tc Dll* is reduced and restricted to the distal end of the appendages in *Tc Sp8/Tc btd* RNAi embryos (J). Anterior is up in all panels. Abbreviations: lr, labrum; an, antenna; md mandible; mx, maxilla; lb, labium; t1, prothoracic leg; t2, mesothoracic leg; t3 metathoracic leg.

Tables

Table 1. Statistical analysis of RNAi phenotypes. The phenotypes have been grouped into classes according to their severity. Explanations of the classes given in the table are as follows. *Tc Sp1-4* RNAi: class I includes larvae with reduced head, but a still discernible head capsule; class II includes larvae with no head capsule and a varying degree of loss of head appendages, including headless larvae. *Sp8* RNAi: class I includes shortened legs with still some leg segment borders visible; class II includes shortened legs with no visible leg segment borders. *Tc btd* RNAi: only one class of phenotype: fused femur and tibiotarsus. *Tc Sp1-4* + *Tc Sp8* RNAi and *Tc Sp1-4* + *Tc btd* RNAi: the two classes correspond to the two classes of *Tc Sp1-4* RNAi. *Tc Sp8* + *Tc btd* RNAi: the two classes correspond to the two classes of *Tc Sp8* RNAi.

Gene	WT *n* (%)	class I *n* (%)	class II *n* (%)	Total *n*
Sp1-4	22 (3,2%)	176 (25,3%)	490 (71,5%)	696
Sp8	126 (19,3%)	328 (50,1%)	200 (30,6%)	654
btd	231 (42,8%)	309 (57,2%)	n.a.	540
Sp1-4+Sp8	76 (10,8%)	332 (47,2%)	296 (42%)	704
Sp1-4+btd	12 (1,8%)	158 (39%)	392 (59,2%)	662
Sp8+btd	90 (13,1%)	230 (33,4%)	368 (53,5%)	688

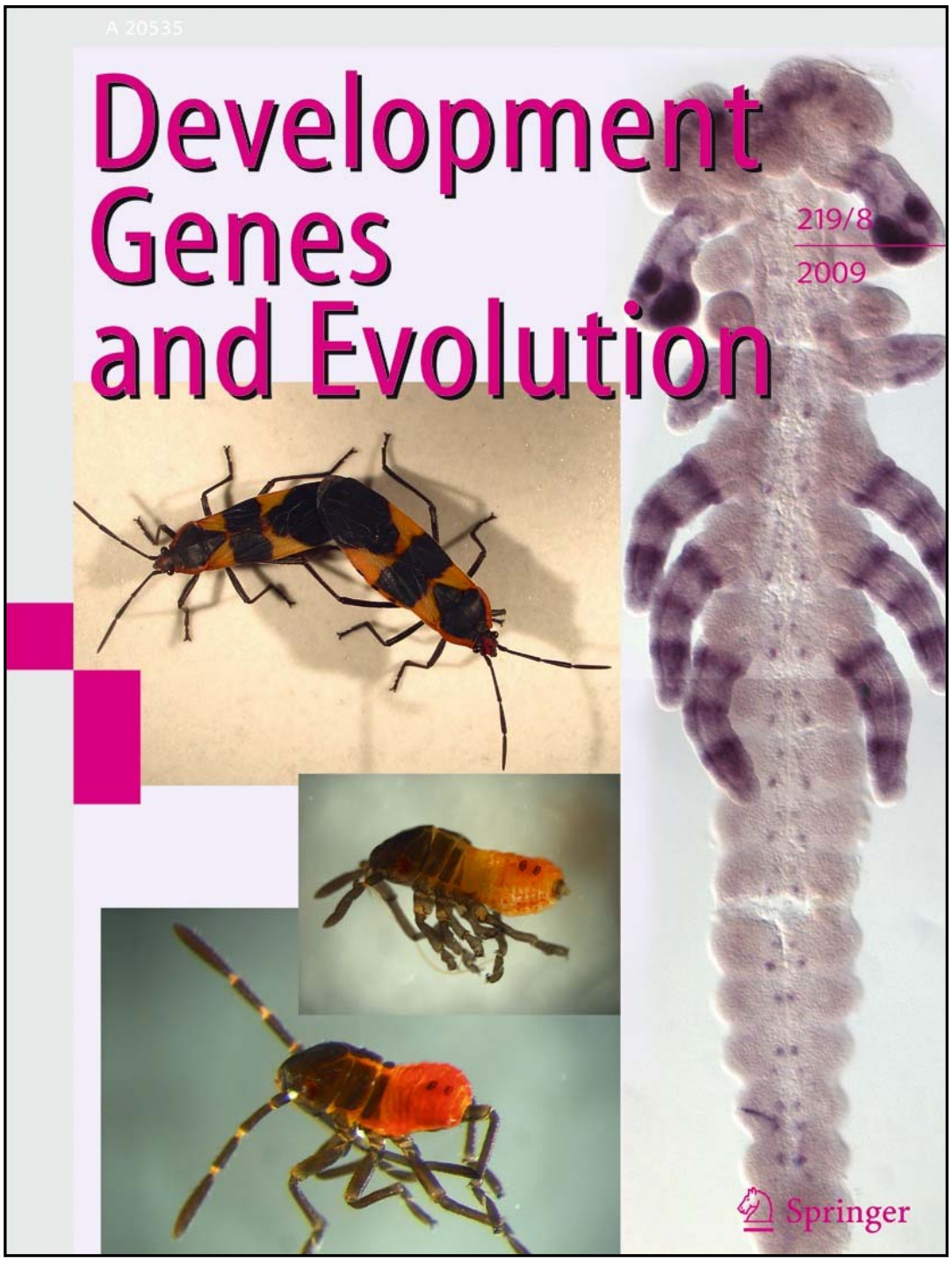
3.3 A conserved function of the zinc finger transcription factor *Sp8/9* in allometric appendage growth in the milkweed bug *Oncopeltus fasciatus.*

The purpose of this work was the analysis of the expression pattern and function of the Sp transcription factor *Sp8/9* in the milkweed bug *Oncopeltus fasciatus.* Therefore, a *Sp8/9* homolog was isolated from *O. fasciatus* and the spatio-temporal expression pattern was analyzed. Also parental RNAi experiments were performed, followed by embryonic and nymphal phenotype analyses.

Nina D. Schäper, Nikola-Michael Prpic, Ernst A. Wimmer

Status: published in “Development, Genes and Evolution”
(2009) 219:427-439

Author contribution to the practical work:
Nina Schäper performed all the experiments


A 20535
Development Genes and Evolution
219/8
2009
Springer

Dev Genes Evol (2009) 219:427–435
DOI 10.1007/s00427-009-0301-0

SHORT COMMUNICATION

A conserved function of the zinc finger transcription factor Sp8/9 in allometric appendage growth in the milkweed bug *Oncopeltus fasciatus*

Nina D. Schaeper · Nikola-Michael Prpic · Ernst A. Wimmer

Received: 9 June 2009 / Accepted: 30 August 2009 / Published online: 16 September 2009

Abstract The genes encoding the closely related zinc finger transcription factors Buttonhead (Btd) and D-Sp1 are expressed in the developing limb primordia of *Drosophila melanogaster* and are required for normal growth of the legs. The *D-Sp1* homolog of the red flour beetle *Tribolium castaneum*, *Sp8* (appropriately termed *Sp8/9)*, is also required for the proper growth of the leg segments. Here we report on the isolation and functional study of the *Sp8/9* gene from the milkweed bug *Oncopeltus fasciatus*. We show that *Sp8/9* is expressed in the developing appendages throughout development and that the downregulation of *Sp8/9* via RNAi leads to antennae, rostrum, and legs with shortened and fused segments. This supports a conserved role of *Sp8/9* in allometric leg segment growth. However, all leg segments including the claws are present and the expression of the leg genes *Distal-less*, *dachshund*, and *homothorax* are proportionally normal, thus providing no evidence for a role of *Sp8/9* in appendage specification.

Keywords Sp transcription factors · Leg development · Allometric organ growth · Appendage evolution · Insect development

Communicated by S. Roth

N. D. Schaeper · N.-M. Prpic · E. A. Wimmer (✉)
Georg-August-Universität,
Johann-Friedrich-Blumenbach-Institut für Zoologie
und Anthropologie, Abteilung Entwicklungsbiologie, GZMB,
Ernst-Caspari-Haus, Justus-von-Liebig-Weg 11,
37077 Göttingen, Germany
e-mail: ewimmer@gwdg.de

Introduction

The members of the *Sp* gene family encode evolutionarily conserved proteins, which are characterized by the presence of three zinc finger motifs (for nucleic acid binding) as well as an additional short conserved motif called Btd box (a transactivation domain), and are involved in a variety of developmental processes in both insects and vertebrates (reviewed in Zhao and Meng 2005). In *Drosophila*, two different Sp factor encoding genes, *buttonhead* (*btd*) and *D-Sp1*, have been shown to be expressed during leg development (Wimmer et al. 1996; Schöck et al. 1999; Estella et al. 2003). The *btd* gene is required for the specification and formation of the entire leg disc primordia (Estella et al. 2003; McKay et al. 2009). The gene is acting upstream of the well-characterized leg specification gene *Distal-less* (*Dll*) and seems to be capable of activating the entire leg developmental cascade when expressed ectopically (Estella et al. 2003). The role of the *D-Sp1* gene during leg development is less clear, but seems to be partially redundant with *btd* (Schöck et al. 1999; Estella et al. 2003). The homolog of *D-Sp1* in the red flour beetle *Tribolium castaneum*, *Sp8*, has been shown to be involved in the control of allometric growth of the leg segments, but the evidence for a role of *Sp8* in the specification of the entire leg primordium is ambiguous (Beermann et al. 2004). Orthologous genes to *D-Sp1* and *Tc Sp8* have also been identified in the mouse, termed *Sp8* and *Sp9* (Bell et al. 2003; Treichel et al. 2003; Kawakami et al. 2004). Please note that the murine *Sp8* gene has also been called inaccurately *mBtd* (Treichel et al. 2003), because it was named at a time when its orthology to either *Drosophila btd* or *Drosophila D-Sp1* was unclear. Intriguingly, both murine genes are also involved in limb outgrowth.

Here we present the isolation of an Sp factor encoding gene from the milkweed bug *O. fasciatus*. Although our

analysis of the *Oncopeltus* gene confidently places it as a homolog of the *Drosophila D-Sp1*, *Tribolium Sp8*, and the murine *Sp8* and *Sp9* genes, the exact orthology within this group is unclear and we therefore designate the *Oncopeltus* gene as *Sp8/9*. We have studied the role of the *Oncopeltus Sp8/9* ortholog in order to investigate its evolutionary conservation in insect appendage development. We find that *Sp8/9* is expressed in the appendages throughout development and the downregulation of *Sp8/9* expression via RNAi leads to shortened legs, rostrum, and antennae. These data show that *Sp8/9* is involved in the allometric growth of the appendages. However, we find no evidence for a more global role of *Sp8/9* in appendage specification.

Materials and methods

Animal husbandry and embryology

Milkweed bugs were reared as described previously (Hughes and Kaufman 2000). Collected eggs were kept at 25°C. Embryos of all stages were fixed as reported in Liu and Kaufman (2004a). *Oncopeltus* embryo dissections before in situ staining were performed under a fluorescence stereomicroscope using SYTOX Green nucleic acid stain (Invitrogen) (Liu and Kaufman 2004b).

Isolation of *Sp8/9* and sequence analysis

Oncopeltus embryos from 0 to 96 h were used for mRNA isolation using the MicroPoly(A)Purist kit (Ambion). This mRNA served as template for double-stranded (ds) cDNA synthesis (SMART PCR cDNA Synthesis kit, Clontech) and RACE template synthesis (SMART RACE cDNA Amplification Kit, Clontech). PCR with the primer pair Fw_GGC MGG GCI ACI TGY GAY TGY CCI AAY TG (GRATCDCPNC) and Rev_ARR TGR TCI SWI CKC ATR AAI CKY AA (LHDSRMFRK) resulted in a fragment of 311 bp. PCR fragments were cloned into the pCRII vector (Invitrogen). Additional sequence information was obtained by 5′ RACE PCR using the reverse primer CAG GTG AGC CTT GAG GTG CGA GGT C. Phylogenetic analysis of different Sp factor sequences was performed as described previously (Prpic et al. 2005). The *Oncopeltus Sp8/9* sequence is available from the EMBL nucleotide database under the accession number FN396612.

In situ hybridization

The longest 5′ RACE fragment of *Sp8/9* (1,078 bp comprising 181 bp 5′UTR and 897 bp ORF) served as template for the synthesis of digoxygenin-labeled RNA probes (Roche). In situ hybridization was performed as described previously (Liu and Kaufman 2004a).

Parental RNA interference

The template for dsRNA synthesis was prepared by PCR with T7 (GAA TTG TAA TAC GAC TCA CTA TAG G) and Sp6-T7 (TAA TAC GAC TCA CTA TAG GAT TTA GGT GAC ACT ATA GA) primers from the longest 5′ RACE fragment of *Sp8/9* that has also been used for probe generation. Double-stranded RNA (dsRNA) was generated using the MEGAscript T7 Kit (Ambion) and resuspended in 1× injection buffer (1.4 mM NaCl, 0.07 mM Na_2HPO_4, 0.03 mM KH_2PO_4, 4 mM KCl) at a concentration of 4 μg/μl. RNA injections in adult virgin *Oncopeltus* females were performed as described previously (Liu and Kaufman 2004a). Injection of 1× injection buffer served as the negative control. To verify the RNAi phenotypes obtained with the full fragment, we repeated the parental RNAi with two shorter non-overlapping fragments of *Of Sp8/9*. A different 5′ RACE *Of Sp8/9* fragment of 1,057 bp (133 bp 5′ UTR, 924 bp ORF) was cut with XhoI which resulted in two fragments of 421 and 636 bp. Parental RNAi experiments with dsRNA transcribed from these two fragments resulted in the same phenotype with a similar frequency as for dsRNA injections with the full longest 5′ RACE fragment (data not shown). As independent RNAi controls, we performed injections of dsRNA of EGFP and *Of eve* which resulted in no abnormal phenotype or the same phenotypes as previously published for *Of eve*, respectively (Liu and Kaufman 2005) (data not shown).

Results and discussion

Isolation of the *Sp8/9* homolog of *O. fasciatus*

A short fragment of the *Oncopeltus* homolog of the *Sp8/9* gene was isolated by PCR. More sequence information was obtained by 5′ RACE PCR. The available sequence comprises 181 bp of 5′ UTR followed by 1,008 bp of protein coding sequence also encoding the Btd box motif and the zinc finger motifs ZF 1, ZF 2, and a portion of ZF 3. An alignment of these conserved domains with the nine Sp factors from the mouse, as well as D-Sp1 and Btd from *Drosophila* and *Tc* Sp8 and *Tc* Btd (Schinko et al. 2008) from *Tribolium*, shows that the Btd box motif and the available sequences of the zinc finger motifs are almost identical in all Sp8/9 proteins, but clearly different from the homologous motifs in other Sp proteins (Fig. 1a). The linker regions between these motifs are also highly conserved in all Sp8/9 proteins, whereas differences in sequence and linker length exist in other Sp proteins (Fig. 1a).

Dev Genes Evol (2009) 219:427–435

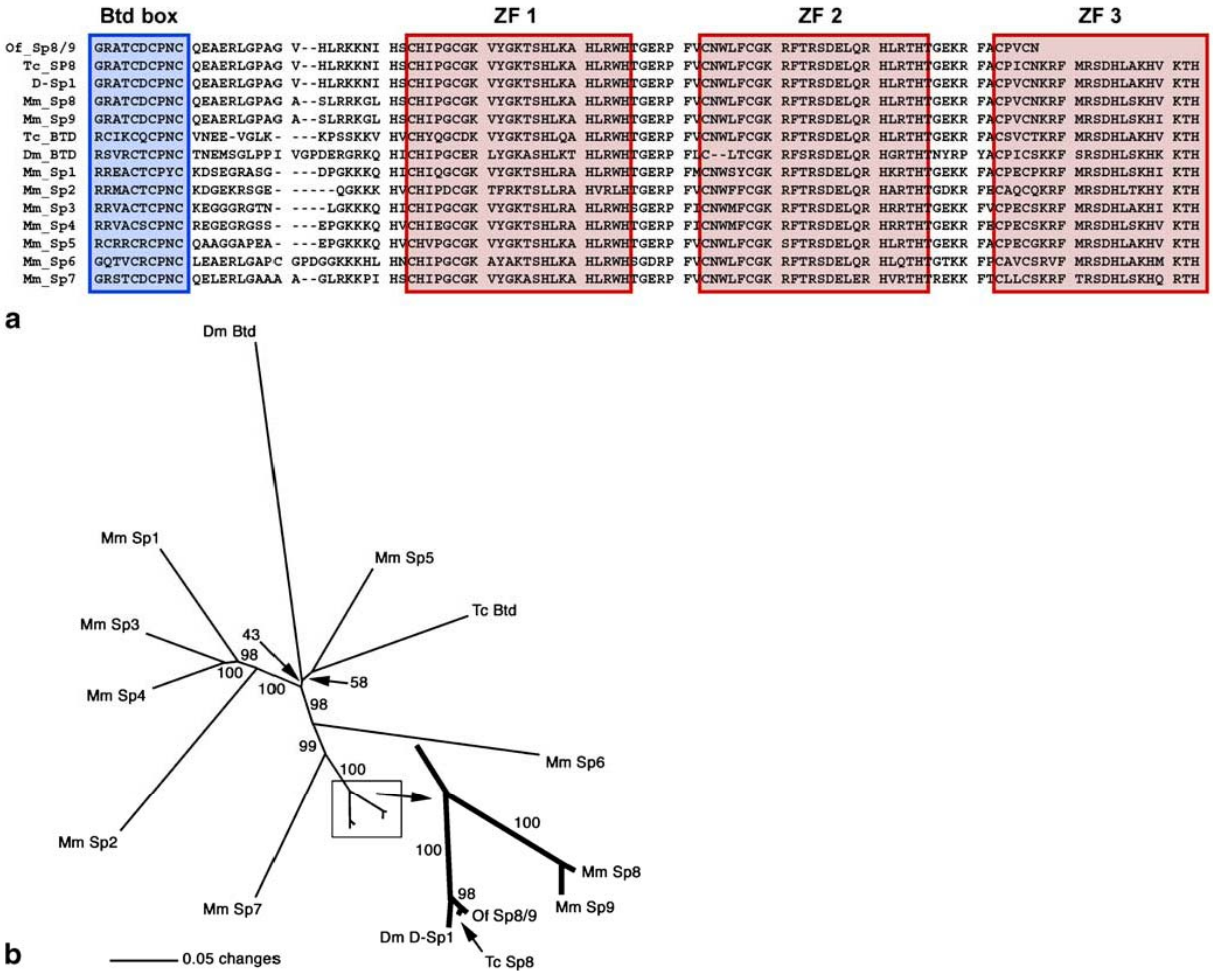


Fig. 1 Alignment and phylogenetic analysis of the Sp8/9 sequence from *Oncopeltus*. **a** Alignment of Sp proteins from *Drosophila*, *Oncopeltus*, *Tribolium*, and mouse. Shown is the protein region containing the Btd box motif (*blue*) and the three zinc finger motifs (*red*). The sequences of the Sp8/9 proteins including the D-Sp1 protein are almost identical and clearly different from other Sp proteins. **b** Phylogenetic analysis of all murine Sp factors and Sp factors from *Tribolium*, *Drosophila*, and *Oncopeltus* based on the alignment shown in (**a**). Shown is the unrooted majority rule consensus computed from 1,000 intermediate trees produced with the quartet puzzling method (Strimmer and von Haeseler 1996). The numbers at the tree edges indicate the reliability values. GenBank accession numbers: *Dm_btd* NP_511100, *D-Sp1* NP_572579, *Tc_btd* NP_001107792, *Tc_Sp8* NP_001034509, *Mm_Sp1* NP_038700, *Mm_Sp2* NP_084496, *Mm_Sp3* NP_035580, *Mm_Sp4* NP_033265, *Mm_Sp5* NP_071880, *Mm_Sp6* NP_112460, *Mm_Sp7* NP_569725, *Mm_Sp8* NP_796056, *Mm_Sp9* NM_001005343. Species abbreviations: *Mm Mus musculus*, *Tc Tribolium castaneum*, *Of Oncopeltus fasciatus*, *Dm Drosophila melanogaster*

In order to further corroborate the orthology of the *Sp8/9* fragment isolated from *Oncopeltus*, we also performed a phylogenetic analysis (Fig. 1b) using the alignment in Fig. 1a in a maximum likelihood analysis with Tree Puzzle (Strimmer and von Haeseler 1996). Most edges in the phylogenetic tree are well supported with reliability values above 95. The Sp8/9 factors from *Oncopeltus*, *Tribolium*, and mouse cluster together in a group supported by the maximum reliability value of 100 and with very short edges, indicating that these genes are closely related. This grouping also includes D-Sp1 from *Drosophila*. These results further support the orthology of the *Oncopeltus Sp8/9* gene with the other *Sp8/9* genes and also give additional evidence to the previously published conclusion that the *Drosophila D-Sp1* gene is actually the *Drosophila Sp8/9* homolog (Beermann et al. 2004).

Sp8/9 is expressed during appendage development in *O. fasciatus*

Next we examined the expression profile of *Sp8/9* in *Oncopeltus* embryos using whole mount *in situ* hybridization. During early stages of germband elongation, strong expression of *Sp8/9* is seen in the limb buds of the thoracic

legs and the antennae (Fig. 2a). Interestingly, the expression in the thoracic limb buds extends from there into the lateral tissue of the thoracic segments posterior to the limb buds, and thus is not restricted to the limb buds (Fig. 2a, arrows). A weaker expression is detected in the maxillary and labial segment, where distinct buds are not yet developed. A separate expression domain is present in the anterior head in the region where the labrum will develop. As the germband grows, a punctate expression in the ventral nervous system appears and the expression in the thorax becomes restricted to the leg limb buds (Fig. 2b). During later germband elongation stages, the punctate expression pattern in the ventral nervous system becomes stronger and some cells in the brain also express *Sp8/9* (Fig. 2c). These brain cells are located at the rim of the head lobes in the area where the optic centers develop and also nearer to the ventral midline in the protocerebrum (Fig. 2c). All limb buds, including the emerging buds of the labrum, the mandibles, and the maxillae, now express *Sp8/9*. In fully elongated embryos, the expression pattern persists, but the

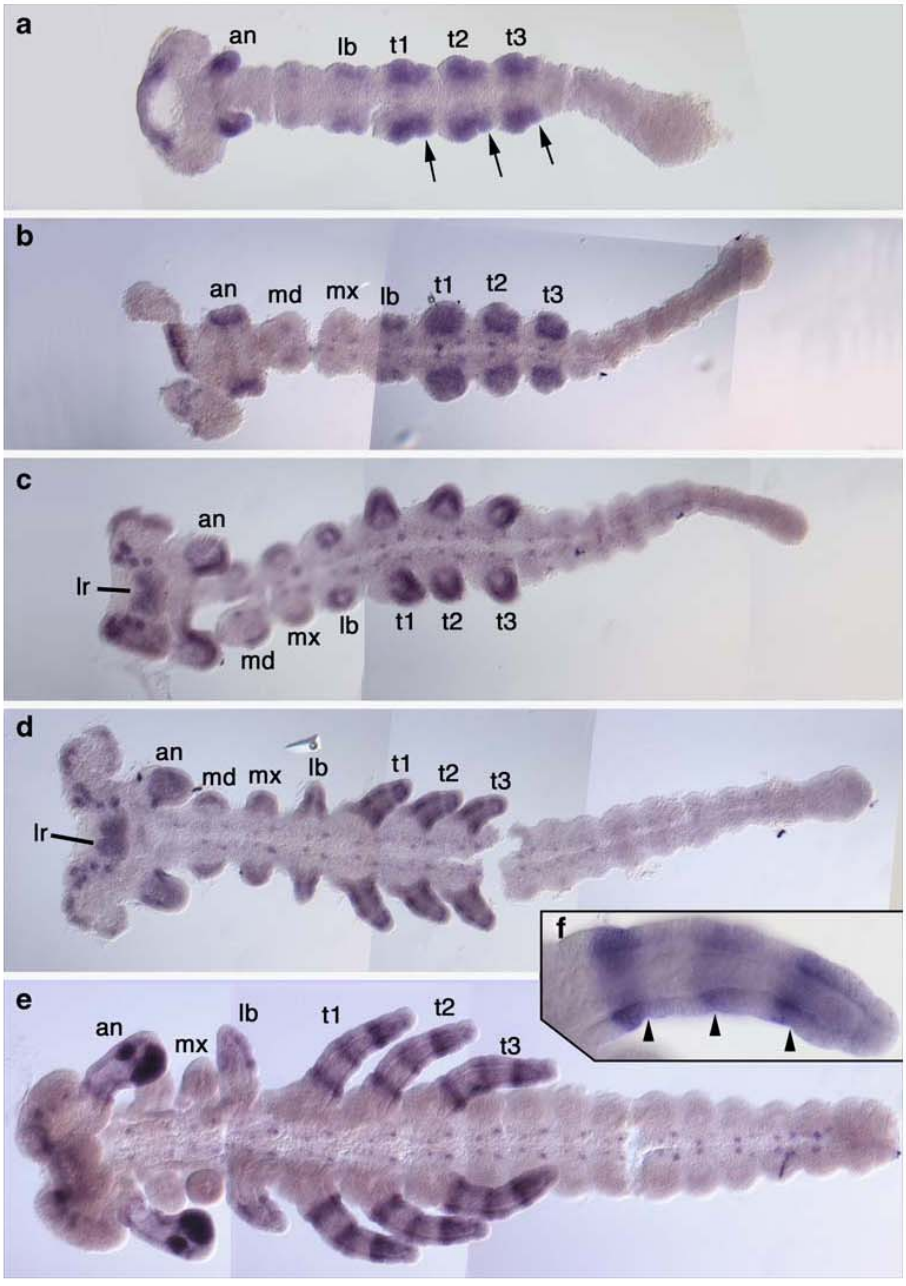


Fig. 2 Expression of *Sp8/9* in *Oncopeltus* embryos. **a** Early germband elongation stage. *Arrows* point to expression extending from the thoracic limb buds into the lateral tissue of the thoracic segments posterior to the limb buds. **b** Mid-elongation germband stage. **c** Embryo approaching the final stages of germband elongation. **d** Fully elongated germband stage. **e** Retracted germband stage. **f** Magnification of a leg at the retracted germband stage showing the expression rings of Sp8/9 abutting the presumptive joint constrictions (*arrowheads*). All embryos are oriented with the anterior end to the *left*. Abbreviations: *lr* labrum, *an* antenna, *md* mandible, *mx* maxilla, *lb* labium, *t1* prothoracic leg, *t2* mesothoracic leg, *t3* metathoracic leg

expression in the thoracic legs becomes slightly non-homogeneous (Fig. 2d), indicating the emergence of iterated rings in the legs at later stages. At fully retracted germband stages, the expression pattern in the appendages is more restricted (Fig. 2e). In the antenna, *Sp8/9* is restricted to a large ventral patch at the distal end and a smaller dot dorsal to it. The expression in the mandible has ceased and the expression in the maxilla is very weak. In the labium, *Sp8/9* expression is largely confined to a faint ring near the middle of the appendage. In the thoracic legs, the expression is seen in three rings. These expression rings are located proximally adjacent to the constrictions that mark the presumptive joints between the leg segments (podomeres; note that not all future podomeres are yet developed at this stage of development) (Fig. 2f).

Functional analysis of *Sp8/9* using RNAi

In order to establish the function of *Sp8/9* during development of *Oncopeltus*, we have applied parental RNAi (Liu and Kaufman 2004a). The results of these injections are summarized in Table 1. In wild-type embryos, shortly before hatching, the appendages lie flat on the body and reach the posterior end of the abdomen (Fig. 3d, e). In *Sp8/9* RNAi embryos, the appendages are severely shortened. They stick out (Fig. 3a) and do not reach further than the middle of the body (Fig. 3a, b).

In wild-type hatchlings the appendages are long and composed of several segments (podomeres) (Fig. 3f). The rostrum is a complex of four appendages (Fig. 3g). The labrum is thin and sharply pointed, the mandibles and the maxillae are long and thread-like, and the labium consists of four segments (Fig. 3g). The rostrum of the *Sp8/9* RNAi animals is malformed (Fig. 3h). The labium is shortened; the distal segments are fused, bent, and enlarged at the tip. Therefore, the filiform mandibles and maxillae protrude from the labium at the distal end, while normally they are entirely ensheathed by it. The overall morphology of the mandibles, maxillae, and the labrum is normal in *Sp8/9* RNAi animals, but they are shorter than in the wild type (Fig. 3h).

The wild-type antennae consist of a basal antennifer and four antennal segments (Fig. 3i). The antennae are severely affected in all *Sp8/9* RNAi animals (Fig. 3c, j). The antennifer and the first segment of the antenna are roughly identical in size and shape to the wild type, but the three distal antennal segments are entirely fused and the antennal appendage as a whole is severely shortened. In about half of the cases, the fused distal antennal portion displays small ectopic outgrowths (arrow in Fig. 3j).

The wild-type thoracic legs consist of a short coxa, a trochanter which is closely attached to the femur, a tibia, and a two-segmented tarsus with two claws (Fig. 3k). The legs of *Sp8/9* RNAi animals are much shorter than in the wild type (Fig. 3c, l, m). All leg segments are present but some of them are fused together. The tarsal segments are always fused, but they can still be distinguished because constrictions indicating the rudimentary joints are still present. The trochanter and femur are also always fused. In some specimens, the proximal podomeres are so severely malformed and fused that they cannot be distinguished anymore (Fig. 3m). In contrast to these joints, the joints between femur and tibia and between tibia and tarsus are always present.

Expression of *Distal-less*, *dachshund*, and *homothorax* in *Sp8/9* RNAi embryos

Since the results of the RNAi experiments indicated a role of *Sp8/9* in appendage development, we next studied the expression of leg developmental genes expressed at different positions along the proximal–distal leg axis.

The gene *Distal-less* (*Dll*) is expressed in the distal portion of all appendages except for the mandible (Fig. 4a, b) (Angelini and Kaufman 2004). In the head appendages, *Dll* is expressed in the tip of the labrum, the entire distal portion of the antenna, and the tips of maxilla and labium (Fig. 4a). The thoracic legs show two separate domains of *Dll* expression: a large distal domain ("sock") and a ring of expression proximal to it (Fig. 4b). The antennae are shorter in *Sp8/9* RNAi embryos, but they express *Dll* in their distal region (Fig. 4c). In addition, *Dll* is expressed in the distal tips of labrum, maxilla, and labium in *Sp8/9* RNAi embryos (Fig. 4c). The thoracic legs are much shorter in *Sp8/9* RNAi embryos than in the wild type, but they show a "ring and sock" pattern of *Dll* expression similar to the wild-type legs (Fig. 4d).

The *dachshund* (*dac*) gene is in the wild-type expressed in a medial ring in the thoracic legs, and in

Table 1 Parental RNAi with *Sp8/9*

	WT *n* (%)	Unspecific phenotype *n* (%)	Appendage phenotype *n* (%)	Total *n*
Sp8/9 pRNAi	4 (2.4%)	6 (3.6%)	160 (94%)	170
Control (injection buffer)	104 (95%)	5 (5%)	0 (0%)	109

Given are the total numbers (*n*) of specimens in each category. *WT* stands for wild-type specimens after injection. Unspecific phenotypes were specimens that showed some development, but had died before any discernible morphology was formed. These phenotypes had a similar frequency in the *Sp8/9* pRNAi and in the control injections. The appendage phenotype was obtained only in the *Sp8/9* pRNAi

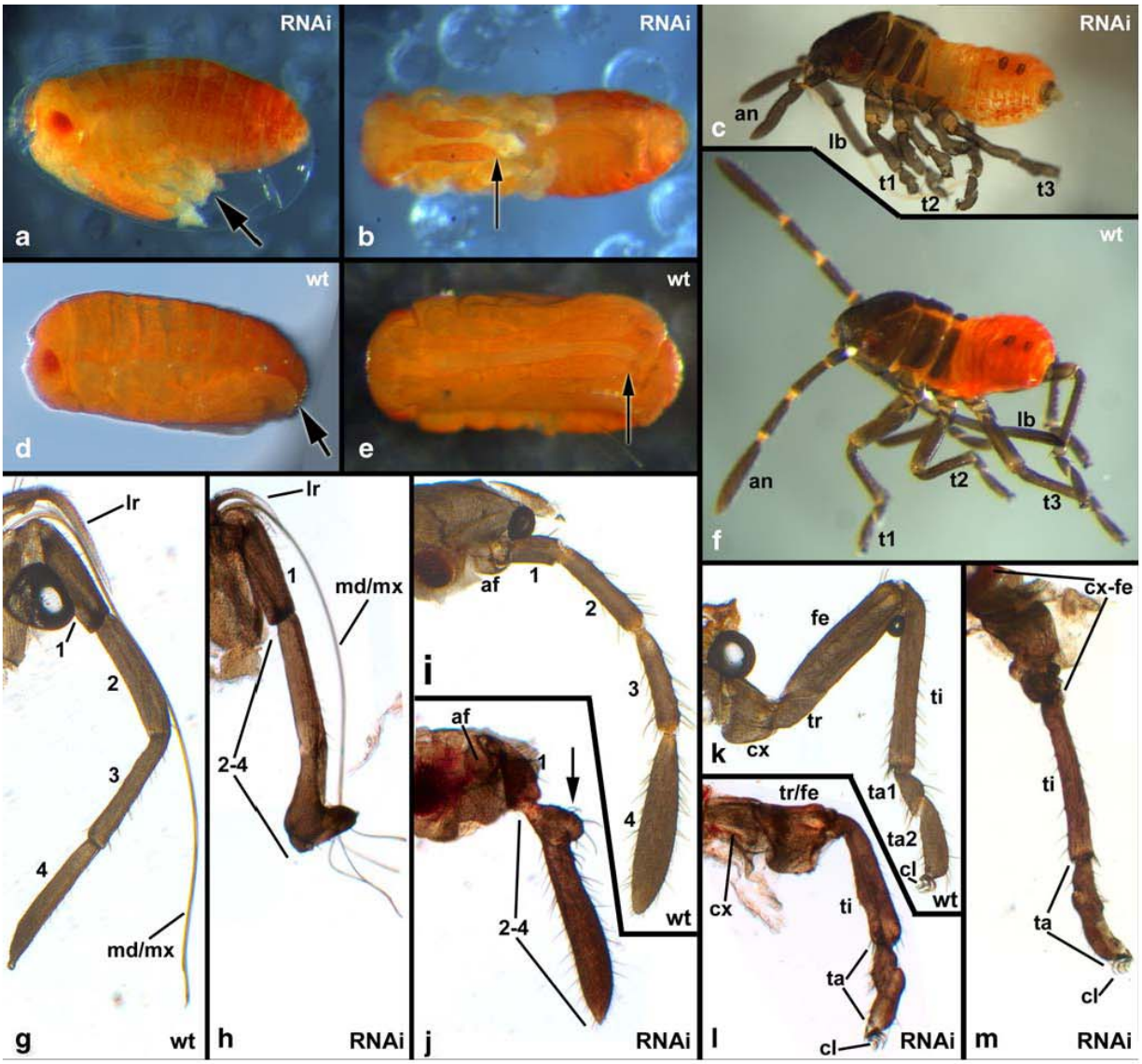


Fig. 3 Parental RNAi with *Sp8/9* in *Oncopeltus*. **a**, **b** Embryos after *Sp8/9* RNAi shortly before hatching in lateral (**a**) and ventral (**b**) aspect. **c** Hatchling after *Sp8/9* RNAi. **d**, **e** Wild-type embryos shortly before hatching in lateral (**d**) and ventral (**e**) aspect. **f** Wild-type hatchling. The *arrows* in (**a**) and (**d**) denote the tips of the legs. The *arrows* in (**b**) and (**e**) denote the tips of the antennae. All animals in (**a**)–(**f**) are oriented with the anterior end to the *left*. **g**–**m** Preparations of the appendages of wild-type hatchlings (**g**, **i**, **k**) and hatchlings after *Sp8/9* RNAi (**h**, **j**, **l**, **m**). The wild-type rostrum (**g**) comprises the four-segmented labium, filiform maxillae and mandibles, and the pointed labrum. The rostrum of RNAi animals (**h**) is malformed with fused distal segments. The wild-type antenna (**i**) comprises an antennifer and four additional segments. The antenna of RNAi animals is severely shortened with fused segments (**j**). The *arrow* in (**j**) points to a small ectopic outgrowth on the antenna seen in about half of the hatchlings after *Sp8/9* RNAi. The wild-type legs (**k**) comprise a coxa, trochanter, femur, tibia, and a two-segmented tarsus with two claws. The legs of RNAi animals are shortened with fused tarsal segments (**l**, **m**). The more proximal segments coxa, trochanter, and femur may be partially (**l**) or fully fused together (**m**). Abbreviations: *lr* labrum, *an* antenna, *af* antennifer, *ro* rostrum, *md* mandible, *mx* maxilla, *lb* labium, *t1* prothoracic leg, *t2* mesothoracic leg, *t3* metathoracic leg, *cx* coxa, *tr* trochanter, *fe* femur, *ti* tibia, *ta* tarsus, *cl* claw

a thin ring near the base in the labium and in the antenna (Fig. 4e) (Angelini and Kaufman 2004). The *dac* gene is also expressed very strongly in the mandible and in the maxilla (Fig. 4e). In the *Sp8/9* RNAi animals, this pattern is not significantly altered. The expression ring in the antenna is thinner, but the expression in the mandible, maxilla, and labium is unchanged (Fig. 4f, g). The thoracic legs are significantly shortened, but the medial ring

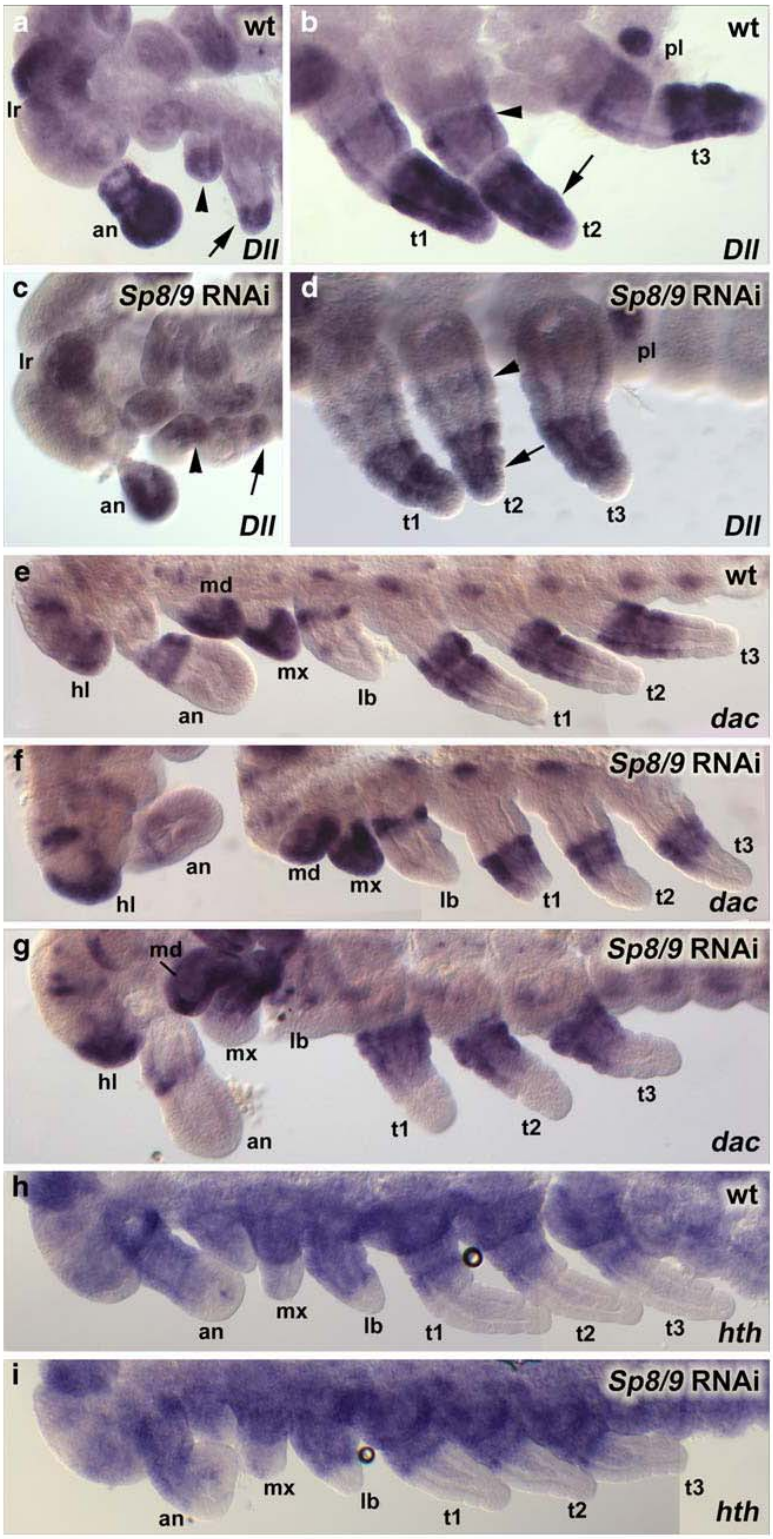


Fig. 4 Expression of *Distal-less*, *dachshund*, and *homothorax* in *Sp8/9* RNAi embryos. **a**, **b** Wild-type *Dll* expression in the head region (**a**) and the thoracic legs (**b**). **c**, **d** Expression of *Dll* in *Sp8/9* RNAi embryos. Please note that the brownish staining in the head and the bases of all appendages (including the mandibles) in (**c**, **d**) is unspecific background sometimes produced during the staining procedure. Embryos in (**a**)–(**d**) are at the retracted germband stage. Note the expression in the tips of the maxilla and labium (*arrowhead* and *arrow*, respectively, in (**a**) and (**c**)). Also note the expression in a distal "sock" and a proximal "ring" in the thoracic legs (*arrow* and *arrowhead*, respectively, in (**b**) and (**d**)). **e**–**g** Expression of *dac* in wild-type (**e**) and *Sp8/9* RNAi embryos (**f**, **g**). Please note that the embryo in (**f**) has been damaged during preparation (the head has come off and the t1 leg is damaged). The embryos in (**e**) and (**f**) are at the retracted germband stage and the embryo in (**g**) is at an even later retracted germband stage. **h**–**i** Expression of *hth* in wild-type (**h**) and *Sp8/9* RNAi embryos (**i**). Both embryos are at the retracted germband stage. All animals are oriented with the anterior end to the *left*. Abbreviations: *hl* head lobe, *lr* labrum, *an* antenna, *md* mandible, *mx* maxilla, *lb* labium, *t1* prothoracic leg, *t2* mesothoracic leg, *t3* metathoracic leg, *pl* pleuropodium

of *dac* expression is present as in the wild-type legs (Fig. 4f, g).

The gene *homothorax* (*hth*) is in the wild type expressed in the proximal area of all appendages (Fig. 4h), but the distal extension of expression is different in the different appendage types (Angelini and Kaufman 2004). Gene expression is restricted to the proximal third of the legs, but is expressed in the proximal two thirds of the appendage in antennae, mandible, and labium and fills the proximal half of the maxilla (Fig. 4h). This proportional pattern is identical in *Sp8/9* RNAi animals although the legs are shorter than in the wild type (Fig. 4i).

A conserved role in appendage axis elongation

Previous work in *Drosophila* has shown that the *D-Sp1* gene is expressed in the thoracic limb primordia in the embryo and in the leg imaginal discs in the larva (Wimmer et al. 1996; Schöck et al. 1999; Estella et al. 2003). In the leg discs, *D-Sp1* is expressed in concentric rings that roughly correspond to the position of the future joints between the leg segments. This expression pattern is very similar to the late expression of *Sp8/9* in the legs of the beetle *T. castaneum*. In this insect species, *Sp8/9* is expressed in up to four segmental rings that lie at a similar location in the legs as the expression rings of the *Serrate* (*Ser*) gene, which encodes the ligand of the Notch (N) receptor (Beermann et al. 2004). RNAi experiments with *Sp8/9* have shown that the podomeres are severely shortened and sometimes fused, and that the number of *Ser* rings is also reduced. Based on these data, Beermann et al. (2004) suggested that *Sp8/9* is involved in the control of allometric growth of the individual leg segments, probably by interfering with the Notch pathway, which is known to control allometric podomere growth in *Drosophila* and other arthropod species (de Celis et al. 1998; Rauskolb and Irvine 1999; Bishop et al. 1999; Prpic and Damen 2009). It has been noted previously that the process of leg segment growth is tightly linked with the process of joint formation, because both processes are regulated by the Notch pathway and its targets (Milán and Cohen 2000), and this would then also explain the observed podomere fusions.

Our results in *Oncopeltus* support an evolutionarily conserved role of *Sp8/9* in the control of allometric podomere growth. First, the *Sp8/9* expression pattern in the legs after full germband elongation is very similar to the pattern in *Tribolium* and consists of several rings adjacent to the constrictions of the future leg joints. Second, after RNAi, all leg segments are still present but severely shortened compared to the wild-type legs. In addition, the podomere fusions observed in the legs and in the antennae are compatible with the notion that the Notch pathway is also affected, because similar podomere fusion phenotypes are observed in *Drosophila* when members of effectors of the Notch pathway are impaired (e.g., de Celis et al. 1998; Rauskolb and Irvine 1999; Bishop et al. 1999).

No evidence for a role of *Sp8/9* in appendage specification

There is some evidence in *Drosophila* that *D-Sp1* has a role in leg specification by activating *Dll* expression in the embryonic leg primordia together with Wg and Dpp signaling and *Dll* autoregulation (Estella et al. 2003; McKay et al. 2009). However, this evidence is not conclusive because the experiments were not able to discriminate between the effect of *D-Sp1* and the effect of the neighboring gene *btd*. Beermann et al. (2004) suggest that *Dll* is a target gene of *Sp8/9* in *Tribolium*, although *Dll* is still expressed in the legs after *Sp8/9* RNAi. In summary, the evidence for a role of *Sp8/9* in leg specification and *Dll* activation in the insects studied so far is inconclusive.

Our results with *Sp8/9* RNAi in *Oncopeltus* provide no evidence for a role of *Sp8/9* in appendage specification or *Dll* activation. All appendages are present in *Sp8/9* RNAi animals and the legs consist of all podomeres including the distal claws. This suggests that the specification of the appendages and their overall proximal–distal patterning is not disrupted. This is further evidenced by the expression of the proximal–distal marker genes *Dll*, *dac*, and *hth* in *Sp8/9* RNAi animals. Although the legs in these animals are much shorter than in the wild type, the expression of *Dll*, *dac*, and *hth* is proportionally identical to the wild-type expression, indicating that proximal, medial, and distal fates are present. These data also suggest that *Dll*, *dac*, and *hth* expression is not dependent on activation by *Sp8/9*. We note, however, that we found very weak residual expression of *Sp8/9* after *Sp8/9* RNAi (data not shown), and thus we cannot exclude the possibility that the phenotypes we obtained do not represent the loss-of-function (null) phenotype.

Acknowledgments NDS would like to thank Paul Z. Liu for introducing her to *Oncopeltus* work and Nipam H. Patel for support and advice. This work has been supported by an Education Abroad Program Grant and a Fulbright Travel Grant (both to NDS), the European Community's Marie Curie Research Training Network ZOONET under contract MRTN-CT-2004-005624 (to EAW) and by a DFG Emmy Noether Program grant PR 1109/1-1 (to NMP).

References

Angelini DR, Kaufman TC (2004) Functional analyses in the hemipteran *Oncopeltus fasciatus* reveal conserved and derived aspects of appendage patterning in insects. Dev Biol 271:306–321

Beermann A, Aranda M, Schröder R (2004) The Sp8 zinc-finger transcription factor is involved in allometric growth of the limbs in the beetle *Tribolium castaneum*. Development 131:733–742

Bell SM, Schreiner CM, Waclaw RR, Campbell K, Potter SS, Scott WJ (2003) *Sp8* is crucial for limb outgrowth and neuropore closure. PNAS 100:12195–12200

Bishop SA, Klein T, Martinez Arias A, Couso JP (1999) Composite signalling from *Serrate* and *Delta* establishes leg segments in *Drosophila* through *Notch*. Development 126:2993–3003

de Celis JF, Tyler DM, de Celis J, Bray SJ (1998) Notch signalling mediates segmentation of the *Drosophila* leg. Development 125:4617–4626

Estella C, Rieckhof G, Calleja M, Morata G (2003) The role of *buttonhead* and *Sp1* in the development of the ventral imaginal discs of *Drosophila*. Development 130:5929–5941

Hughes CL, Kaufman TC (2000) RNAi analysis of *Deformed*, *proboscipedia* and *Sex combs reduced* in the milkweed bug *Oncopeltus fasciatus*: novel Hox genes in the Hemipteran head. Development 127:3683–3694

Liu PZ, Kaufman TC (2004a) *hunchback* is required for suppression of abdominal identity, and for proper germband growth and segmentation in the intermediate germband insect *Oncopeltus fasciatus*. Development 131:1515–1527

Liu PZ, Kaufman TC (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 PZ, Kaufman TC (2005) *even-skipped* is not a pair-rule gene but has segmental and gap-like functions in *Oncopeltus fasciatus*, an intermediate germband insect. Development 132:2081–2092

Kawakami Y, Esteban CR, Matsui T, Rodriguez-Leon J, Kato S, Izpisua Belmonte JC (2004) *Sp8* and *Sp9*, two closely related *buttonhead*-like transcription factors, regulate *Fgf8* expression and limb outgrowth in vertebrate embryos. Development 131:4763–4774

McKay DJ, Estella C, Mann RS (2009) The origins of the *Drosophila* leg revealed by the cis-regulatory architecture of the Distalless gene. Development 136:61–71

Milán M, Cohen SM (2000) Subdividing cell populations in the developing limbs of *Drosophila*: do wing veins and leg segments define units of growth control? Dev Biol 217:1–9

Prpic NM, Janssen R, Damen WGM, Tautz D (2005) Evolution of dorsal–ventral axis formation in arthropod appendages: *H15* and *optomotor-blind/bifid*-type T-box genes in the millipede *Glomeris marginata* (Myriapoda: Diplopoda). Evol Dev 7:51–57

Prpic NM, Damen WGM (2009) *Notch*-mediated segmentation of the appendages is a molecular phylotypic trait of the arthropods. Dev Biol 326:262–271

Rauskolb C, Irvine KD (1999) *Notch*-mediated segmentation and growth control of the *Drosophila* leg. Dev Biol 210:339–350

Schinko JB, Kreuzer N, Offen N, Posnien N, Wimmer EA, Bucher G (2008) Divergent functions of *orthodenticle*, *empty spiracles* and *buttonhead* in early head patterning of the beetle *Tribolium castaneum* (Coleoptera). Dev Biol 317:600–613

Schöck F, Purnell BA, Wimmer EA, Jäckle H (1999) Common and diverged functions of the *Drosophila* gene pair *D-Sp*1 and *buttonhead*. Mech Dev 89:125–132

Strimmer K, von Haeseler A (1996) Quartet puzzling: a quartet maximum likelihood method for reconstructing tree topologies. Mol Biol Evol 13:964–969

Treichel D, Schöck F, Jäckle H, Gruss P, Mansouri A (2003) *mBtd* is required to maintain signalling during murine limb development. Genes Dev 17:2630–2635

Wimmer EA, Frommer G, Purnell BA, Jäckle H (1996) *buttonhead* and *D-Sp1*: a novel *Drosophila* gene pair. Mech Dev 59:53–62

Zhao C, Meng A (2005) Sp1-like transcription factors are regulators of embryonic development in vertebrates. Develop Growth Differ 47:201–211

3.4 Expression pattern of three Sp homologous genes in the cnidarian *Nematostella vectensis.*

To collect additional data about the spatio-temporal expression pattern and thus further elucidate the potential ancestral function of the three different Sp genes, these were analyzed in the diploblastic sea anemone *N. vectensis*. *Sp1-4*, *Sp5/btd* and *Sp6-9* could be isolated from a cDNA pool and via in situ hybridization the spatio-temporal expression pattern was determined.

Nina D. Schäper and Ernst A. Wimmer

Status: work in progress

Author contribution to the practical work:
Nina Schäper did all the experiments.

Expression pattern of three Sp homologous genes in the cnidarian *Nematostella vectensis.*

Nina D. Schaeper and Ernst A. Wimmer*

Georg-August-Universität, Johann-Friedrich-Blumenbach-Institut für Zoologie und Anthropologie, Abteilung Entwicklungsbiologie, GZMB, Ernst-Caspari-Haus, Justus-von-Liebig-Weg 11, 37077 Göttingen, Germany

(*) Author for correspondence:
Ernst A. Wimmer
Georg-August-Universität, Johann-Friedrich-Blumenbach-Institut für Zoologie und Anthropologie, Abteilung Entwicklungsbiologie, GZMB, Ernst-Caspari-Haus, Justus-von-Liebig-Weg 11, 37077 Göttingen, Germany

Tel.: +49 551 39 22889
FAX: +49 551 39 5416
e-mail: ewimmer@gwdg.de

Introduction

Members of the Sp family of transcription factors are present in different numbers in vertebrates and invertebrates (e.g. Suske et al., 2005). Recent studies have shown that members of the Sp family of transcription factors are already present in the cnidarian *Nematostella vectensis* (Putnam et al., 2007) and a set of three Sp genes is ancestral in the metazoans (Schaeper et al., in prep. (see 3.1)). Gene expression data and functional data for these zinc finger transcription factors are available from several bilaterian species, but data from the cnidaria are missing so far. Investigations in this early branching group of the metazoa would help to understand the ancestral function of these transcription factors.

The body plan of cnidarians is relatively simple. The adult organism consists of two epithelia (ectoderm and endoderm). These form a tube with a mouth opening at the oral end and a foot structure at the aboral end. A ring of tentacles emerge around the oral opening which are used for feeding. Cnidarians already possess a number of different specified cell types. Data from the cnidarian hydroid *Hydra* showed that interstitial stem cells (i-cells) can give rise to two major cell lines: nerve cells and cnidocytes (nematocytes) to maintain cell complexity (Sproull and David, 1979; Siebert et al., 2007).

The anthozoan cnidarian *Nematostella vectensis* has become a suitable organism for not only ecological, but also developmental studies. Its genome is sequenced and several methods like in situ hybridization, morpholino-based RNA knockdown, as well as transgenesis are established (e.g. Putnam et al., 2007; Rentsch et al., 2008).

In order to begin an investigation into the possible ancestral function of the Sp gene triplet *Sp1-4, Sp5/btd* and *Sp6-9*, homologs of these genes were isolated from of the sea anemone *N. vectensis* and their embryonic expression pattern was analyzed.

Materials and Methods

Gene cloning

Embryos of all developmental stages were used for mRNA isolation using the MicroPoly(A)Purist kit (Ambion). Double-stranded (ds) cDNA template synthesis was performed using the SMART PCR cDNA Synthesis kit (Clontech). Gene specific primers were made based on sequences of *Nv*_Sp1-4 [GenBank: XP_001635004], *Nv*_Sp5/Btd [GenBank: XP_001635002], *Nv*_Sp6-9 [GenBank: XP_001634948].

The primer pair NvSp1-4_fw CCC TCT CCT CTA GCT CTG CTG GCT GC and NvSp1-4_re CAT CCG ATC GCG TAA ACC GTT TCC CAC was used to amplify a 1132bp fragment of *Nv_Sp1-4*, NvSp5/btd_fw GGC GCT CTT AGC GGC AAC CTG CAG CC and NvSp5/btd_re

CTC CGT GGA CAT GTC GGG ATC ATC G for a 1060 bp fragment of *Nv_Sp5/btd*, and the primers NvSp6-9_fw GGG ATG CTA GCT GCA ACT TGT AGT CGC and NvSp6-9_re CAC TCC GTT GAT CCT CCG AGT TCG TGG for amplification of *Nv_Sp6-9* (1139 bp). All fragments were cloned into the pCRII vector (Invitrogen).

In situ hybridization

The isolated *N. vectensis* Sp homologs were used for digoxigenin labeled RNA probe synthesis (Roche). Embryo and planula collection, fixation and in situ staining were performed as described in Rentsch et al. (2008).

Results

It was possible to isolate all three Sp family members identified in the sequenced genome from a *N. vectensis* cDNA pool. This shows that *Nv Sp1-4*, *Nv Sp5/btd* and *Nv Sp6-9* are indeed expressed in this cnidarian. The next step was the analysis of the embryonic expression pattern of *Nv Sp1-4*, *Nv Sp5/btd* and *Nv Sp6-9*.

Generally, all three Sp genes are expressed in a similar way with only small differences. *Nv Sp1-4* is expressed in single cells of the endo- and ectoderm throughout the embryo in gastrula (not shown) and planula stages (Fig. 1 A, D). In the oral region, a narrow belt has a reduce number of *Sp1-4* expressing cells (Fig. 1 A, arrow). Also *Nv Sp5/btd* is expressed in single cells throughout the embryo, but only in the ectoderm and with an additional accumulation of stained cells in a central stripe (Fig. 1 E, arrow). There is also expression in a small number of oral entodermal cells. *Nv Sp6-9* is expressed in single ectodermal cells throughout the embryo in gastrula and planula stages as well as in the budding tentacles of metamorphosing planulas (Fig. 1 C, F).

A closer examination with higher magnification suggests that these Sp gene expressing cells are interstitial cells and nematocytes. These cells can be identified by their shape already at these early stages of development.

Discussion

All three Sp genes could be isolated from a cDNA pool, containing embryos from all developmental stages. Preliminary in situ stainings for these three genes in *N. vectensis* embryos and planulas were performed in the lab of Prof. Ulrich Technau (Dept. for Molecular Evolution and Development, University of Vienna, Austria) and showed similar expression patterns. *Nv Sp1-4*, *Nv Sp5/btd* and *Nv Sp6-9* are expressed in blastula and planula stages in

single scattered cells. These cells are apparently i-cells and nematocytes. Various studies for stem cell growth and differentiation were done in the hydrozoan *Hydra* (Sproull and David, 1997), but i-cells have not been analyzed in *N. vectensis* so far. For the question, whether *Nv Sp1-4*, *Nv Sp5/btd* and *Nv Sp6-9* are indeed expressed in i-cells, markers for this cell type have to be developed for *N. vectensis*.

Furthermore, *Nv Sp6-9* is also expressed in budding tentacles of metamorphosing planulas. In cnidarians, nematocytes are located in high numbers in the tentacles due to feeding behaviour. It might be the case that the expression of *Nv Sp6-9* in the budding tentacles is located in these nematocyte cells. Alternatively, the expression of Sp6-9 in the tentacles might indicate a role of this gene in the development of these outgrowths. Intriguingly, Sp6-9 homologs are also expressed in the appendages of arthropods and vertebrates and are required for their outgrowth (e.g. Beermann et al., 2004; Kawakami et al., 2004) (see also 3.1, 3.2. 3.3). A similar case is known for the *Dll/Dlx* gene. This gene has been studied in several species (e.g. Panganiban et al., 1997). *Dll/Dlx* expression along a proximodistal axis has been observed in onychophoran lobopodia, polychaete annelid parapodia, ascidian ampullae, and even echinoderm tube feet. Therefore it seems that *Dll/Dlx* is used by a variety of organisms for body wall outgrowths.

The expression of *Nv Sp6-9* in budding tentacles also suggests an ancestral function for *Sp6-9* in body wall outgrowths as different as cnidarian tentacles, vertebrate limbs and arthropod appendages. More experiments have to be done for more definite conclusions.

Due to some problems with *N. vectensis* husbandry, perturbed spawning and resulting embryonic developmental problems we were not sure about the validity of these data. Embryos used for a second trial were also impaired in their development and therefore artefact staining cannot be excluded. These preliminary data therefore have to be confirmed. Furthermore, more different developmental stages have to be stained to get a broader overview of the expression pattern of Sp genes. In addition, morpholino-based knockdown experiments could provide insight into the ancestral function of *Sp1-4*, *Sp5/btd* and *Sp6-9*.

Acknowledgments

NDS want to thank Ullrich Technau and his group for their hospitality during her stay in Vienna, and the opportunity to perform the *Nematostella* in situ hybridization in his group in Vienna. We also thank Stefanie Pontasch (group of Prof. Dr. Bert Hobmayer, Institute for Zoology, University Innsbruck) for providing us with *N. vectensis* embryos for RNA isolation.

References

Beermann A, Aranda M, Schröder R. The *Sp8* zinc-finger transcription factor is involved in allometric growth of the limbs in the beetle *Tribolium castaneum*. *Development* 2004, 131:733-742.

Kawakami Y, Esteban CR, Matsui T, Rodríguez-León J, Kato S, Izpisúa Belmonte JC. *Sp8* and *Sp9*, two closely related *buttonhead*-like transcription factors, regulate *Fgf8* expression and limb outgrowth in vertebrate embryos. *Development*. 2004, 131:4763-4774.

Panganiban G, Irvine SM, Lowe C, Roehl H, Corley LS, Sherbon B, Grenier JK, Fallon JF, Kimble J, Walker M, Wray GA, Swalla BJ, Martindale MQ, Carroll SB. The origin and evolution of animal appendages. *PNAS* 1997, 94:5162-5166.

Putnam NH, Srivastava M, Hellsten U, Dirks B, Chapman J, Salamov A, Terry A, Shapiro H, Lindquist E, Kapitonov VV, Jurka J, Genikhovich G, Grigoriev IV, Lucas SM, Steele RE, Finnerty JR, Technau U, Martindale MQ, Rokhsar DS. Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. *Science* 2007, 317:86-94.

Rentzsch F, Fritzenwanker JH, Scholz CB, Technau U. FGF signalling controls formation of the apical sensory organ in the cnidarian *Nematostella vectensis. Development* 2008, 135:1761-1769.

Schaeper ND, Prpic NM, Wimmer EA. A clustered set of three Sp-family genes is ancestral in the Metazoa: evidence from sequence analysis, protein domain structure, chromosomal location and developmental mRNA expression. (in preparation)

Siebert S, Anton-Erxleben F., Bosch TCG. Cell type complexity in the basal metazoan *Hydra* is maintained by both stem cell based mechanisms and transdifferentiation. *Dev Biol* 2007, 313:13-24.

Sproull F, David CN. Stem cell growth and differentiation in *Hydra attenuata*. *J Cell Sci* 1979, 38:155-169.

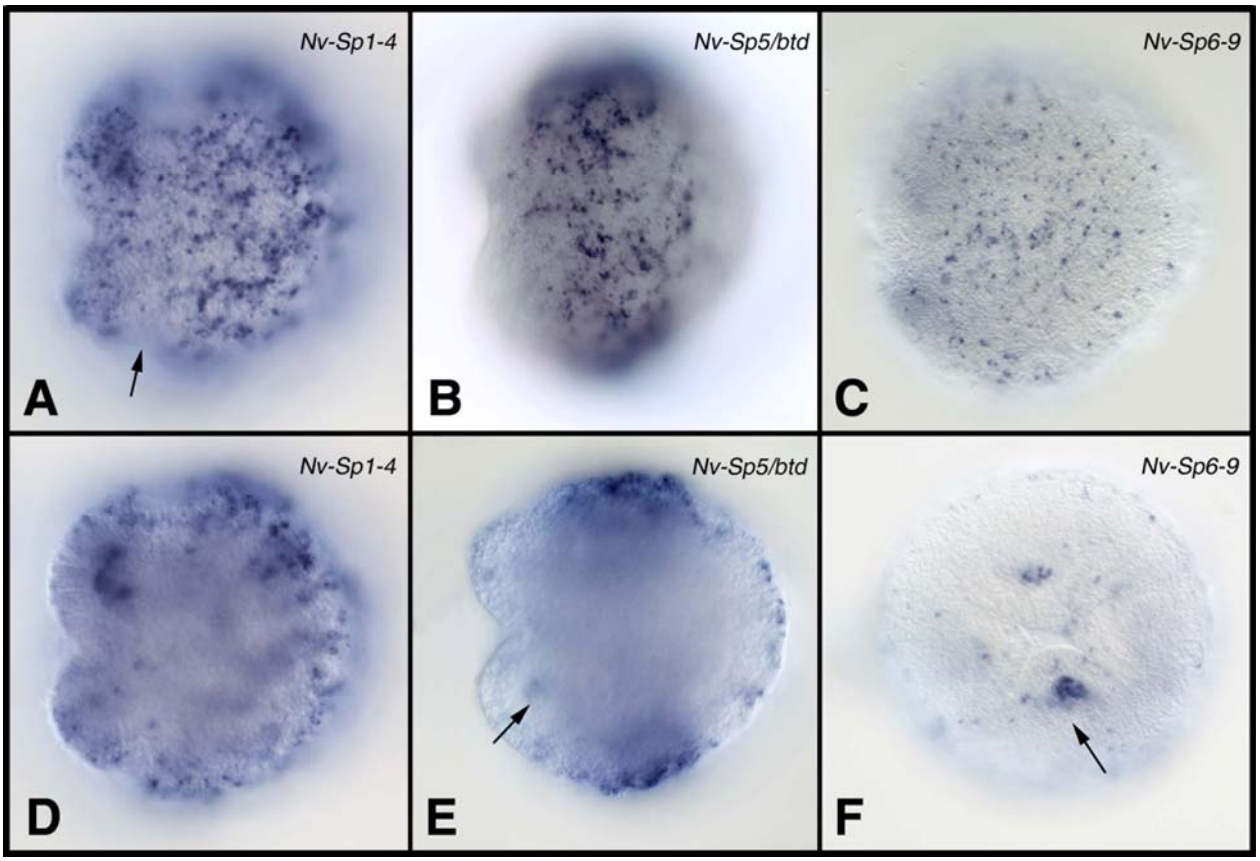


Figure 1. Expression pattern of *Sp1-4*, *Sp5/btd* and *Sp6-9* in *Nematostella vectensis*. Early planula stages orientated with the oral part to the left (A-E) and the oral part in top view (F). (A/D) Expression of *Nv Sp1-4*. A and D show the same specimen in different focal planes. Arrow marks a region with reduced expression (A). (B/E) Expression of *Nv Sp5/btd*. B and E show the same specimen in different focal planes. Expression is located in a central stripe like pattern (B). Arrow in E marks potential entodermal expression. (C/F) Expression of *Nv Sp6-9*. C and F show the same specimen in different orientations. Expression of *Nv Sp6-9* is distributed over the animal (C) and also accumulated in the budding tentacles (F, arrow).

3.5 Evolutionary plasticity of *collier* function in head development in diverse arthropods.

The purpose of this work was to reveal similarities and also differences regarding the expression and function of *collier* – a traget gene of *btd* in *D. melanogaster* – in different arthropods.
Therefore, *col* homologues were isolated from the holometabolous insect *Tribolium castaneum*, the hemimetabolous insect *Oncopeltus fasciatus*, the amphipod crustacean *Parhyale hawaiensis* as well as from the chelicerate *Achaearanea tepidariorum*. To compare *col* among these arthropods, the spatio-temporal expression pattern was analyzed and parental RNAi experiments with *col* in *T. castaneum* and *O. fasciatus* were carried out.

Nina D. Schäper, Matthias Pechmann, Wim G.M. Damen, Nikola-Michael Prpic, Ernst A. Wimmer

Status: submitted to “Developmental Biology” (on November 26, 2009)

Author contribution to the practical work:
Nina Schäper isolated *collier* from *Tribolium castaneum*, *Oncopeltus fasciatus*, and *Parhyale hawaiensis*, did in situ hybridizations for these three organisms and performed RNAi for *T. castaneum* and *O. fasciatus*.
Matthias Pechmann isolated *collier* from *Achaearanea tepidariorum* and performed in situ hybridization in this species.

From: ees.developmentalbiology.0.56e9f.23b581fc@eesmail.elsevier.com
[mailto:ees.developmentalbiology.0.56e9f.23b581fc@eesmail.elsevier.com]
On Behalf Of Developmental Biology
Sent: Wednesday, December 02, 2009 7:54 PM
To: Wimmer, Ernst

Subject: Manuscript Number Assigned: DBIO-09-768

Ms. No.: DBIO-09-768

Title: Evolutionary plasticity of collier function in head development of diverse arthropods

Corresponding Author: Dr. Ernst A. Wimmer

Authors: Nina D Schaeper; Matthias Pechmann; Wim G Damen; Nikola-Michael Prpic;

Dear Dr. Wimmer,

Your submission, referenced above, has been assigned the following manuscript number: DBIO-09-768

You will be able to check on the progress of your paper by logging on to the Elsevier Editorial System as an author:
http://ees.elsevier.com/developmentalbiology/

Thank you for submitting your work to Developmental Biology.

Kind regards,

Sandra Tokashiki
Journal Manager
Developmental Biology

Elsevier Science
525 B St., Ste. 1900
San Diego, CA 92101-4495 USA
tel: 619-699-6351
fax: 619-699-6211
db@elsevier.com

Manuscript for Developmental Biology
Section "Evolution of Developmental Control Mechanisms"

Evolutionary plasticity of *collier* function in head development of diverse arthropods.

Nina D. Schaeper[1], Matthias Pechmann[1], Wim G.M. Damen[2], Nikola-Michael Prpic[1], Ernst A. Wimmer[1]*

1 Georg-August-Universität, Johann-Friedrich-Blumenbach-Institut für Zoologie und Anthropologie, Abteilung Entwicklungsbiologie, GZMB, Ernst-Caspari-Haus, Justus-von-Liebig-Weg 11, 37077 Göttingen, Germany

2 Friedrich-Schiller-University, Department of Genetics, Philosophenweg 12, 07743 Jena, Germany

(*) Author for correspondence:
Ernst A. Wimmer
Georg-August-Universität, Johann-Friedrich-Blumenbach-Institut für Zoologie und Anthropologie, Abteilung Entwicklungsbiologie, GZMB, Ernst-Caspari-Haus, Justus-von-Liebig-Weg 11, 37077 Göttingen, Germany

Tel.: +49 551 39 22889
FAX: +49 551 39 5416
e-mail: ewimmer@gwdg.de

Number of words: 8310 (in total)
Number of figures: 10
Number of tables: none
Supplementary Information: 2 tables

Abstract

The insect intercalary segment represents a small and appendage-less head segment that is homologous to the second antennal segment of Crustacea and the pedipalpal segment in Chelicerata. Despite their divergence in external appendages, these homologous segments give rise to the tritocerebrum and are therefore in general referred to as ‘tritocerebral segment’ irrespective of the arthropod clade. In *Drosophila*, the gene *collier* (*col*) has an important role for the formation of the intercalary segment. Here we show that the loss of *col* in the beetle *Tribolium castaneum* leads to increased cell death in the intercalary segment and a larval phenotype with the animals holding their heads abnormally bent. In the milkweed bug *Oncopeltus fasciatus*, the loss of *col* function has a more severe effect in lacking the intercalary segment and also affecting the adjacent mandibular and antennal segments. By contrast, *col* is not expressed early in the second antennal segment in the crustacean *Parhyale hawaiensis* or in the pedipalpal segment of the spider *Achaearanea tepidariorum*. This suggests that the early expression of *col* in a stripe and its role in tritocerebral segment development is an insect novelty that correlates with the specific appendage-less morphology of the intercalary segment in insects.

Key words: arthropod head evolution, intercalary segment, *collier*, *Tribolium castaneum*, *Oncopeltus fasciatus*

Introduction

The segmental composition of the arthropod head is highly conserved (Damen et al., 1998; Telford and Thomas, 1998). The specific morphology of the head segments, however, can be quite diverse. A prime example for this diversity is the tritocerebral segment. Originally, the tritocerebral segment is a proper segment equipped with a pair of appendages, and this ancestral condition is conserved in several extant arthropod groups (e.g. the pedipalpal segment in spiders and the second-antennal segment in crustaceans). In the insects, however, the tritocerebral segment is smaller than the remaining head segments and lacks appendages completely. This reduced tritocerebral segment of the insect head is called intercalary segment. The developmental genetic mechanisms leading to the appendage-less condition of the intercalary segment are not known, but a small number of genes have been isolated in *D. melanogaster* that are expressed in the intercalary segment and, when mutated, affect its morphology. One of these genes is *collier* (*col*; also known as *knot* (*kn*)).

In *D. melanogaster*, *col* is expressed early in parasegment 0 (PS0) and is crucial for the formation of head structures derived from this area, namely the intercalary and mandibular segment (Crozatier et al., 1996; Crozatier et al., 1999; Seecoomar et al., 2000). The *col* gene is also required for the development of the central and peripheral nervous system (e. g. Crozatier et al., 1996; Baumgardt et al., 2007; Crozatier and Vincent, 2008), is involved in wing patterning (Nestoras et al., 1997; Vervoort et al., 1999; Mohler et al., 2000; Crozatier et al., 2002; Hersh and Carroll, 2005), and functions in somatic muscle formation (Crozatier and Vincent, 1999; Dubois et al., 2007; Crozatier and Vincent, 2008).

Genes related to *col* are also known from other animals. Members of this COE transcription factor family (named after the factors Collier/Olf/EBF) are evolutionarily conserved proteins and are present throughout the Metazoa. One single gene of the COE family can be found in invertebrates including also the cnidarian *Nematostella vectensis*. Vertebrates, however, possess several paralogous genes of the COE family. Teleost fish, amphibians, and birds have three, mammals have four members of the COE family (reviewed in Dubois and Vincent (2001) and Liberg et al. (2002)).

The COE family is involved in a variety of processes during embryonic development. The *Caenorhabditis elegans* COE family member *unc-3* plays a role in ventral nerve cord development, chemosensory neuron differentiation and axon guidance (Prasad et al., 1998). Similarly, Pang et al. (2003) proposed a function in chemoreception for the homologous *coe* gene from *Nematostella vectensis*, which indicates an ancestral role of COE factors in chemoreception. This is also consistent with the expression of the murine Olf/EBF genes in the

developing nervous system, in immature olfactory neuronal precursors and mature neurons of the adult olfactory epithelium (reviewed in Dubois and Vincent (2001)). It was shown that mouse *Olf-1/EBF-1* is necessary for B-cell differentiation (reviewed in Liberg et al. (2002)). Moreover, the EBF genes also play a role in limb development of mouse and chicken (Mella et al., 2004).

Here we report the isolation of COE homologs (designated below as *col*) of four different arthropod species, the red flour beetle *Tribolium castaneum*, the milkweed bug *Oncopeltus fasciatus*, the amphipod crustacean *Parhyale hawaiensis* and the common house spider *Achaearanea tepidariorum*. All four arthropod species show a late expression pattern of *col* in the nervous system, but only in *T. castaneum* and *O. fasciatus* an early expression in the head is seen. Loss-of-function studies using parental RNA interference (pRNAi) in *T. castaneum* and *O. fasciatus* reveal a role for *col* in the patterning and maintenance of the intercalary segment. The highly conserved neural expression in all arthropods suggests that this represents the ancestral function of *col* in the Arthropoda, whereas the early expression of *col* in the head and its role in intercalary segment development seem to be insect-specific. We propose that the evolution of this novel functional domain of *col* in the tritocerebral segment of the insect ancestor correlates with the origin of the specific appendage-less morphology of this – intercalary – segment in insects.

Materials and Methods

Arthropod husbandry and embryology

The *T. castaneum* cultures were kept as described in Berghammer et al. (1999). Milkweed bugs *O. fasciatus* were cultured according to Hughes and Kaufman (2000). Embryos of all stages were fixed as described previously (Liu and Kaufman, 2004a). Dissections of *O. fasciatus* embryos before in situ staining were performed under a fluorescence stereomicroscope using SYTOX Green nucleic acid stain (Liu and Kaufman, 2004b). The *P. hawaiensis* cultures were kept at 26°C in shallow plastic boxes filled with artificial seawater (30g/l of synthetic sea salt) and a thin layer of crushed coral substrate. Membrane pumps constantly ventilated the water. The amphipods were fed with dry fish flakes twice a week. Water was changed every two weeks. Gravid *P. hawaiensis* females were anaesthesized with clove oil (10µl per 50ml seawater) and embryos were collected out of the brood pouch with forceps. *P. hawaiensis* embryos were staged, dissected and fixed as described before (Browne et al., 2005; Browne et al., 2006). Embryos of the common house spider *A. tepidariorum* were treated as previously described (Akiyama-Oda and Oda, 2003).

Isolation of *collier* homologs from different arthropods

T. castaneum embryos aged 0 to 72h, *O. fasciatus* embryos aged 0 to 96h, *P. hawaiensis* embryos of all described stages (Browne et al., 2005), and *A. tepidariorum* embryos (stages 0 to 10) were used for mRNA isolation using the MicroPoly(A)Purist kit (Ambion; Applied Biosystems, Darmstadt, Germany). Double-stranded (ds) cDNA was synthesized using the SMART PCR cDNA Synthesis kit (Clontech; BD Biosciences, Heidelberg, Germany). The *P. hawaiensis* and *A. tepidariorum* mRNA also served as template for RACE template synthesis (SMART RACE cDNA Amplification Kit, Clontech). A 1130bp fragment of *Tc col* (LOC662502, accession number XM_968593) was isolated using the gene specific primers Tccol_fw (GGG CGC ACT TCG AGA AAC AGC CTC CC) and Tccol_re (GCT GAC AGC CAG TTG CCC CGT ATAC G). Two non-overlapping *Tc col* fragments were isolated from this 1130bp template with the primer pair Tccol_fw1 (GGC GCA CTT CGA GAA ACA GCC TC) and Tccol_re1 (CTT CTT GTC GCA GCA ACG GCT GCA C) (fragment length: 353bp), as well as Tccol_fw2 (GCA ATC CCA GAG ACA TGA GAC GG) and Tccol_re2 (GCT GAC AGC CAG TTG CCC CGT ATA CG) (fragment length: 674bp). The degenerate primer pair fw_AHFEKQP (GCI CAY TTY GAR AAR CAR CC) and re_DNMFVHNN (TTR TTR TGI ACR AAC ATR TTR TC) was used for the isolation of *Of col*. The PCR resulted in a 516bp fragment. Two non-overlapping fragments of *Of col* were amplified using the gene specific primer pair Ofcol1_fw3 (CCC TCG AAT CTG AGG AAG TCC AAC) and Ofcol1_re3 (CCT CGC AG ATG ATT GCC TGT TTC G) (fragment length: 253bp) as well as Ofcol2_fw4 (CCG AGA TGT GCA GAG TAC TTC TGA CC) and Ofcol2_re4 (GTC GGA GAT GGC GAG CAG AGG ACC) (fragment length: 248bp). A 432 bp fragment of *Ph col* was obtained using the degenerate primers fw_AGQPI/VEIE (GCI GGI CAR CCI RTI GAR ATH GA) and re_DNMFVHN (TTR TTR TGI ACR AAC ATR TTR TC). Additional sequence information was obtained by RACE PCR. The complete sequence of *Ph col* comprises 1883bp, including 166bp 5'-UTR, 1506bp ORF (501 amino acids) and 211bp 3'-UTR. A 393bp fragment of *At col* was obtained using the primers fw_AGQPI/VEIE and re_AGNPRDM (GAT RTC NCK NGG RTT NCC NGC). 5' RACE PCR resulted in a 839bp fragment (240bp 5'-UTR, 599bp ORF). The orthology of all fragments with *D. melanogaster col* was confirmed by phylogenetic analysis as described before (Prpic et al., 2005). All PCR fragments were subcloned into the pCRII vector (Invitrogen; Karsruhe, Germany). The isolated arthropod *collier* sequences were submitted to the EMBL database (Accession numbers: *Of col* FN557065, *Ph col* FN557064, *At col* FN557066).

In situ Hybridization and TUNEL detection

The 1130bp fragment of *Tc col*, the 516bp fragment of *Of col,* the 1883bp fragment of *P. hawaiensis col* and 839bp fragment of *A. tepidariorum col* were used for digoxygenin-labeled RNA probe synthesis. For *T. castaneum* double stainings, fluorescein-labeled probes were used

for INT/BCIP (Roche; Mannheim, Germany) stainings. In situ hybridization was performed as described in Wohlfrom et al. (2006) for *T. castaneum*, in Liu and Kaufman (2004a) for *O. fasciatus*, in Browne et al. (2006) for *P. hawaiensis*, and in Prpic et al. (2008) for *A. tepidariorum*. Detection of cell death via the TUNEL method was performed with minor modifications as described previously for the spider *Cupiennius salei* (Prpic and Damen, 2005).

Parental RNA interference

Double-stranded RNA (dsRNA) for *O. fasciatus* and *T. castaneum* parental RNA interference (RNAi) was synthesized using the MEGAscript T7 Kit (Ambion; Applied Biosystems, Darmstadt, Germany) and resuspended in 1x injection buffer (1,4mM NaCl, 0,07mM Na_2HPO_4, 0,03mM KH_2PO_4, 4mM KCl) at a concentration of 4 µg/µl. DsRNA of the 516bp fragment of *Of col*, as well as the two non-overlapping fragments of 253bp und 248bp were injected into female virgin bugs as described previously (Liu and Kaufman, 2004a). For *Tc col*, dsRNA of the 1130bp fragment, as well as the two non-overlapping fragments of 353bp and 674bp were injected into adult beetles and pupae. Virgin adult female red flour beetles were immobilized on ice, held carefully with forceps and injected laterally into the ovipositor. Injected beetles were put on flour and kept at 32°C. Males were added the next day, embryos were collected every 24h. Injections into female pupae were performed as described in Bucher et al. (2002). Hatched larval progeny were embedded into a Hoyer's medium/lactic acid mixture (1:1). Injection of 1x injection buffer served as a negative control for *O. fasciatus* and *T. castaneum* RNAi. Additional independent control injections for *O. fasciatus* were performed as described previously (Schaeper et al., 2009).

Results

Expression of *collier* in the flour beetle *Tribolium castaneum*

In order to investigate the conservation of the function of *col* in insects other than *D. melanogaster*, we have first cloned the *col* homologue of the red flour beetle *T. castaneum* and determined its expression pattern. A strong stripe of *Tc col* expression is already detected in very early germ band stages (Fig. 1A, B). This expression domain is located in the anterior head, but clearly behind the ocular region as indicated by the expression of the ocular *Tc wingless* (*Tc wg*) stripe. Later when the head lobes begin to form, the *Tc col* stripe can be seen at the junction between the head lobes and the sack-shaped rest of the germ band (Fig. 1C). When primordial body segments become indicated by the segmental expression of *Tc wg*, it is evident that the *Tc col* stripe lies between the ocular and the mandibular *Tc wg* stripe (Fig. 1D). A strong internal (probably mesodermal) expression extends from there along the ventral midline back almost to the maxillary segment (Fig. 1D, arrow). Slightly later this internal expression retracts again and

becomes restricted to the mandibular segment (Fig. 1E). The *Tc col* stripe now abuts the mandibular *Tc wg* stripe, and is thus located in the tissue of the anterior mandibular segment and tissue anterior to it (Fig. 1E). In the further course of development the *Tc col* stripe undergoes refinement (Fig. 1F-G), spans the furrow anterior to the mandibular segment, and thus is located in tissue that will contribute to the intercalary segment. The anterior and posterior borders of the *Tc col* stripe become fuzzy when the embryo approaches the full germ band extension stage (Fig. 1H) and at the end of germ band extension the *Tc col* stripe is almost completely restricted to the furrow anterior to the mandibular segment (Fig. 1I). During germ band retraction the *Tc col* stripe in the head becomes even more restricted and is gradually replaced by an emerging expression in the central nervous system of the head (Fig. 1J-L). In early retraction stage embryos the stripe splits (Fig. 1J) and is now clearly located in presumptive intercalary tissue. Slightly later the split *Tc col* stripe gets weaker (Fig. 1K) and finally dissolves into the punctate pattern of the central nervous system of the intercalary segment (Fig. 1L).

In addition to the head stripe, *Tc col* also has other expression domains. In early germ band retraction embryos spots of *Tc col* expression appear laterally in the mesoderm of the thorax (Fig. 2A) and later similar spots also appear in the anterior abdominal segments (Fig. 2B) and subsequently in all abdominal segments (Fig. 2C, D). As soon as the dorsal extension of the germ band has started, these lateral spots move dorsally with the growing dorsal tissue and simultaneously become weaker (Fig. 2E). In embryos approaching dorsal closure the dorso-lateral spots have almost completely faded (Fig. 2F). In the central nervous system the expression pattern of *Tc col* increases in complexity during development. Initial neural expression is detected in the head in the intercalary and antennal segment (Fig. 1J; Fig. 2A), but soon additional cell groups in the protocerebrum also express *Tc col*: first in the ocular region (Fig. 1K), later in the entire protocerebrum (Fig. 1L; Fig. 2B). Gradually, punctate expression is also detected in the gnathal segments (Fig. 2B), thoracic and anterior abdominal segments (Fig. 2C), and finally in all segments (Fig. 2D). This expression pattern increases in complexity mirroring the further development and compaction of the central nervous system (Fig. 2E) including brain formation (Fig. 2F).

Phenotypic effects of *Tc col* RNAi in *Tribolium castaneum*

The gene expression data from *D. melanogaster* and *T. castaneum* show that the two species share very similar *col* expression patterns. In order to see whether *col* also shares similar functions in the two species we have tested the function of *Tc col* in *T. castaneum* using RNAi (Supplementary Table 1). RNAi embryos do not show any obvious malformations (Fig. 3, Fig. 4). However, using the TUNEL method in germ band elongation and retraction stages we detected

significantly elevated cell death levels in the intercalary segment area in *Tc col* RNAi embryos compared to the wildtype. In late germ band elongation stages we detected a patch of dead cells in the area in front of the mandibular segment (compare Fig. 3A and B). In early germ band retraction stages elevated cell death levels were detected in a stripe spanning the area in front of the mandibular segment (compare Fig. 3C and D). This stripe of elevated cell death levels in front of the mandibular segment persists in mid germ band retraction stages (compare Fig. 3E and F).

In this area of elevated cell death levels, we also detected changes in gene expression of additional genes known to be expressed during head development (Fig. 4). The genes *Tc engrailed* (*Tc en*) and *Tc wg* are expressed in segmental stripes including small dots in the intercalary segment (Fig. 4A, C, C'). In *Tc col* RNAi embryos the *Tc en* and *Tc wg* pattern is normal, except that the dots in the intercalary segment are missing (Fig. 4B, D). The gene *Tc hedgehog* (*Tc hh*) is expressed in a similar fashion to *Tc en* also including spots in the intercalary segment (Fig. 4E, G). These intercalary spots are missing in the *Tc col* RNAi embryos (Fig. 4F, H). The gene *Tc cap´n´collar* (*Tc cnc*) is expressed strongly in the mandibular segment and in a ring around the stomodeum including the labrum (Fig. 4I). This pattern is only slightly altered in *Tc col* RNAi embryos, in which the separation between the stomodeal and mandibular expression domains is less clear and both domains touch each other (Fig. 4J). The Hox gene *Tc labial* (*Tc lab*) is expressed strongly in the intercalary segment (Fig. 4K). We could not detect any visible difference between the *Tc lab* expression in wildtype and *Tc col* RNAi embryos (Fig. 4L). This is similar to *D. melanogaster*, in which loss of *col* function does not have an effect on *lab* expression either (Crozatier et al., 1996; Seecoomar et al., 2000).

We have then studied the larval phenotype of *Tc col* RNAi animals. Recently, a map of cuticular bristle markers for the *T. castaneum* head has become available (Schinko et al., 2008). This bristle map has recently been used to identify defects in the intercalary segment (Posnien and Bucher, 2009), a segment that is difficult to identify in larvae on a morphological basis. However, even with these marker bristles, our *Tc col* RNAi cuticles were undistinguishable from wildtype cuticles and the marker bristles of the so-called "gena triplet" and "maxilla escort" (Schinko et al., 2008) are all present as in the wildtype (data not shown).

Although the cuticle preparations of wildtype and *Tc col* RNAi animals are identical, we observed a striking difference in living larvae. While wildtype larvae hold their body straight and their head up (Fig. 5A), the *Tc col* RNAi larvae assume a very peculiar body posture: they hold their head bent down such that their mouthparts almost thouch the ventral side of the thorax (Fig. 5B). These larvae are able to assume a normal body posture when teased with a brush, but soon after return to their bent posture. This peculiar bending of the head does not seem to be caused

by internal deformations of muscles, because we could not detect any obvious differences in larval muscle architecture in untreated and *Tc col* RNAi animals of the transgenic line Pig-19, a line that expresses EGFP in muscles (data not shown) (Lorenzen et al., 2003).

Expression of *collier* in the milkweed bug *Oncopeltus fasciatus*

Both *D. melanogaster* and *T. castaneum* are holometabolous insects. In order to see whether the role of *col* is also conserved in hemimetabolous insects we cloned *col* from the milkweed bug *O. fasciatus* and investigated its expression. During blastoderm stages, *Of col* is expressed in a stripe in the anterior portion of the embryo (Fig. 6A, B). When the germ band has started extending, this stripe of *Of col* expression is located at the junction of the trunk and the head lobes, which are relatively undeveloped at this stage of *O. fasciatus* development (Fig. 6C). Later during germ band elongation the *Of col* stripe narrows and is now located between the antennal and the mandibular segment (Fig. 6D). During the following elongation stages the furrows between the segments become more pronounced (Fig. 6E, F). The *Of col* stripe splits (Fig. 6E) and separates into two spots in the furrow between the mandibular and the intercalary segment (Fig. 6F). In embryos approaching the end of germ band elongation additional spots appear along the head segments, which are similar to the initial *Of col* spots with respect to their shape and relative segmental position, but are expressed at a weaker level (Fig. 6G, H). In embryos during germ band retraction, the *Of col* expression pattern consists of a complex punctate pattern in the central nervous system including the brain and additional spots along the edge of the abdomen presumably in mesodermal tissue (Fig. 6I).

Phenotypic effects of *col* RNAi in *Oncopeltus fasciatus*

We have then also tested the function of *Of col* in *O. fasciatus* using RNAi (Liu and Kaufman 2004a) (Supplementary Table 2). Embryos from females subjected to *Of col* RNAi treatment appear morphologically relatively normal when examined before dorsal closure. In the *Of col* RNAi embryos, the maxillary segment and appendages are morphologically normal, but the mandibular segment and its appendage buds are slightly reduced and in addition the distance between the mandibular and antennal appendages is shorter than in wildtype embryos at the same stage, indicating missing intercalary tissue (Fig. 7B, D, F, H, J).

In these areas we have also detected changes in gene expression. The Hox gene *Of lab* is in the wildtype expressed in the intercalary segment (Fig. 7A) (Angelini et al, 2005). In *Of col* RNAi embryos the *Of lab* gene is still expressed, but the expression area is severely reduced (Fig. 7B). The gene *Of proboscipedia* (*Of pb*) is normally expressed in the intercalary and mandibular segment (an additional domain is present in the distal part of the labial appendages) (Fig. 7C)

(Angelini et al, 2005). In *Of col* RNAi embryos the two spots of *Of pb* expression in the posterior mandibular segment are still present, but the expression in the anterior mandibular segment and in the intercalary segment is missing (Fig. 7D). The *Of Deformed* (*Of Dfd*) gene is normally expressed in the mandibular and maxillary segments including the appendages (Fig. 7E) (Angelini et al, 2005). The expression of *Of Dfd* in the *Of col* RNAi embryos is similar to the wildtype expression pattern, except that the total length of the *Of Dfd* domain is reduced owing to the reduction of the mandibular segment and its appendages (Fig. 7F). The gene *en* is expressed in the posterior portion of all segments (Fig. 7G) (Liu and Kaufman, 2004b). In the *Of col* RNAi embryos the expression is relatively normal, but the two intercalary *en* stripes are missing and the mandibular *en* stripe is smaller than in the wildtype (Fig. 7H).

In wildtype embryos after dorsal closure the mandibular appendages start to assume their stylet shaped morphology (Fig. 8A). In *Of col* RNAi embryos at the same stage the mandibular stylets are somewhat thicker and shorter than in the wildtype (Fig. 8B). An effect not expected on the basis of the *Of col* expression pattern was observed in embryos shortly before hatching. In the wildtype the antennae reach almost the posterior end of the embryo (Fig. 8C) and the mandibular stylets are already very thin pin-shaped structures (Fig. 8A). In *Of col* RNAi animals, however, the antennae are much shorter, often twisted in a corkscrew manner, and consist mainly of necrotic tissue as indicated by their conspicuous black color (Fig. 8D). In some specimens the antennae are necrotic, thickened and bent, but not strongly twisted and the mandibular stylets are normal (Fig. 9B). In other specimens the antennae are necrotic and severely twisted and the mandibular stylets are of normal shape, but necrotic as well (Fig. 9C). Finally, some specimens show severely shortened and bloated necrotic antennae and necrotic mandibular stylets that still have the morphology of earlier embryonic stages (Fig. 9D).

We have also tested for a selection of appendage genes since the downregulation of *Of col* via RNAi later also affects the development of the antennal appendages. The distal appendage marker gene *Of Distal-less* (*Of Dll*) is expressed in the distal part of all appendages except for the mandible (Angelini and Kaufman, 2004). In the antenna it is expressed in a medial ring and a distal spot (Fig. 7I). This expression pattern is unchanged in the antennae of *Of col* RNAi embryos even though the adjacent intercalary and mandibular segments are morphologically clearly affected (Fig. 7J). We have also tested the expression of the appendage marker genes *Of dachshund* (*Of dac*), *Of homothorax* (*Of hth*) (Angelini and Kaufman, 2004), and *Of Sp8/9* (Schaeper et al., 2009), and we also could not detect any differences in gene expression of these three genes in wildtype and *Of col* RNAi embryos (data not shown).
These results agree with the morphologically normal appearance of the antennae at these embryonic stages. Thus the necrosis observed at later stages does not seem to be caused by

patterning defects within the antennae, this rather appears to be a secondary effect of intercalary segment malformation triggered only late in embryogenesis.

Expression of *collier* in non-insect arthropods

The data from *D. melanogaster*, *T. castaneum* and *O. fasciatus* indicate that the role of *col* in head development is conserved in the insects. However, differences in the regulation of *lab* suggest a certain degree of evolutionary plasticity of *col* function in the insects. In order to further investigate the degree of evolutionary plasticity of *col* expression outside of the insects, we cloned *col* homologues of the beach hopper *P. hawaiensis*, a representative of the Crustacea, and the common house spider *A. tepidariorum*, a representative of the Chelicerata, and studied their expression. In both cases we could not detect any early expression of *col* in the embryo (not shown). In *P. hawaiensis Ph col* expression was first detected during germ band elongation in the central nervous system including the brain (Fig. 10A, B) and the posterior body portion (pleosoma and urosoma) (Fig. 10C). Additional expression was detected in lateral (=future dorsal tissue) segmental spots along the trunk (Fig. 10C, arrows). In *A. tepidariorum At col* expression was first detected in fully extended germ band stages (not shown) and persists through the process of germ band inversion (Fig. 10D-G). As in *P. hawaiensis*, *At col* expression is always confined to the nervous system and spots in the dorsal portion of the trunk, and the punctate pattern of the segmental expression in the nervous system is virtually identical in all segments (Fig. 10D, E) including the pedipalpal segment that corresponds to the second antennal segment and the intercalary segment in crustaceans and insects, respectively. The expression in the brain is restricted to a small number of cells at the anterior edge (Fig. 10F). In the opisthosoma (abdomen) the neural expression does not reach into the posterior-most body segments (Fig. 10G), probably correlating with the rudimentary nature of the nervous system in this body part.

Discussion

The role of *collier* in insect head development

Previous work in *D. melanogaster* has shown that *col* functions at the interface between the patterning mechanisms of the head and the trunk (Crozatier et al., 1996; Crozatier et al., 1999, Seecoomar et al., 2000). It is expressed in PS0 which gives rise to the posterior part of the intercalary segment and the anterior part of the mandibular segment, and the loss of *col* leads to the loss of these head parts (Crozatier et al., 1996; Crozatier et al., 1999). Our *col* expression data from *T. castaneum* agree with the results of Economou and Telford (2009) and show that *Tc col* is expressed at the junction between the mandibular and the intercalary segment and thus

in a similar region as in *D. melanogaster*. The elevated cell death levels in the intercalary segment area in *Tc col* RNAi embryos at germ band elongation and retraction stages show that the intercalary anlage is gradually reduced by cell death. Gene expression in this segment in *Tc col* RNAi embryos is partially wildtype (normal *Tc lab* expression), but other genes are no longer activated in this area (lacking intercalary *Tc wg*, *Tc en*, and *Tc hh* spots) or are no longer repressed in this area (fused *Tc cnc* domains). The larval bent head phenotype suggests that at this stage of development in *Tc col* RNAi animals, the intercalary segment is already severely reduced if not absent. However, the bristle markers that have been attributed to the intercalary segment on the basis of their absence in *Tc lab* RNAi larvae (Posnien and Bucher, 2009) are all present in *Tc col* RNAi larvae. This suggests that the area deleted in the heads of *Tc lab* RNAi larvae is larger than the area deleted in *Tc col* RNAi animals. Interestingly, in *D. melanogaster* the loss of *lab* also affects an area that is larger than its expression domain in the intercalary segment: loss of *lab* eliminates the intercalary segment, but secondarily also leads to cell death in adjacent segments resulting in the loss of mandibular, maxillary and labial structures (Merrill et al., 1989). Posnien and Bucher (2009) did not test for cell death in the *Tc lab* RNAi animals, but we presume the effect of *Tc lab* loss in *T. castaneum* to be similar to the effect in *D. melanogaster*, leading to the loss of the intercalary segment as well as portions of adjacent segments. This is supported by the fact that some of the marker bristles are not only missing after *Tc lab* RNAi, but are also absent when other head segments are missing.

This interpretation is also supported by our results with *Of col* RNAi in *O. fasciatus*. *Of col* is also expressed spanning the junction between intercalary and mandibular segment, and thus at a similar location as *col* in *D. melanogaster* and *T. castaneum*. The lack of the intercalary expression of *Of pb* and *Of en* shows that pattern formation in the intercalary segment is disturbed in *Of col* RNAi embryos. The lack of the *Of pb* expression in the anterior mandibular segment suggests that this portion is affected as well. In contrast to *D. melanogaster* (Crozatier et al., 1996; Seecoomar et al., 2000) and *T. castaneum* (this work), *Of col* RNAi also reduces the expression of *Of lab*. At later stages of embryogenesis large portions of the antennal and mandibular segments including their appendages degenerate by necrosis. Thus, the loss of *Of col* function in *O. fasciatus* leads to the loss of more segments than would be expected from its expression domain. This is caused by necrotic cell death and seems to be a secondary effect of *Of col* loss, reminiscent of the effect of *lab* loss in *D. melanogaster* (Crozatier et al., 1996; Seecoomar et al., 2000). The triggering of cell death in adjacent head segments upon the loss of larger parts of the intercalary segment – either by direct loss of *lab* (in *D. melanogaster* and *T. castaneum*) or secondary loss of *lab* after loss of *col* function (in *O. fasciatus*) – appears to be a common theme in the insects.

Evolution of *col* function and the origin of the appendage-less character of the tritocerebral segment in insects

The late expression in the central nervous system is highly conserved in all arthropod species that have been investigated, and thus likely represents an ancestral role of *col* in the Arthropoda. A role of *col* related genes in nervous system development is also present in animals as diverse as nematodes, vertebrates, and cnidarians, and thus this role is ancestral in the entire Metazoa (Prasad et al., 1998; Dubois and Vincent, 2001; Pang et al., 2003). The late expression in the mesoderm (muscle primordia) is also present in all arthropods investigated. However, this feature is apparently restricted to arthropods (including data from *D. melanogaster*) (Crozatier and Vincent, 2008).

In *D. melanogaster*, *col* has an early expression domain in the head region that fulfils a crucial role in head development, being necessary for pattern formation and the proper formation of the intercalary and mandibular segments. This function is conserved in *T. castaneum* and *O. fasciatus*, although in *O. fasciatus* differences exist in the regulation of the Hox gene *lab* (discussed above). The function of *col* in the head appears to be an insect novelty, because there is no early head expression of *col* in the spider *A. tepidariorum* and the crustacean *P. hawaiensis*. We propose that the ancestral function of *col* in the arthropods was in the development of the nervous system and in parts of the mesoderm. This condition has been conserved in extant chelicerates and crustaceans, as evidenced by the *col* expression pattern in the spider *A. tepidariorum* and the crustacean *P. hawaiensis*. We speculate that in the insect lineage *col* expression in the tritocerebral segment has come under the control of the head patterning mechanisms and has gained a novel early function in the development of this segment. The occurrence of this novel function of *col* and the ocurrence of a true intercalary segment – an appendage-less tritocerebral segment – coincide, which suggests that *col* in insects has a role in causing the appendage-less state of this segment, probably by controling the normal segment formation machinery (e.g. *hh*, *en*, *wg*) or the appendage formation mechanisms (e. g. *Dll*). To test this idea sophisticated manipulations in *D. melanogaster* are necessary in order to identify the genes and regulatory sequences controlled by *col*.

Acknowledgements

We would like to thank Marco Winkler for assistance with beetle work and Ralf Janssen for providing us with degenerate primers for *collier.* We also thank Paul Z. Liu and Nipam H. Patel for help with the *Oncopeltus* work, and Gregor Bucher for running the *Tribolium* facility in Göttingen. This work has been supported by a University of California Education Abroad

Program Grant and a Fulbright Travel Grant (both to NDS), the European Community's Marie Curie Research Training Network ZOONET under contract MRTN-CT-2004-005624 (to EAW and WGMD) and by a DFG Emmy Noether Program grant PR 1109/1-1 (to NMP).

References

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

Angelini, D.R., Kaufman, T.C., 2004. Functional analyses in the hemipteran *Oncopeltus fasciatus* reveal conserved and derived aspects of appendage patterning in insects. Dev Biol 271,306-321.

Angelini, D.R., Liu, P.Z., Hughes, C.L., Kaufman, T.C., 2005. Hox gene function and interaction in the milkweed bug *Oncopeltus fasciatus* (Hemiptera). Dev Biol 287,440-455.

Baumgardt, M., Miguel-Aliaga, I., Karlsson, D., Ekman, H., Thor, S., 2007. Specification of neuronal identities by feedforward combinatorial coding. PLOS Biol doi:10.1371/journal.pbio.0050037.

Berghammer, A., Bucher, G., Maderspacher, F., Klingler, M., 1999. A system to efficiently maintain embryonic lethal mutations in the flour beetle *Tribolium castaneum*. Dev. Genes Evol. 209,382-389.

Browne, W.E., Price, A.L., Gerberding, M., Patel, N.H., 2005. Stages of embryonic development in the amphipod crustacean, *Parhyale hawaiensis*. Genesis 42,124-149.

Browne, W.E., Schmid, B.G.M., Wimmer, E.A., Martindale, M.Q., 2006. Expression of otd orthologs in the amphipod crustacean, *Parhyale hawaiensis.* Dev. Genes Evol. 216,581–595.

Bucher, G., Scholten, J., Klinger, M., 2002. Parental RNAi in *Tribolium* (Coleoptera). Curr. Biol. 12,R85-R86.

Crozatier, M., Valle, D., Dubois, L., Ibnsouda, S., Vincent, A., 1996. *collier*, a novel regulator of *Drosophila* head development, is expressed in a single mitotic domain. Curr. Biol. 6:6,707-718.

Crozatier, M., Vincent, A., 1999. Requirement for the *Drosophila* COE transcription factor Collier in formation of an embryonic muscle: transcriptional response to Notch signalling. Development 126,1495-1504.

Crozatier, M., Valle, D., Dubois, L., Ibnsouda, S., Vincent, A., 1999. Head versus trunk patterning in the *Drosophila* embryo; *collier* requirement for formation of the intercalary segment. Development 126,4385-4394.

Crozatier, M., Glise, B., Vincent, A., 2002. Connecting Hh, Dpp and EGF signalling in patterning the *Drosophila* wing; the pivotal role of *collier/knot* in the AP organiser. Development 129,4261-4269.

Crozatier, M., Vincent, A., 2008. Control of multidendritic neuron differentiation in *Drosophila*: The role of Collier. Dev. Biol. 315,232-242.

Damen, W.G.M., Hausdorf, M., Seyfarth, E.A., Tautz, D., 1998. The expression pattern of Hox genes in the spider *Cupiennius salei* suggests a conserved mode of head segmentation in arthropods. Proc. Natl. Acad. Sci. USA 95, 10665-10670.

Dubois, L., Vincent, A., 2001. The COE- Collier/Olf1/EBF- transcription factors: structural conservation and diversity of developmental functions. Mech. of Dev. 108,3-12.

Dubois, L., Enriquez, J., Daburon, V., Crozet, F., Lebreton, G., Crozatier, M., Vincent, A., 2007. *collier* transcription in a single *Drosophila* muscle lineage: the combinatorial control of muscle identity. Development 134,4347-4355.

Economou, A.D., Telford, M.J., 2009. Comparative gene expression in the heads of *Drosophila melanogaster* and *Tribolium castaneum* and the segmental affinity of the *Drosophila* hypopharyngeal lobes. Evol. and Dev. 11:1,88-96.

Hersh, B.M., Carroll, S., 2005. Direct regulation of *knot* gene expression by Ultrabithorax and the evolution of cis-regulatory elements in *Drosophila*. Development 132,1567-1577.

Hughes, C.L., Kaufman, T.C., 2000. RNAi analysis of *Deformed*, *proboscipedia* and *Sex combs reduced* in the milkweed bug *Oncopeltus fasciatus*: novel Hox genes in the Hemipteran head. Development 127,3683-3694.

Liberg, D., Sigvardsson, M., Akerblad, P., 2002. The EBF/Olf/Collier family of transcription factors: regulators of differentiation in cells originating from all three embryonal germ layers. Mol. and Cell. Biol. 8389-8397.

Liu, P.Z., 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 131,1515-1527.

Liu, P.Z., Kaufman, T.C., 2004b. *Krüppel* 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.

Lorenzen M.D., Berghammer A.J., Brown S.J., Denell R.E., Klingler M., Beeman R.W., 2003. piggyBac-mediated germline tranformation in the beetle *Tribolium castaneum*. Insect Mol Biol 12,433-440.

Mella, S., Soula, C., Morello, D., Crozatier, M., Vincent, A., 2004. Expression patterns of the *coe/ebf* transcription factor genes during chicken and mouse limb development. Gene Expression Patterns 4,537-545.

Merrill, V.K.L., Diederich, R.J., Turner, F.R., Kaufman, T.C., 1989. A genetic and developmental analysis of mutations in *labial*, a gene necessary for proper head formation in *Drosophila melanogaster*. Dev. Biol. 135,376-391.

Mohler, J., Seecoomar, M., Agarwal, S., Bier, E., Hsai, J., 2000. Activation of *knot* (*kn*) specifies the 3-4 intervein region in the *Drosophila* wing. Development. 127,55-63.

Nestoras, K., Lee, H., Mohler, J., 1997. Role of *knot (kn)* in wing patterning in *Drosophila*. Genetics 147,1203-1212.

Pang, K., Matus, D.Q., Martindale, M.Q., 2004. The ancestral role of COE genes may have been in chemoreception: evidence from the development of the sea anemone, *Nematostella vectensis* (Phylum Cnidaria; Class Anthozoa). Dev. Genes Evol. 214,134-138.

Posnien, N., Bucher, G., (2009). The Hox1 ortholog *Tc-labial* is required for embryonic development of the intercalary segment giving rise to lateral parts of the head in Tribolium. Dev. Biol. doi:10.1016/j.ydbio.2009.11.010.

Prasad, B.C., Ye, B., Zackhary, R., Schrader, K., Seydoux, G., 1998. *unc-3*, a gene required for axonal guidance in *Caenorhabditis elegans*, encodes a member of the O/E family of transcription factors. Development 125,1561-1568.

Prpic, N.M., Damen, W.G.M., 2005. Cell Death During Germ Band Inversion, Dorsal Closure, and Nervous System Development in the Spider *Cupiennius salei.* Dev. Dynamics 234,222-228.

Prpic, N.M., Janssen, R., Damen, W.G.M., Tautz, D., 2005. Evolution of dorsal–ventral axis formation in arthropod appendages: *H15* and *optomotor-blind/bifid*-type T-box genes in the millipede *Glomeris marginata* (Myriapoda: Diplopoda). Evol. Dev. 7,51–57.

Prpic, N.M., Schoppmeier, M., Damen, W.G.M., 2008. Whole-mount in situ hybridization of spider embryos. Cold Spring Harb Protoc. doi:10.1101/pdb.prot5068.

Schaeper, N.D., Prpic, N.M., Wimmer, E.A., 2009. A conserved function of the zinc finger transcription factor *Sp8/9* in allometric appendage growth in the milkweed bug *Oncopeltus fasciatus*. Dev. Genes Evol. 219,427-435.

Schinko, J.B., Kreuzer, N., Offen, N., Posnien, N., Wimmer, E.A., Bucher, G., 2008. Divergent functions of *orthodenticle, empty spiracles* and *buttonhead* in early head patterning of the beetle *Tribolium castaneum* (Coleoptera). Dev. Biol. 317,600-613.

Seecoomar, M., Agarwal, S., Vani, K., Yang, G., Mohler, J., 2000. *knot* is required for the hypopharyngeal lobe and its derivatives in the *Drosophila* embryo. Mech. Dev. 91,209-215.

Telford, M.J., Thomas, R.H., 1998. Expression of homeobox genes shows chelicerate arthropods retain their deutocerebral segment. Proc. Natl. Acad. Sci. USA 95, 10671-10675.

Vervoort, M., Crozatier, M., Valle, D., Vincent, A., 1999. The COE transcription factor Collier is a mediator of short-range Hedghog-induced patterning of the *Drosophila* wing. Curr. Biol. 9,632-639.

Wohlfrom, H., Schinko, J.B., Klingler, M., Bucher, G., 2006. Maintenance of segment and appendage primordia by the *Tribolium* gene *knödel*. Mech. Dev. 123,430-439.

Figures

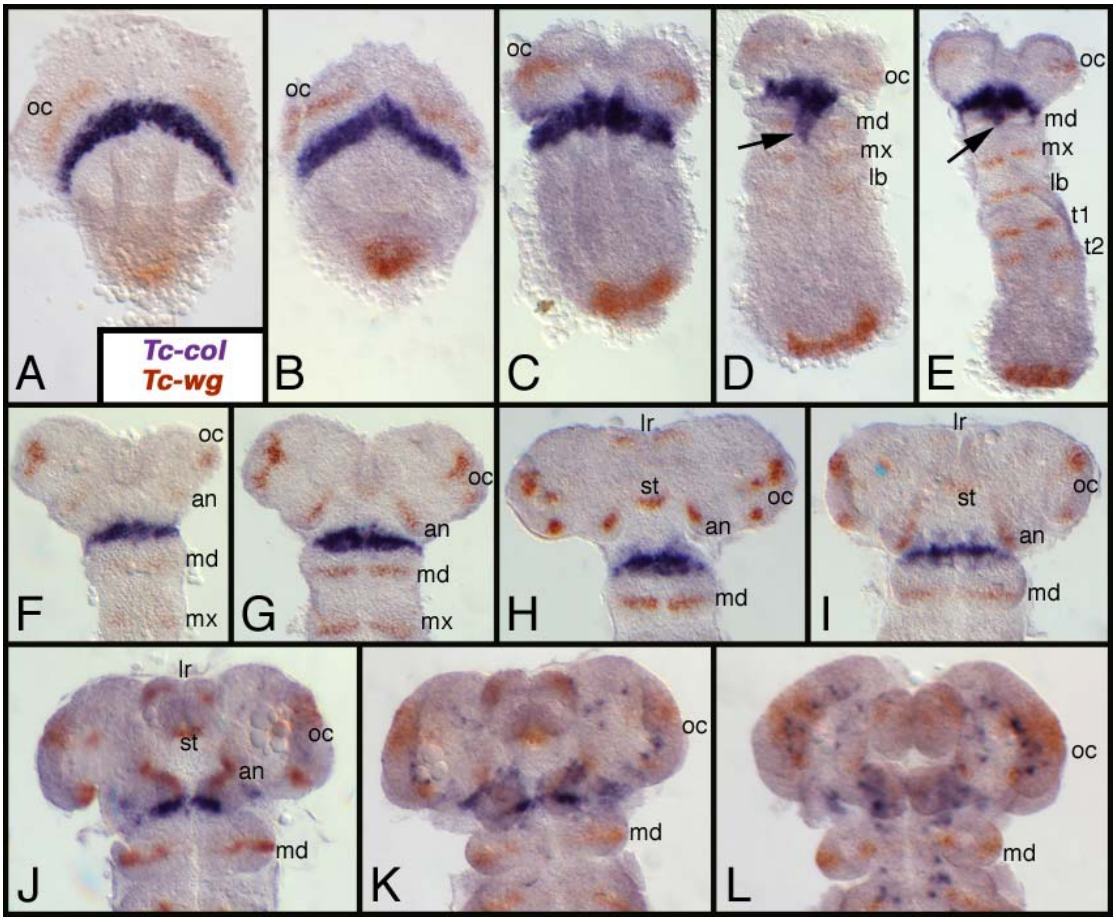


Figure 1. Expression of *Tc col* (blue) and *Tc wg* (red) in *T. castaneum* embryos. (A) Gastrulating germ band. (B) Young germ band, slightly more advanced than the embryo in A. (C) Embryo before serosal closure. (D) Embryo at serosal closure. The arrow points to internal expression likely in the mesoderm. (E) Embryo at the onset of germ band elongation. The arrow points to the internal expression that has already retracted anteriorly compared to earlier stages. (F-L) Anterior ends of embryos at increasingly older developmental stages.

(F) Embryo at early germ band elongation. (G) Mid germ band elongation embryo. (H) Embryo shortly before full elongation. (I) Fully elongated embryo. (J) Embryo at early germ band retraction. (K) Embryo slightly more advanced than that in J. (L) Embryo approaching mid germ band retraction. Anterior is to the top in all panels. Abbreviations: oc, ocular region; md, mandibular segment; mx, maxillary segment; lb, labial segment; t1, prothoracic segment; t2, mesothoracic segment; lr, labrum; st, stomodeum; an, antennal segment.

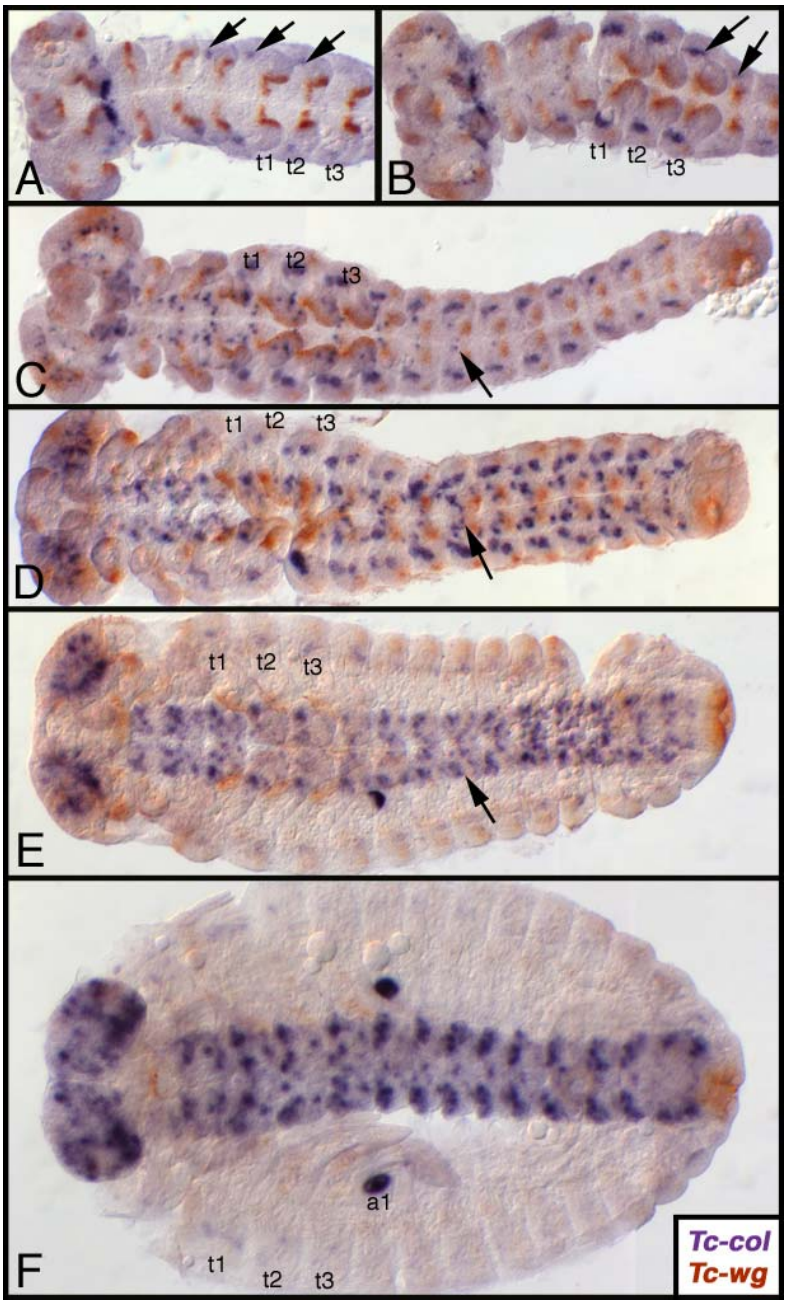


Figure 2. Expression of *Tc col* (blue) and *Tc wg* (red) in *T. castaneum* embryos. (A) Embryo at early germ band retraction. Note the lateral spots of expression in the thorax (arrows). (B) Embryo slightly more advanced than that in A. The lateral expression is now in the thorax and the anterior abdomen (arrows). (C) Embryo approaching mid germ band retraction. The arrow denotes the beginning expression in the ventral nervous system. (D) Embryo at mid germ band retraction. The punctate pattern in the ventral nervous system is denoted by the arrow. (E) Embryo approaching the end of germ band retraction. The arrow points to the expression in the ventral nervous system. (F) Embryo approaching dorsal closure. Anterior is to the left in all panels. Abbreviations: t1, prothoracic segment; t2, mesothoracic segment; t3, metathoracic segment; a1, first abdominal segment. Please note that the dark staining in the appendage (pleuropodium) of the first abdominal segment is a known artefact obtained with any probe at older developmental stages.

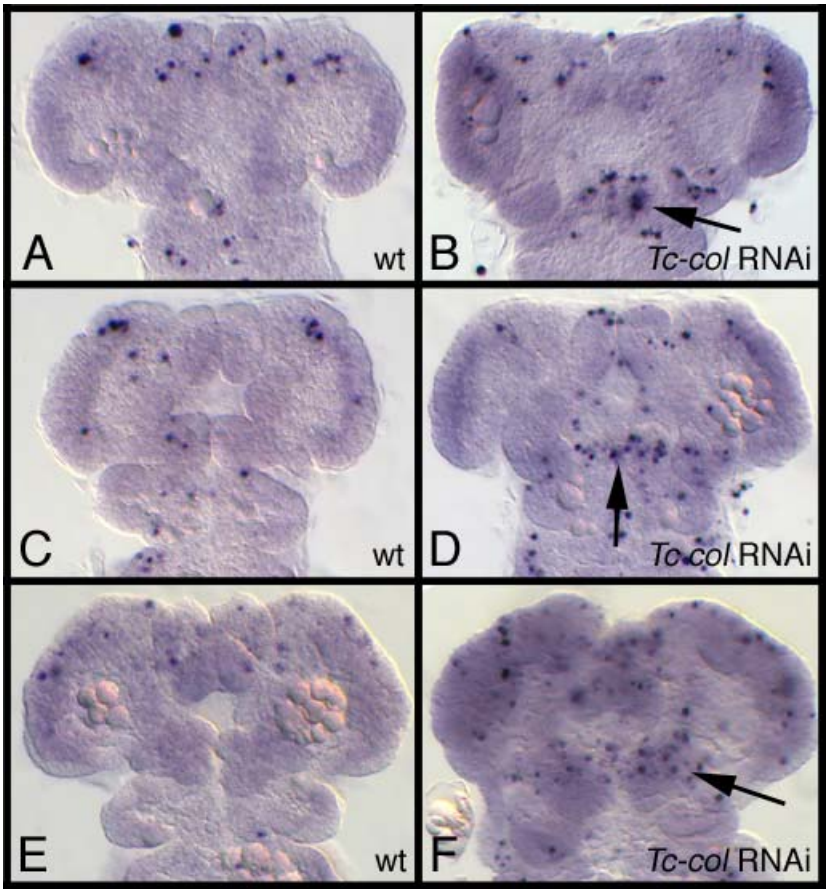


Figure 3. Cell death in *Tc col* RNAi embryos detected by TUNEL. (A, C, E) Anterior ends of wildtype embryos. (B, D, F) Anterior ends of *Tc col* RNAi embryos. (A, B) Embryos at late germ band elongation. The arrow in B points to the patch of labeled cells in the intercalary segment area. (C, D) Early germ band retraction embryos. The arrow in D denotes the band of TUNEL positive cells anterior to the mandibular segment. (E, F) Mid germ band retraction embryos. The arrow in F points to the band of TUNEL positive cells anterior to the mandibular segment and covered by the antennal appendages. Anterior is up in all panels.

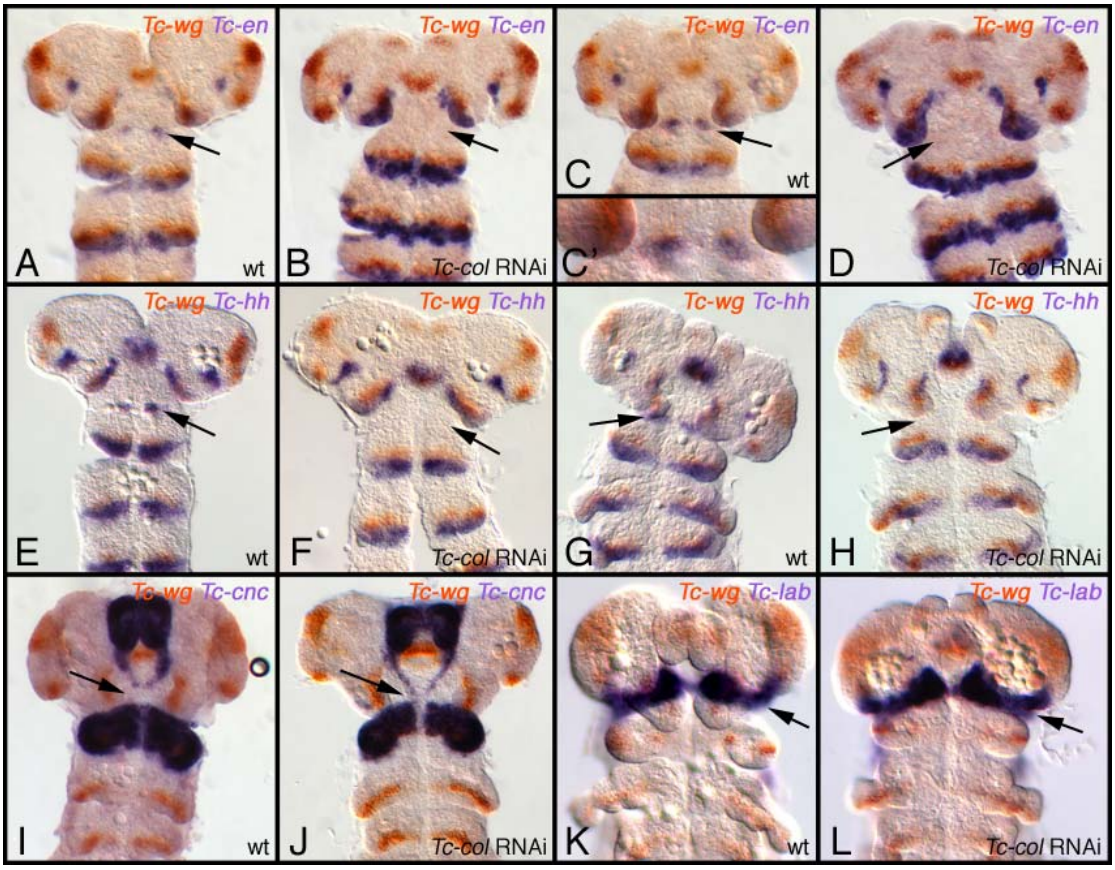


Figure 4. Gene expression in *Tc col* RNAi embryos. All panels show anterior ends of *T. castaneum* embryos with anterior up. (A-D) Expression of *Tc wg* (red) and *Tc en* (blue). (E-H) Expression of *Tc wg* (red) and *Tc hh* (blue). (I, J) Expression of *Tc wg* (red) and *Tc cnc* (blue). (K, L) Expression of *Tc wg* (red) and *Tc lab* (blue). In germ band elongation stages (A) *Tc en* is expressed in two dots in the intercalary segment area (arrow in A). In *Tc col* RNAi embryos at the same age these dots are missing (arrow in B). In germ band retraction stages (C) *Tc wg* is additionally expressed in two dots anteriorly adjacent to the *Tc en* dots (arrow in C, and magnified in C´). These dots are missing in *Tc col* RNAi embryos at the same stage (arrow in D). In germ band elongation (E) and retraction stages (G) *Tc hh* is expressed in two dots in the intercalary segment (arrows). These spots are missing in *Tc col* RNAi embryos of the same age (arrows in F and H). *Tc cnc* is expressed in two separate domains around the stomodeum and in the mandibular segment and the intercalary segment area is free of *Tc cnc* expression (arrow in I). In *Tc col* RNAi embryos the two separate *Tc cnc* domains are joined (arrow in J). *Tc lab* is expressed in a stripe in the intercalary segment area in wildtype (arrow in K) and *Tc col* RNAi embryos (arrow in L).

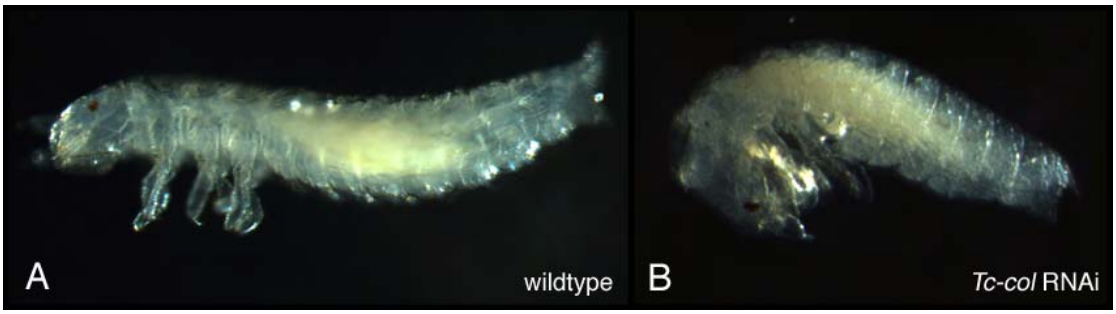


Figure 5. Larval phenotype of *Tc col* RNAi. (A) Live *T. castaneum* wildtype larva. (B) Live *Tc col* RNAi larva with the typically bent head. Anterior is left in all panels.

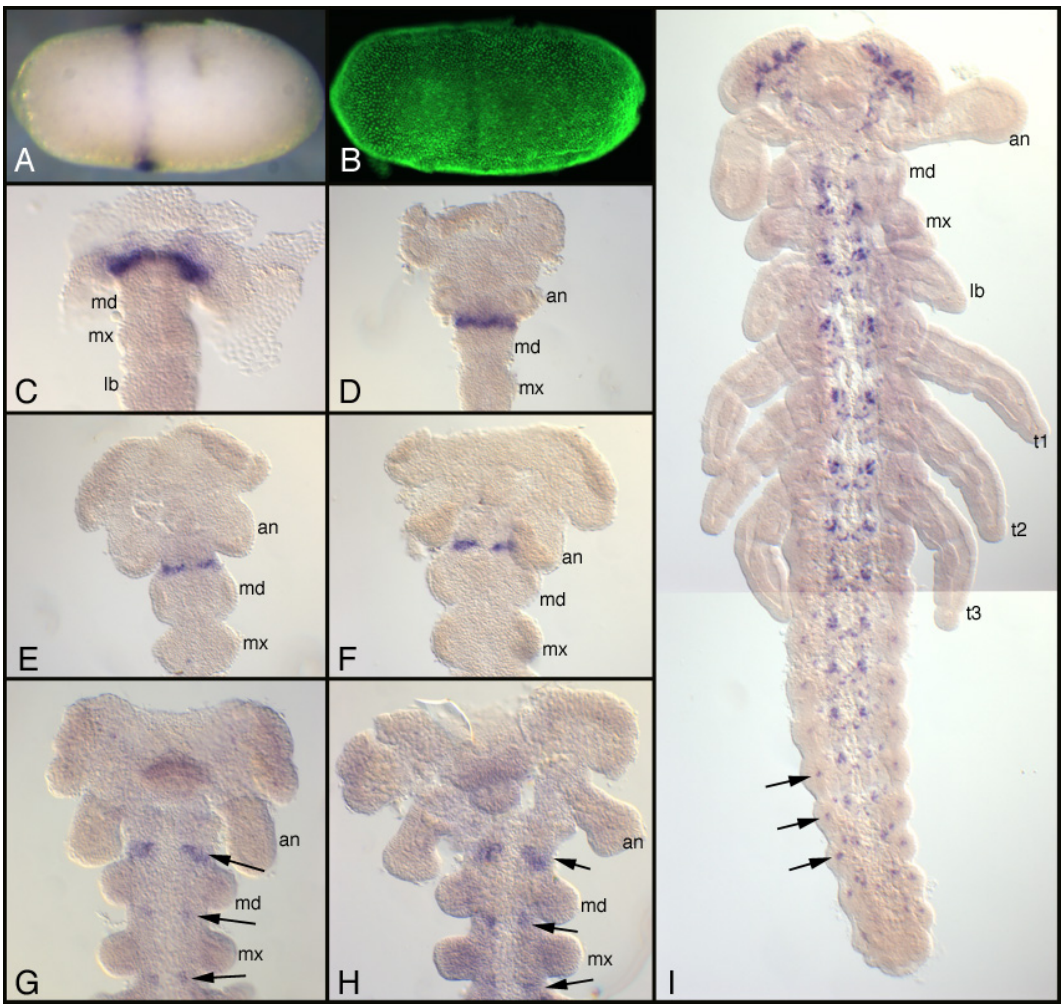


Figure 6. Expression of *Of col* in *O. fasciatus* embryos. (A, B) Blastoderm embryo in brightfield microscopy (A) and fluorescent microscopy in order to visualize the Sytox Green stained nuclei (B). Anterior is to the left in A, B. (C-H) Anterior portion of embryos at different developmental stages. (C) Young germ band. (D) Embryo at early germ band elongation. (E) Embryo at mid germ band elongation. (F) Embryo slightly more advanced than the one in E. (G) Embryo at the end of germ band elongation. Note the expression in the developing ventral nervous system (arrows). (H) Embryo slightly more advanced than the one in G. The expression in the ventral nervous system is stronger (arrows). (I) Embryo at mid germ band retraction. Arrows denote lateral, presumably mesodermal, expression dots. Anterior is up in C - I. Abbreviations: an, antennal segment; md mandibular segment; mx, maxillary segment; lb, labial segment; t1, prothoracic segment; t2, mesothoracic segment; t3, metathoracic segment.

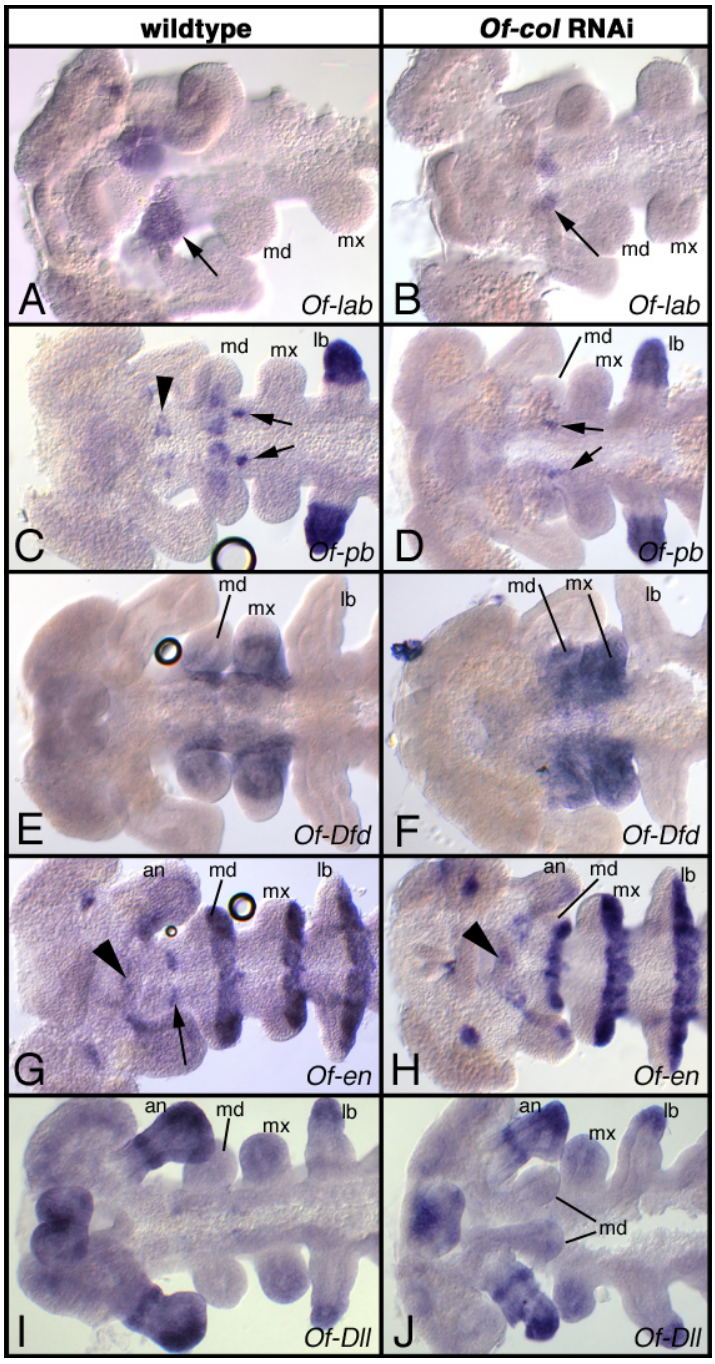


Figure 7. Gene expression and early embryonic phenotype in *Of col* RNAi embryos. (A, B) Expression of *Of lab* in the intercalary segment is reduced in *Of col* RNAi embryos (arrows). (C, D) Expression of *Of pb* in the anterior mandibular segment and the intercalary segment (arrowhead) is missing in *Of col* RNAi embryos (D). The *Of pb* dots in the posterior mandibular segment (arrows) are present in wildtype and *Of col* RNAi embryos. (E, F) Expression of *Of Dfd* in the mandibular and maxillary segment in wildtype (E) and *Of col* RNAi embryos (F). (G, H) The stomodeal expression of *Of en* (arrowhead) is present in wildtype (G) and *Of col* RNAi embryos (H), but the intercalary *en* dots (arrow) are missing in *Of col* RNAi animals. (I, J) Expression of *Of Dll* is indistinguishable in wildtype (I) and *Of col* RNAi embryos (J). Anterior is to the left in all panels. Abbreviations see Fig. 6.

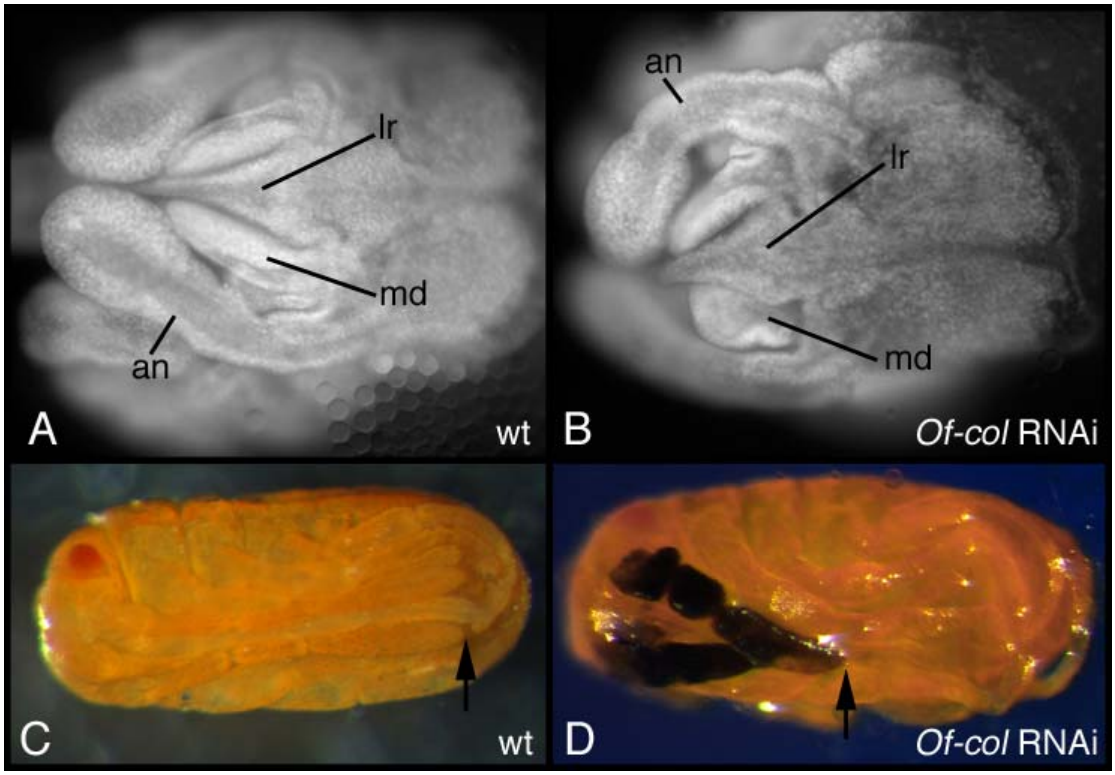


Figure 8. Late embryonic phenotype of *Of col* RNAi embryos. (A, B) Fluorescence micrographs using Sytox Green showing embryos after dorsal closure in frontal view. (A) Wildtype embryo. (B) *Of col* RNAi embryo. Note the malformed mandibles. (C, D) Embryos shortly before hatching. Note the long antennae of the wildtype (arrow in C denotes distal end) and the shortened, necrotic antennae in *Of col* RNAi embryos (arrow in D points to the distal end). Anterior is to the left in C, D. Abbreviations: lr, labrum; an, antenna; md, mandible.

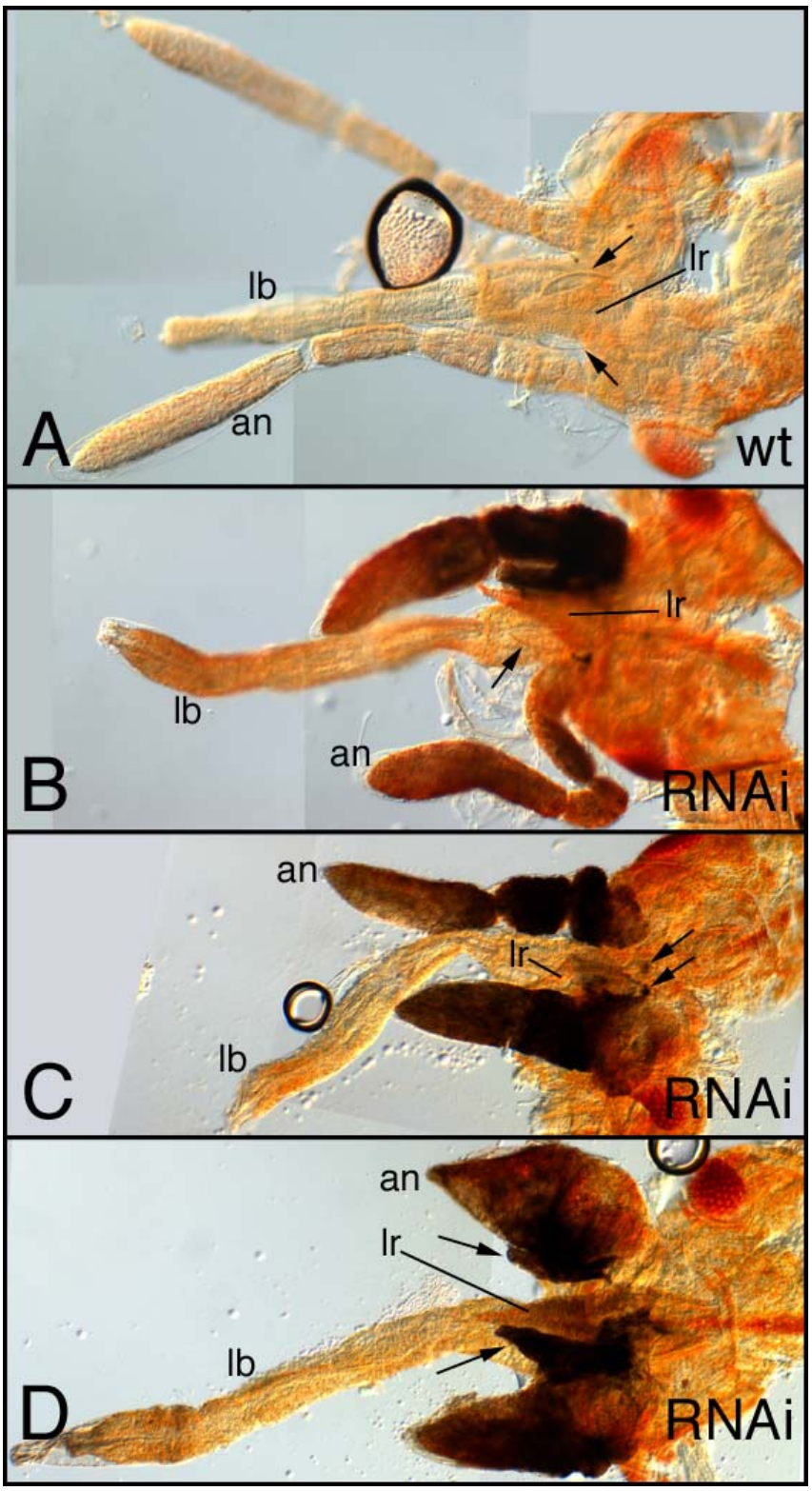


Figure 9. Late embryonic phenotype of *Of col* RNAi embryos. (A) Head of a wildtype embryo shortly before hatching. The arrows point to the stylet-shaped mandibles. (B-D) Heads of *Of col* RNAi embryos shortly before hatching. Arrows point to the mandibles (if visible in the preparation). Note the blackish appearance of the necrotic tissue of the antennae. Anterior is to the left in all panels. Abbreviations: lr, labrum; an, antenna; lb, labium.

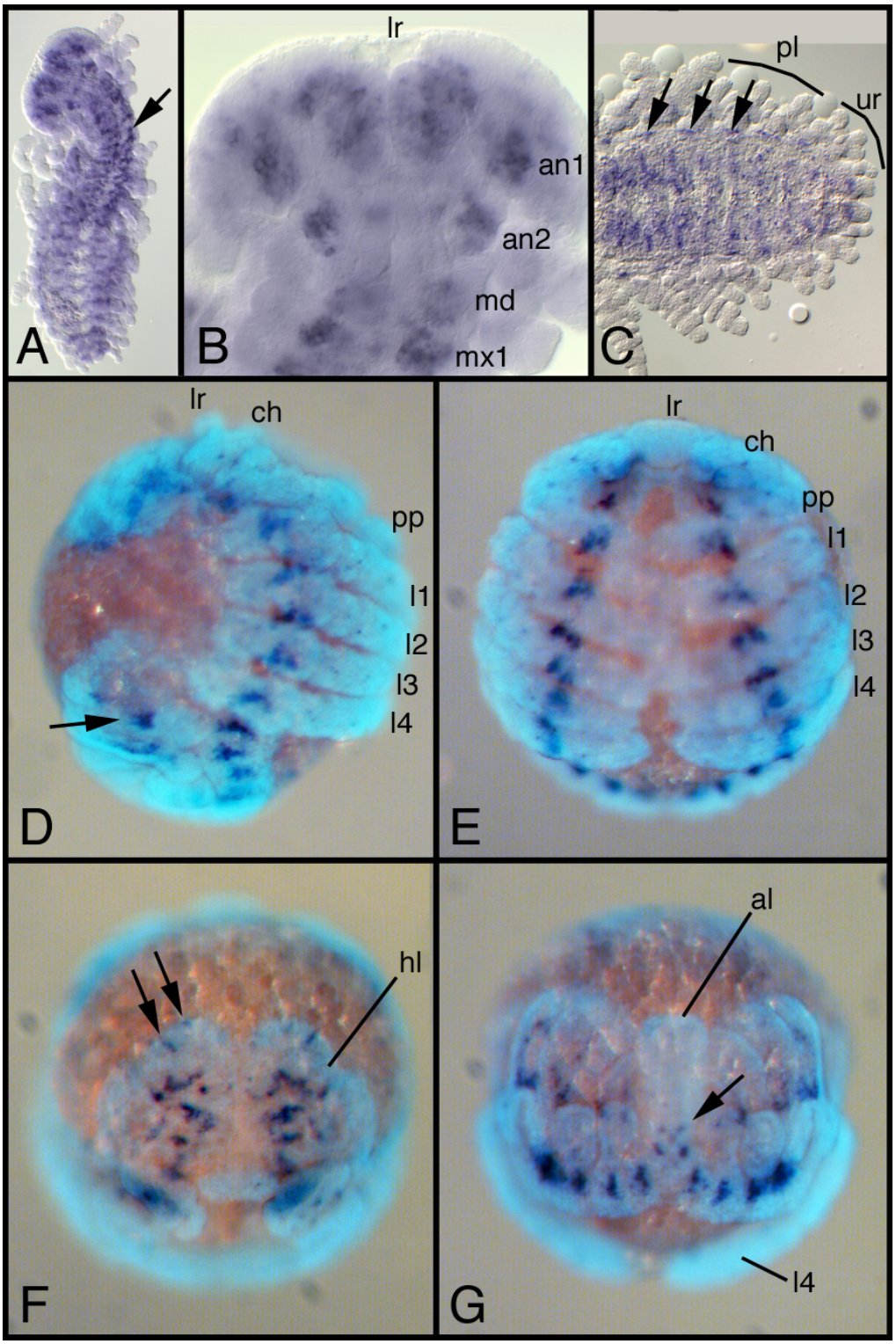

A
B
lr
an1
an2
md
mx1
C
pl
ur
D
lr
ch
pp
l1
l2
l3
l4
E
lr
ch
pp
l1
l2
l3
l4
F
hl
G
al
l4

Figure 10

Figure 10. Expression of *col* in non-insect arthropods. (A-C) Expression of *Ph col* in the crustacean *P. hawaiensis*. (A) Embryo at stage S22 (Browne et al. 2005). The arrow points to the expression in the ventral nervous system. Anterior is up. (B) Magnification of the anterior portion of the embryo in A. Anterior is up. (C) Posterior end of an embryo at stage S23 showing expression in the ventral nervous system and in lateral, presumably mesodermal spots (arrows). Anterior is to the left. (D-G) Expression of *At col* in embryos of the spider *A. tepidariorum* at late inversion. (D) Lateral view and anterior is up. The arrow denotes lateral expression. (E) Ventral aspect and anterior is up. (F) Frontal view. Arrows point to spots in the anterior rim of the head lobes. (G) Posterior view. Note that the expression in the ventral nervous system reaches only up to the 8th opisthosomal segment (arrow) and does not extend to the posterior-most segments. Abbreviations: lr, labrum; an1, first antennal segment; an2, second antennal segment; md, mandibular segment; mx1, first maxillary segment; pl, pleosoma; ur, urosoma; ch, chelicera; pp, pedipalp; l1 - l4, walking leg 1 to 4; hl, head lobe; al, anal lobe.

Supplementary Information

Supplementary Table 1: Statistical analysis of parental *Tc col* RNAi

	WT n (%)	empty egg shells n(%)	bent head phenotype n(%)	total *n*
Tc col pRNAi	293 (33,1%)	120 (13,5%)	473 (53,4%)	886
control (injection buffer)	432 (84,9%)	77 (15,1%)	0 (0%)	509

Supplementary Table 2: Statistical analysis of parental *Of col* RNAi

	WT n (%)	unspecific* n(%)	necrotic antennae phenotype n(%)	total *n*
Of col pRNAi	74 (19,5%)	36 (9,5%)	270 (71%)	380
control (injection buffer)	123 (94%)	8 (6%)	0 (0%)	131

Given are the total numbers (n) of specimens in each category. WT stands for wild-type specimens after injection.

* unspecific phenotypes were specimens that showed some development, but had died before any discernible morphology was formed. These phenotypes had a similar frequency in the pRNAi and in the control injections.

4 Discussion

4.1 The evolution of the Sp transcription factor family

Previous studies in *Drosophila melanogaster* suggested that the two Sp factor genes *D-Sp1* and the head gap-like gene *btd* arose from a recent gene duplication. Evidence for this hypothesis were the similar postblastodermal expression patterns of both genes as well as partially redundant developmental functions, and genomic mapping to the same cytological band on the X-chromosome (Wimmer et al., 1996; Schöck et al., 1999). Besides these two Sp genes, also a third Sp gene, namely *CG5669*, is present in the genome of *D. melanogaster*. No functional data are available for this particular gene at present.
Two scenarios of the origin of these three Sp factor genes are possible. First, a set of two Sp genes is ancestral and *btd* arose recently from a gene duplication or second, three Sp genes were already present in ancestral bilaterian species.

4.1.1 Phylogenetic sequence analysis, protein domain structure and chromosomal location of Sp genes reveal their homology

To better understand the evolution of the Sp family genes, we first searched for Sp factors in the sequenced insect genomes of *Drosophila pseudoobscura*, *Anopheles gambiae*, *Bombyx mori*, *Apis mellifera*, *Nasonia vitripennis* and *Tribolium castaneum* in order to establish the number of Sp family genes in holometabolous insects. In all these holometabolous insects three Sp factor genes are present (see 3.1).
To include more arthropod species also from outside the Holometabola, we performed a PCR based search in cDNA collections from the hemimetabolous insect *O. fasciatus*, the basal hexapods *T. domestica* and *F. candida*, and the cruastacean *P. hawaiensis*. In *T. domestica* and *F. candida*, also three Sp genes could be found, whereas only two could be isolated from *O. fasciatus* and *P. hawaiensis*. We could also detect three Sp genes in the completely sequenced genome of the crustacean *Daphnia pulex*. The most likely explanation for the fact that only two Sp genes could be isolated from *O. fasciatus* and *P. hawaiensis* is that we missed the third Sp gene in our PCR survey. It is unlikely that these transcripts were missing in our cDNA pool because a whole range of embryos from all developmental stages were used. Maybe, the sequence of this Sp homolog is highly derived in these species so that our degenerate primers could not bind properly.

The fact, that an expanded screen for Sp genes outside the arthropods also identified three Sp genes in the completely sequenced genomes of the echinoderm *Strongylocentrotus purpuratus*, the chordate *Branchiostoma floridae* and also in the cnidarian *N. vectensis* as well as the placozoan *Trichoplax adhaerens* supports the idea, that a set of three Sp genes is ancestral in the Metazoa (see 3.1).

The phylogentic analysis of all obtained Sp factors, using the conserved region of the Btd-box and the three zinc fingers revealed three monophyletic groups. One group contains all the vertebrate *Sp1*, *Sp2*, *Sp3* and *Sp4* genes as well as a single representative of each of the invertebrate species. Therefore we termed this group *Sp1-4* clade. The second group contains the vertebrate *Sp5* genes and one invertebrate representative of each species including *D. melanogaster btd*. Hence we termed this group *Sp5/btd* clade. The third group contains the remaining vertebrate Sp genes *Sp6*, *Sp7*, *Sp8* and *Sp9* as well as one invertebrate Sp factor of each species and thus it was termed *Sp6-9* clade. Indeed, the *D. melanogaster D-Sp1* groups into the *Sp6-9* clade which supports the idea for closer homology of this gene to the vertebrate *Sp6-9* genes.

The third *D. melanogaster* Sp factor besides *D-Sp1* and *btd*, *CG5669*, groups in the *Sp1-4* clade and thus each of the three monophyletic group of Sp factors obtained in the phylogenetic analysis contains one Sp gene from *D. melanogaster* (see 3.1). The two isolated Sp genes each from *O. fasciatus* and *P. hawaiensis* group to the *Sp1-4* clade and *Sp6-9* clade, respectively. This indicates that the *Sp5/btd* homologous gene of each species was missed in the PCR survey. The fact that each single invertebrate Sp representative groups into each one of these clades strongly supports the existence of three Sp genes in the ancestral state of the Metazoa, namely one *Sp1-4* gene, one *Sp5/btd* and one *Sp6-9* gene. The higher number of Sp genes in vertebrates is due to independent duplication events in the vertebrate lineage.

Next, protein domains of Sp factors were analyzed to find further evidence for the classification of the subgroups. Bouwman and Philipsen (2002) already grouped the human Sp factors into two groups, one containing the Sp factors Sp1, Sp2, Sp3 and Sp4, and the other group containing the (to that time known) remaining Sp factors Sp5, Sp6, Sp7 and Sp8. In our analysis, we also used the protein sequences from human Sp factors (including Sp9) as well as the three Sp factor sequences of the cnidarian *N. vectensis*, and the arthropods *D. melanogaster*, *A. mellifera*, *T. castaneum*, *F. candida*, *T. domestica*, *O. fasciatus* and *P. hawaiensis*. By considering their protein structure domains other than the conserved part of the Btd-box and the three zinc fingers, we found evidence for subdivision of this protein family in two larger subgroups. One contains all Sp1-4 factors and the other all Sp5/Btd as well as Sp6-9 (see 3.1).

In addition we found characters by which we could further subdivide the Sp5-9/Btd group into a Sp5/Btd and a Sp6-9 group, as also indicated by the phylogenetic analysis (see 3.1). Furthermore, our analyses let us to the detection of slightly different protein structure domains within the N-termini of the human Sp factors Sp6 and Sp7, consisting mainly of Threonine and

Proline instead of only Proline as stated by Bouwman and Philipsen (2002). Therefore, human Sp 6 and Sp7 fit well into the Sp6-9 group, in which the Sp factors contain domains of mainly Serine and Proline or Serine and Threonine. All Sp factors of the Sp1-4 clade are highly conserved in their amino acid composition and show Glutamine as well as Serine and Threonine rich regions.

Besides, the chromosomal location of Sp genes suggests an ancestral triplet. In the cnidarian *Nematostella vectensis*, all three Sp genes are located on one single scaffold (scaffold 53) (Putnam et al., 2007). The close proximity of *Sp1-4*, *Sp5/btd* and *Sp6-9* suggest that these genes form a gene cluster of closely related genes which evolved by tandem gene duplication similar to the Hox gene cluster. In the human genome nine Sp genes are present, which are distributed to only four chromosomes. Interestingly, on the human chromosome 2 there are also three Sps in close proximity to each other: *Sp3*, *Sp5* and *Sp9*. Like in *N. vectensis* one gene of the *Sp1-4* clade, one of the *Sp5/btd* clade and also one of the *Sp6-9* clade form a triplet gene cluster. Due to genome duplications in the vertebrate lineage, these three genes were subject of several duplication events. Thus, besides the three Sp genes on human chromosome 2, *Sp4* and *Sp8* are found on chromosome 7, *Sp1* and *Sp7* on chromosome 12 and *Sp2* and *Sp6* on chromosome 17 in humans.

The genomes of arthropods like *D. melanogaster*, *A. gambiae*, *A. mellifera*, *T. castaneum* and *D. pulex* show that in each case, the *Sp5/btd* and *Sp6-9* genes directly face each other on the same chromosome whereas the *Sp1-4* gene is located on another chromosome. Interestingly, Sp genes are often located close to the Hox genes. In human, Sp3, Sp5 and Sp9 are linked to the Hox D cluster and the remaining Sp genes are arranged in duplets of one Sp1-4 clade gene and one Sp6-9 clade gene which are directly linked to the remaining three Hox clusters (see 3.1). In *D. melanogaster* and *A. gambiae Sp1-4* is linked to the Hox cluster. Unfortunately, *Sp1-4* in *T. castaneum* und *A. mellifera* is placed in unassembled reads. Therefore we cannot make a conclusion about their location relative to the Hox genes for now. The scaffold data from *N. vectensis* were partially assembled into putative ancestral metazoan linkage groups (Putnam et al., 2007; Ryan et al., 2007). These data also suggest that the Sp genes are located next to Hox genes on the hypothetical ancestral linkage group PAL A. This suggests that the Hox gene cluster and the Sp cluster were ancestrally located next to each other and might have co-evolved.

It would be interesting to search for Sp genes in the genomes of sponges (e.g. *Amphimedon queenslandica*) and choanoflagellates (e.g. *Monosiga brevicollis*). The genome sequencing of these two species will be finished soon. The set of Sp genes in these species could be analyzed to draw further conclusions about the ancestral state of Sp factors. It might be the case that a set of two Sp genes were ancestral, namely *Sp1-4* and *Sp5-9*, and then *Sp5-9* got duplicated later which resulted in the triplet Sp set already existing in *Trichoplax adhaerens* und *Nematostella vectensis*. It would also be important to further assemble and annotate the sequenced basal

metazoan genomes to draw conclusions about the relationship of Sp genes to Hox genes and putative ancestral Hox clusters.

4.1.2 Expression pattern analyses of Sp genes further support their orthology

Besides, the analyzed expression patterns of arthropod Sp genes support their distribution to the three clades of Sp factors and the orthology to the vertebrate Sp genes. The analyzed arthropod *Sp1-4* genes are expressed ubiquitously throughout embryogenesis like it is known for their vertebrate homologs and therefore supporting their orthology. *Sp5/btd* is very similarly expressed in *D. melanogaster*, *T. castaneum* und *T. domestica*, namely in an early anterior stripe and later a metameric pattern emerges. In older stages an expression in the appendages and the nervous system can be detected. We were not able to fix early *F. candida* embryos but the later expression pattern of *Sp5/btd* is similar to those of the other analyzed arthropods with expression domains in the nervous system and appendages (see 3.1).
Zebrafish *Sp5* is expressed in a head stripe along the midbrain-hindbrain boundary, in the otic vesicles, diencephalon, tail bud and also in the somites (Tallafuss et al., 2001). *Sp5-like* in zebrafish is expressed in a partially overlapping pattern in ectodermal and mesodermal tissue, brain, trunk crest cells and somites (Zhao et al., 2003). Mouse *Sp5* is also expressed in a head stripe in the midbrain-hindbrain boundary, the primitive streak, and later on in the tail bud, otic vesicles, pharyngeal region, central nervous system, somites and limb buds (Treichel et al., 2001). In summary, vertebrate expression patterns of *Sp5/btd* genes are similar to the arthropod *Sp5/btd* expression and thus support the orthology of this clade.
Sp6-9 expression in insects and crustaceans is highly conserved. The expression pattern becomes first visible when the appendages start growing. In *D. melanogaster*, *D-Sp1* is expressed in the developing leg and antennal discs (Estella et al., 2003). Later on, *Sp6-9* is also expressed in the nervous system. The arthropod *Sp6-9* expression patterns are similar to the pattern of vertebrate *Sp8* and *Sp9*. Data from mouse, chicken and zebrafish revealed an expression for both genes in the limbs and fins and for *Sp8* also in the nervous system (e.g. Treichel et al., 2003; Kawakami et al., 2004). Functional data for arthropod genes of this clade are available for *Of Sp8/9* and *Tc Sp8* (Beerman et al., 2004; Schaeper et al., 2009 (see 3.3)). In both insects these genes are crucial for limb outgrowth, like it is the case for the vertebrate *Sp8* and *Sp9* genes (e.g. Treichel et al., 2003; Kawakami et al., 2004). Therefore, *Sp6-9* has an ancestral role in appendage formation irrespective of the homology of the appendages. This is also supported by the expression pattern of *Sp6-9* in the tentacles of the sea anemone *N. vectensis* (see 3.4).

Insights into the possible ancestral function of Sp genes are provided by the sea anemone *N. vectensis*. First analyses for the three Sp genes in *N. vectensis* showed expression in single scattered cells throughout the embryo in gastrula and early planula stages (see. 3.4). These cells seem to be interstitial cells and nematocytes. Besides, *Nv Sp6-9* is expressed in the budding tentacles of metamorphosing planulas. This expression might indicate an ancestral role for *Sp6-9* in body wall outgrowths, as it was also shown for the appendage gene *Distal-less* (*Dll*) (Panganiban et al., 1997). Due to problems with *N. vectensis* husbandry and perturbed spawning and resulting embryonic developmental problems we were not sure about the validity of these data. Therefore, these preliminary data have to be confirmed.

In summary, the data strongly support the idea that a set of three Sp genes is ancestral within the metazoans. According to this, the recent gene duplication theory proposed for *D. melanogaster btd* and *D-Sp1* has to be regarded as unlikely. However, as mentioned above, it is possible that more basal organisms than *T. adhaerens* possess only two Sp genes, one *Sp1-4* gene and one *Sp5-9* gene. This *Sp5-9* ancestor then became duplicated which led to the eumetazoan triplet of Sp factors. Thus, *btd* and *D-Sp1* might be descendants from one ancestral gene, but this duplication just took place much earlier in evolution than proposed.

4.2 Influence of Sp family members in arthropod head development

In *D. melanogaster*, *btd* is a gap-like head gene which causes severe cephalic defects in the antennal, mandibular and intercalary region when mutated (Cohen and Jürgens, 1990; Wimmer et al., 1993). When knocked down in *T. castaneum*, *btd* does not cause any detectable head defect (Schinko et al., 2008) but affected embryos and larvae show leg defects (see 3.2). The second Sp factor in *T. castaneum*, *Tc Sp8*, is involved in allometric limb growth (Beermann et al., 2004) (see also 3.2). Interestingly, the third Sp factor in *T. castaneum* besides *Tc btd* and *Tc Sp8*, named *Tc Sp1-4* (see 3.1) leads to headless larvae when knocked down via RNAi. These data show that *Tc Sp1-4*, but not *Tc btd*, has a significant role in head development in *T. castaneum* being required for the formation of all head segments. This suggests that the developmental functions can be switched between paralogs of the same gene family.

Sp proteins share a very similar region containing three zinc fingers as putative DNA binding motifs. Therefore, all three proteins might have similar DNA binding activities in vivo and regulate similar sets of target genes. Indeed, Schöck et al. (2000) could demonstrate a complete rescue of the *D. melanogaster btd* mutant phenotype to wildtype by providing an engineered form of Btd in which the *Dm* Btd zinc finger motif had been replaced with the human Sp1 zinc finger motif. Since human Sp1 belongs to the Sp1-4 clade of Sp factors and *Dm* Btd to the Sp5/Btd clade (see 3.1), this indicates that at least zinc finger binding specificities are not

specific to the different Sp clades. Other motifs in these proteins might thus be responsible for their specific functional attributes.

A certain degree of functional redundancy might be possible, as already demonstrated in *D. melanogaster*: *D-Sp1* (the ortholog of *Tc Sp8*) and *btd* can partially substitute for each other in mediating leg growth (Estella et al., 2003). Putative functional redundancies might mask a head role of *Tc btd* or *Tc Sp8*, or might mask an additional role of *Sp1-4* in the formation of the thorax or the abdomen. To exclude this possibility, combinatorial double-injections in *T. castaneum* were performed. Only double injections for *Tc Sp8* and *Tc btd* led to a more severe phenotype than each gene alone, indicating that the two genes have partially redundant functions in leg development, similar to the situation in *D. melanogaster* (Estella et al., 2003). The other Sp combinations seem not to have partially redundant functions. However, RNAi often does not lead to a complete knockdown of gene function. Furthermore, in double injected beetles the amount of each dsRNA is lesser than in single injected animals. This could explain why double injections of two Sp genes do not enhance the phenotype. Therefore *T. castaneum* mutants for these three Sp genes have to be identified in order to fully clarify the question of redundancy. Additional experiments should be carried out to further investigate the *Tc Sp1-4* phenotype.

Recent studies in *T. castaneum* have already indicated that the head patterning mechanisms in this species differ from those in *D. melanogaster*. Intriguingly, the headless phenotypes resulting from *Tc Sp1-4* RNAi are virtually identical to the *Tc otd* RNAi phenotypes (Schröder, 2003; Schinko et al., 2008). This suggests a function for *Tc otd* and *Tc Sp1*-4 on similar levels in the same genetic circuit during head patterning. Schröder (2003) could also show a synergistic influence of *otd* and *hb* in *T. castaneum*. In double RNAi animals, only a few abdominal segments remain. Therefore removal of *hb* severes the *otd* phenotype. Thus, also *hb* should be analyzed in *Tc Sp1-4* RNAi embryos. *Otd* expression in *Tc Sp1-4* RNAi embryos is wildtype. Therefore we suggest that *Sp1-4* acts directly downstream of *otd* but upstream of factors determining smaller areas of the head, like e.g. *Tc col*. To address this question further, *Sp1-4* should be also stained in *Tc otd* embryos. Schoppmeier et al. (2009) could show a function for *Mex-3* in *T. castaneum* head development. The loss of function phenotype of *Mex-3* is also very similar to the *Sp1-4* RNAi phenotype. *Mex-3* and *zen-2*, a gene sharing common ancestry with the dipteran anterior morphogen Bicoid, together repress Caudal. Therefore, *Mex-3*, *zen-1* and *zen-2* should also be stained in *Tc Sp1-4* embryos, and vice versa.

Tc Sp1-4 is ubiquitously expressed and knockdown experiments resulted in a tissue specific phenotype. Also mouse *Sp1*, *Sp3* and *Sp4* are ubiquitously expressed but knock out mice show also tissue specific phenotypes (Marin et al., 1997). This suggests that the function of the *Sp1-4* genes might be regulated on the translational rather than the transcriptional level. Alternatively, the protein activity or stability is regulated in a tissue specific manner by yet unknown co-factors. It would be interesting to study Sp1-4 distribution in *T. castaneum* by generating a protein-specific antibody. Moreover, Wimmer et al. (1997) could show, that misexpression of *Dm btd* in

the entire anterior half of the *D. melanogaster* embryo can rescue the *btd* mutant phenotype to wildtype. Thus, the ectopic expression of *btd* does not disturb anterior patterning. It seems that expression boundaries of a certain gene do not always correlate with its function, which could also be the case for *Tc Sp1-4*.

Due to the interesting head phenotype of *Tc Sp1-4*, orthologs of this gene should also be analysed in further arthropod species. Even in *D. melanogaster* no data are available for the *Sp1-4* homolog *CG5669*. Neither mutant- nor P-element lines are currently available for *CG5669*. First preliminary knock down experiments with transgenic UAS-RNA hairpin lines (Dietzl et al., 2007) for *CG5669* did not show any head phenotype (data not shown). More experiments with different Gal4 drivers have to be performed to reveal the function of *CG5669*. The function of *Sp1-4* should also be investigated in the hemimetabolous insect *O. fasciatus* since RNAi is working in this species (see 3.3 and 3.5).

4.3 Contribution of *collier* to the formation of the tritocerebral segment

The *D. melanogaster* Sp factor and head gap-like gene *btd* is needed for proper activation of the second-level regulator *col* (Crozatier et al., 1999). It was shown that *Dm col* is needed for specifying the region of PS0 and is therefore an important factor working at the interface between procephalon and gnathocephalon. The loss of *col* leads to missing structures from the intercalary and mandibular segments (Crozatier et al., 1996; Crozatier et al., 1999). Our data from *col* homologs of other insects showed that *col* is also expressed early in a stripe in the region of the intercalary and mandibular segments. Later on it is expressed in the nervous system like it is the case for *Dm col*. However, *col* homologs from the crustacean *P. hawaiensis* and the common house spider *A. tepidariorum* are not expressed early but only later in development in the nervous system (see 3.5).

Loss of *col* in *T. castaneum* caused larvae with a bent head phenotype but all head bristles, which can be assigned to certain head segments (Schinko et al., 2008) are still present. Further analyses showed that *Tc col* RNAi leads to increased cell death in the intercalary region and the intercalary spots of the segment polarity genes *en*, *hh* and *wg* are missing and the *cap'n'collar* (*cnc*) expression domains fuse. However, *lab* expression is wildtype in *Tc col* RNAi embryos. The same effects on marker gene expression were observed for *Dm col* deficient flies (Crozatier et al., 1996; Seecomar et al., 2000).

The bent head phenotype, cell death data and the affected marker gene expression indicate that the intercalary segment was affected by RNAi and that *Tc col* has therefore a role in formation of the intercalary segment. On the other hand, we could not observe changes in the *T. castaneum* putative intercalary segment bristles as it was observed in *Tc lab* RNAi larvae (Posnien and Bucher, 2009b). One explanation for these divergent results could be that in *Tc col* RNAi only

the posterior part of the intercalary segment is affected whereas the anterior part remains unaffected. Maybe the posterior part of the intercalary does not give rise to head cuticle bristles. Another explanation would be that the *Tc lab* RNAi bristle phenotype is caused by cell death in the intercalary and adjacent head segments. This was reported for *Dm lab* where loss of *lab* leads to secondary cell death not only in the intercalary segment, but also in the adjacent segments and therefore loss of mandibular, maxillary and labial structures (Merrill et al., 1989). This is also indicated by the fact that some head bristle markers are not only missing in *lab* RNAi larvae, but are also absent when other head segments are missing (Posnien and Bucher, 2009b).

To answer this question, *lab* RNAi embryos should be stained for *col* to test whether *col* is affected. Furthermore, TUNEL should be performed to test for increased cell death in the head region in *lab* RNAi embryos.

col RNAi also causes defects in head segments in *O. fasciatus*. Molecular data from marker gene expression in *col* RNAi embryos indicate that the intercalary segment and the anterior part of the mandibular segment are missing. Interestingly, in contrast to *col* knockdown in *T. castaneum* and *D. melanogaster*, *lab* expression from early on is almost gone in *Of col* RNAi embryos. This suggests that *lab* activation requires *col* in *O. fasciatus*. Later in development it becomes visible that the antennae and mandibles are also affected in their morphology. They are much shorter than in wildtype and become necrotic. This necrotic cell death might also be a secondary effect as described for *Dm lab* and in *O. fasciatus* might actually be caused by the down regulation of *Of lab* after *Of col* RNAi.

In contrast to insects, in the crustacean *P. hawaiensis* and the chelicerate *A. tepidariorum col* is not expressed early in the head, but only late in the nervous system. Data not only from *D. melanogaster*, but also from such diverse animals as nematodes, cnidarians and vertebrates show a function for *col* homologs in the central nervous system (Crozatier et al., 1996; Prasad et al., 1998; Dubois and Vincent, 2001; Pang et al., 2003; Crozatier and Vincent, 2008). Thus, the late function of *col* is ancestral within the metazoa. The observed mesodermal expression of *col* (muscle primordia) seems to be specific for arthropods, as this function of *col* was not observed outside the arthropods so far. Our data suggest that the early head function of *col* is a novelty within the insects where *col* gained a novel function in the control of tritocerebral segment patterning mechanisms. Obviously this new function in the tritocerebral segment coincides with the lack of appendages, probably by controlling the known segment formation machinery (e.g. *en, hh, wg*) and / or the appendage formation mechanisms (e.g. *Dll*).

However, unpublished data from the myriapod *Glomeris marginata* (millipede) show that *col* is also expressed in an early head stripe (Ralf Janssen, personal communication). The tritocerebral segment in myriapods also lacks a pair of appendages like it is the case in insects. These findings can be interpreted in different ways. First, contrary to recent phylogenies (e.g. Hwang et al., 2001) myriapods could indeed be the sister group of insects as it was proposed in

the past (e.g. Snodgrass, 1938; Cisne, 1974). The early *col* expression in the head indicates a common ancestry of these two groups. Second, the early head function of *col* evolved independently within insects and myriapods and is connected with the appendage-less state of the tritocerebral segment in these two groups.
This scenario could thus be one example for a newly gained expression domain that contributes to drastical changes in morphology. In the case of the tritocerebral segment appendages, this change could be correlated with a new mode of life (e.g. the conquest of land). Extended studies should be performed in *D. melanogaster* to find more regulatory sequences and genes controlled by *col*.

4.4 Concluding remarks: Conserved function of conserved genes?

Genetic studies revealed a number of similarities in regulatory genes and signalling molecules that function in growth and patterning of specific structures in insects and vertebrates (e.g. eye formation (Hanson, 2001) or limb outgrowth (Panganiban, 1997)). Striking similarities during limb formation were shown for the *Dll* gene in such diverse organisms as vertebrates, arthropods, onychophorans, annelids and echinoderms (e.g. Panganiban, 1997) despite different anatomies and evolutionary histories of these diverse appendages.
In this study, the examined expression pattern and function of the Sp family member *Sp6-9* in various arthropods and a cnidarian representative (see 3.1, 3.3, 3.3, 3.4) as well as existing vertebrate data (e.g. Kawakami et al., 2004) also revealed a comparable conserved function in appendage development as it was shown for *Dll* (e.g. Panganiban, 1997). Investigations of the *D. melanogaster* head gene *col* showed a conserved function for *col* in the formation of the intercalary segment (see 3.5). However, this early head function of *col* seems to be restricted to appendage-less tritocerebral segments in the arthropods: An early head function for *col* could only be shown in insects and myriapods (Ralf Janssen, personal communication). For other arthropod classes like crustaceans and chelicerates (see 3.5) as well as all other bilaterian species *col* only possesses a late neural function (e.g. Prasad et al., 1998; Pang et al., 2003; Crozatier and Vincent, 2008), which seems to be ancestral. On the other hand, other genes showed less conserved functions, even when compared among the same arthropod class. For example, *Dm btd* is a head gap-like gene crucial for the formation of the antennal, mandibular and intercalary segment (Cohen and Jürgens, 1990; Wimmer et al., 1993). In *T. castaneum btd* obviously has no influence on head metamerization since no head phenotype could be observed in *Tc btd* RNAi knockdown (see 3.2). Furthermore, in *D. melanogaster*, *btd* is required for *col* activation. However, in *T. castaneum btd* RNAi animals, *col* expression is still

wildtype. Moreover, a new function of a hitherto unstudied gene in invertebrates could be shown. The *btd* related gene *Sp1-4* has a role in head development in *T. castaneum* (see 3.2).

These latter data suggest that differences in head pattern formation exist not only between vertebrates and invertebrates, but already within different arthropod classes or even among different insect species. This might be due to adaptation to special conditions in one particular organism or group. To reveal the diversity of the genetic mechanisms used for patterning of the head segments, much more arthropod species and outgroup species (e.g. Onychophora, Tardigrada) have to be included into developmental studies to define which parts are derived and to what extent and which parts are conserved. Furthermore, additional methods have to be developed and adapted to new organisms because the classical candidate gene approach encounters difficulties and new gene functions need to be identified.

5 References

Baumgardt M, Miguel-Aliaga I, Karlsson D, Ekman H, Thor S. Specification of neuronal identities by feedforward combinatorial coding. *PLOS Biol* 2007, doi:10.1371/journal.pbio.0050037.

Beermann A, Aranda M, Schröder R. The *Sp8* zinc-finger transcription factor is involved in allometric growth of the limbs in the beetle *Tribolium castaneum*. *Development* 2004, 131:733-742.

Bell SM, Schreiner CM, Waclaw RR, Campbell K, Potter SS, Scott WJ. *Sp8* is crucial for limb outgrowth and neuropore closure. *Proc Natl Acad Sci USA* 2003, 100:12195-12200.

Berleth T, Burri M, Thoma G, Bopp D, Richtstein S, Frigerio G, Noll M, Nüsslein-Volhard C. The role of localization of *bicoid* RNA in organizing the anterior pattern of the *Drosophila* embryo. *EMBO* J 7:1749-1756.

Black AR, Jensen D, Lin SY, Azizkhan-Clifford J. Growth/Cell Cycle Regulation of Sp1 Phosphorylation. *Journ of Biol Chem* 1999, 274:1207–1215.

Black AR, Black JD, Azizkhan-Clifford J. Sp1 and krüppel-like factor family of transcription factors in cell growth regulation and cancer. *J Cell Physiol* 2001, 188:143-160.

Bouwman P, Göllner H, Elsässer HP, Eckhoff G, Karis A, Grosveld F, Philipsen S, Suske G. Transcription factor Sp3 is essential for post-natal survival and late tooth development. *EMBO J* 2000, 19:655-661.

Bouwman P, Philipsen S. Regulation of the activity of Sp1-related transcription factors. *Mol Cell Endocrinol* 2002, 95:27-38.

Bucher G, Wimmer EA. Investigating embryonic head formation using a novel model organism. *B.I.F Futura* 2005, 20:164-168.

Chen Y, Guo Y, Ge X, Itoh H, Watanabe A, Fujiwara T, Kodama T, Aburatani H. Elevated expression and potential roles of human Sp5, a member of Sp transcription factor family, in human cancers. *Biochem Biophys Res Commun* 2006, 340:758-766.

Cisne JL. Trilobites and the evolution of arthropods. *Science* 1974, 186:13-18.

Cohen SM, Jürgens G. Mediation of *Drosophila* head development by gap-like segmentation genes. *Nature* 1990, 346:482-485.

Crozatier M, Valle D, Dubois L, Ibnsouda S, Vincent A.. *collier*, a novel regulator of *Drosophila* head development, is expressed in a single mitotic domain. *Curr Biol* 1996, 6:707-718.

Crozatier M, Vincent A. Requirement for the *Drosophila* COE transcription factor Collier in formation of an embryonic muscle: transcriptional response to Notch signalling. *Development* 1999, 126:1495-1504.

Crozatier M, Valle D, Dubois L, Ibnsouda S, Vincent A. Head versus trunk patterning in the *Drosophila* embryo; *collier* requirement for formation of the intercalary segment. *Development* 1999, 126:4385-4394.

Crozatier M, Glise B, Vincent A. Connecting Hh, Dpp and EGF signalling in patterning the *Drosophila* wing; the pivotal role of *collier/knot* in the AP organiser. *Development* 2002, 129:4261-4269.

Crozatier M, Vincent A. Control of multidendritic neuron differentiation in *Drosophila*. The role of Collier. *Dev Biol* 2008, 315:232-242.

Damen WGM, Hausdorf M, Seyfarth EA, Tautz D. The expression pattern of Hox genes in the spider *Cupiennius salei* suggests a conserved mode of head segmentation in arthropods. *Proc. Natl. Acad. Sci. USA* 1998, 95:10665-10670.

Dietzl G, Chen D, Schnorrer F, Su KC, Barinova Y, Fellner M, Gasser B, kinsey K, Oppel S, Scheiblauer S, Couto A, Marra V, Keleman K, Dickson BJ. A genome-wide transgenic RNAi library for conditional gene inactivation in *Drosophila*. *Nature* 2007, 448:151-157.

Dubois L, Vincent A. The COE- Collier/Olf1/EBF- transcription factors. structural conservation and diversity of developmental functions. *Mech of Dev* 2001, 108:3-12.

Dubois L, Enriquez J, Daburon V, Crozet F, Lebreton G, Crozatier M, Vincent A. *collier* transcription in a single *Drosophila* muscle lineage: the combinatorial control of muscle identity. *Development* 2007, 134:4347-4355.

Dynan WS, Tjian R. The promoter-specific transcription factor Sp1 binds to upstream sequences in the SV40 early promoter. *Cell* 1983a, 35:79-87.

Dynan WS, Tjian R. Isolation of transcription factors that discriminate between different promoters recognized by RNA polymerase II. *Cell* 1983b, 32:669-680.

Estella C, Rieckhof G, Calleja M, Morata G. The role of *buttonhead* and *Sp1* in the development of the ventral imagical discs of *Drosophila*. *Development* 2003, 130:5929-5941.

Finkelstein R, Perrimon N. The molecular genetics of head development in *Drosophila melanogaster*. *Development* 1991, 112:899-912.

Frohnhöfer HG, Nüsslein-Volhard C. Organization of anterior pattern in the *Drosophila* embryo by the maternal gene *bicoid*. *Nature* 1986, 324:120-125.

Gallitano-Mendel A, Finkelstein R. Novel segment polarity gene iteractions during embryonic head development in *Drosophila*. *Dev Biol* 1997, 192:599-613.

Gallitano-Mendel A, Finkelstein R. Ectopic *orthodenticle* expression alters segment polarity gene expression but not head segment identity in the *Drosophila* embryo. *Dev Biol* 1998, 199:125-137.

Göllner H, Dani C, Phillips B, Philipsen S, Suske G. Impaired ossification in mice lacking the transcription factor Sp3. *Mech Dev* 2001a, 106:77-83.

Göllner H, Bouwman P, Mangold M, Karis A, Braun H, Rohner I, Del Rey A, Besedovsky HO, Meinhardt A, van den Broek M, Cutforth T, Grosveld F, Philipsen S, Suske G. Complex phenotype of mice homozygous for a null mutation in the Sp4 transcription factor gene. *Genes Cells* 2001b, 6:689-697.

Griesel G, Treichel D, Collombat P, Krull J, Zembrzycki A, van den Akker WM, Gruss P, Simeone A, Mansouri A. *Sp8* controls the anteroposterior patterning at the midbrain-hindbrain border. *Development* 2006, 133:1779-1787.

Grossniklaus U, Pearson RK, Gehring WJ. The *Drosophila sloppy paired* locus encodes two proteins involved in segmentation that show homology to mammalian transcription factors. *Genes Dev* 1992, 6:1030-1051.

Grossniklaus U, Cadigan KM, Gehring WJ. Three maternal coordinate systems cooperate in the patterning of the *Drosophila* head. *Development* 1994, 120:3155-3171.

Hanson IM. Mammalian homologues of the *Drosophila* eye specification genes. *Cell and Dev Biol* 2001, 12:475-484.

Harrison SM, Houzelstein D, Dunwoodie SL, Beddington RS. *Sp5*, a new member of the *Sp1* family, is dynamically expressed during development and genetically interacts with *Brachyury*. *Dev Biol* 2000, 227:358-372.

Hersh BM, Carroll S. Direct regulation of *knot* gene expression by Ultrabithorax and the evolution of cis-regulatory elements in *Drosophila*. *Development* 2005, 132:1567-1577.

Hertveldt V, Louryan S, van Reeth T, Drèze P, van Vooren P, Szpirer J, Szpirer C. The development of several organs and appendages is impaired in mice lacking Sp6. *Dev Dyn* 2008, 237:883-892.

Hughes CL, Kaufman TC. Hox genes and the evolution of the arthropod body plan. *Evol and Dev* 2002, 4:459–499.

Hwang UW, Friedrich M, Tautz D, Park CJ, Kim W. Mitochondrial protein phylogeny joins myriapods with chelicerates. *Nature* 2001, 413:154-157.

Ingham PW, Martinez Arias A. Boundaries and fields in early embryos. *Cell* 1992, 68:221-235.

Kaback LA, Soung do Y, Naik A, Smith N, Schwarz EM, O'Keefe RJ, Drissi H. Osterix/Sp7 regulates mesenchymal stem cell mediated endochondral ossification. *J Cell Physiol* 2008, 214:173-182.

Kadonaga JT, Carner KR, Masiarz FR, Tjian R. Isolation of cDNA encoding transcription factor Sp1 and functional analysis of the DNA binding domain. *Cell* 1987, 51:1079-1090.

Kawakami Y, Esteban CR, Matsui T, Rodríguez-León J, Kato S, Izpisúa Belmonte JC. *Sp8* and *Sp9*, two closely related *buttonhead*-like transcription factors, regulate *Fgf8* expression and limb outgrowth in vertebrate embryos. *Development*. 2004, 131:4763-4774.

Liberg D, Sigvardsson M, Akerblad P. The EBF/Olf/Collier family of transcription factors: regulators of differentiation in cells originating from all three embryonal germ layers. *Mol and Cell Biol* 2002, 8389-8397.

Marin M, Karis A, Visser P, Grosveld F, Philipsen S. Transcription factor Sp1 is essential for early embryonic development but dispensable for cell growth and differentiation. *Cell* 1997, 89:619-267.

Mella S, Soula C, Morello D, Crozatier M, Vincent A. Expression patterns of the *coe/ebf* transcription factor genes during chicken and mouse limb development. *Gene Expression Patterns* 2004, 4:537-545.

Merrill VKL, Diederich RJ, Turner FR, Kaufman TC. A genetic and developmental analysis of mutations in *labial*, a gene necessary for proper head formation in *Drosophila melanogaster*. *Dev Biol* 1989, 135:376-391.

Milona MA, Gough JE, Edgar AJ. Expression of alternatively spliced isoforms of human Sp7 in osteoblast-like cells. *BMC Genomics* 2003, 4:43.

Mohler J, Seecoomar M, Agarwal S, Bier E, Hsai J. Activation of *knot* (*kn*) specifies the 3-4 intervein region in the *Drosophila* wing. *Development* 2000, 127:55-63.

Moorefield KS, Fry SJ, Horowitz JM. Sp2 DNA binding activity and trans-activation are negatively regulated in mammalian cells. *J Biol Chem* 2004, 279:13911-13924.

Nakamura T, Unda F, de-Vega S, Vilaxa A, Fukumoto S, Yamada KM, Yamada Y. The Krüppel-like factor epiprofin is expressed by epithelium of developing teeth, hair follicles, and limb buds and promotes cell proliferation. *J Biol Chem* 2004, 279:626-634.

Nakashima K, Zhou X, Kunkel G, Zhang Z, Deng JM, Behringer RR, de Crombrugghe B. The novel zinc finger-containing transcription factor Osterix is required for osteoblast differentiation and bone formation. *Cell* 2002, 108:17–29.

Nestoras K, Lee H, Mohler J. Role of *knot (kn)* in wing patterning in *Drosophila*. *Genetics* 1997, 147:1203-1212.

Pang K, Matus DQ, Martindale M.Q. The ancestral role of COE genes may have been in chemoreception: evidence from the development of the sea anemone, *Nematostella vectensis* (Phylum Cnidaria; Class Anthozoa). *Dev Genes Evol* 2004, 214:134-138.

Panganiban G, Irvine SM, Lowe C, Roehl H, Corley LS, Sherbon B, Grenier JK, Fallon JF, Kimble J, Walker M, Wray GA, Swalla BJ, Martindale MQ, Carroll SB. The origin and evolution of animal appendages. *PNAS* 1997, 94:5162-5166.

Pechmann M, McGregor AP, Schwager EE, Feitosa NM, Damen WGM. Dynamic gene expression is required for anterior regionalization in a spider. *PNAS* 2009, 106:1468-1472.

Philipsen S, Suske G. A tale of three fingers: the family of mammalian Sp/XKLF transcription factors. *Nucleic Acids Res* 1999, 27:2991-3000.

Posnien N, Bashasab F, Bucher G. The insect upper limb (labrum) is a nonsegmental appendage-like structure. *Evol and Dev* 2009a, 11:5,480-488.

Posnien N, Bucher G. The Hox1 ortholog Tc-*labial* is required for embryonic development of the intercalary segment giving rise to lateral parts of the head in Tribolium. *Dev Biol* 2009b, doi:10.1016/j.ydbio.2009.11.010.

Prasad BC, Ye B, Zackhary R, Schrader K, Seydoux G. *unc-3*, a gene required for axonal guidance in *Caenorhabditis elegans*, encodes a member of the O/E family of transcription factors. *Development* 1998, 125:1561-1568.

Putnam NH, Srivastava M, Hellsten U, Dirks B, Chapman J, Salamov A, Terry A, Shapiro H, Lindquist E, Kapitonov VV, Jurka J, Genikhovich G, Grigoriev IV, Lucas SM, Steele RE, Finnerty JR, Technau U, Martindale MQ, Rokhsar DS. Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. *Science* 2007, 317:86-94.

Rogers BT, Kaufman TC. Structure of the insect head in Ontogeny and Phylogenie: a view from *Drosophila*. *Int Rev of Cyt* 1997, 174:1-84.

Safe S, Abdelrahim M. Sp transcription factor family and its role in cancer. *Eur J Cancer* 2005, 41:2438-2448.

Saffer JD, Jackson SP, Annarella MB. Developmental expression of *Sp1* in the mouse. *Mol Cell Biol* 1991, 11:2189-2199.

Schaeper ND, Prpic NM, Wimmer EA. A conserved function of the zinc finger transcription factor *Sp8/9* in allometric appendage growth in the milkweed bug *Oncopeltus fasciatus*. *Dev Genes Evol* 2009 219:427-439.

Schinko JB, Kreuzer N, Offen N, Posnien N, Wimmer EA, Bucher G. Divergent functions of *orthodenticle, empty spiracles* and *buttonhead* in early head patterning of the beetle *Tribolium castaneum* (Coleoptera). *Dev Biol* 2008, 317:600-613.

Schmidt-Ott U, Technau GM. Expression of *en* and *wg* in the embryonic head and brain of *Drosophila* indicates a refolded band of seven segments remnants. *Development* 1992, 116:111-125.

Schöck F, Purnell BA, Wimmer EA, Jäckle H. Common and diverged functions of the *Drosophila* gene pair *D-Sp1* and *buttonhead*. *Mech Dev* 1999, 89:125-132.

Schöck F, Reischel J, Wimmer E, Taubert H, Purnell BA, Jäckle H. Phenotypic suppression of *empty spiracles* is prevented by *buttonhead*. *Nature* 2000, 405:351-354.

Scholtz G, Edgecombe GD. The evolution of arthropod heads: reconciling morphological, developmental and palaeontological evidence. *Dev Genes Evol* 2006, 210:395-415.

Schoppmeier M, Fischer S, Schmitt-Engel C, Löhr U, Klingler M. An ancient anterior patterning system promotes *caudal* represion and head formation in ecdysozoa. *Curr Biol* 2009, 19:1811-1815.

Schröder R. The genes *orthodenticle* and *hunchback* substitute for *bicoid* in the beetle *Tribolium*. *Nature* 2003, 422:621-625.

Scohy S, Gabant P, Van Reeth T, Hertveldt V, Drèze PL, Van Vooren P, Rivière M, Szpirer J, Szpirer C. Identification of KLF13 and KLF14 (SP6), novel members of the SP/XKLF transcription factor family. *Genomics* 2000, 70:93-101.

Seecoomar M, Agarwal S, Vani K, Yang G, Mohler J. *knot* is required for the hypopharyngeal lobe and its derivatives in the *Drosophila* embryo. *Mech Dev* 2000, 91:209-215.

Snodgrass RE. Evolution of Annelida, Onychophoran and Arthropoda. *Smithson Misc Coll* 1938, 138:1-77.

St Johnston D, Nüsslein-Volhard C. The origin of pattern and polarity in the *Drosophila* embryo. *Cell* 1992, 68:201-219.

Supp DM, Witte DP, Branford WW, Smith EP, Potter SS. Sp4, a member of the Sp1-family of zinc finger transcription factors, is required for normal murine growth, viability, and male fertility. *Dev Biol* 1996, 176:284-299.

Suske G. The Sp-family of transcription factors. *Gene* 1999, 238:291-300.

Suske G, Bruford E, Philipsen S. Mammalian SP/KLF transcription factors: bring in the family. *Genomics* 2005, 85:551-556.

Tallafuss A, Wilm TP, Crozatier M, Pfeffer P, Wassef M, Bally-Cuif L. The zebrafish buttonhead-like factor *Bts1* is an early regulator of *pax2.1* expression during mid-hindbrain development. *Development* 2001, 128:4021-4034.

Telford MJ, Thomas RH. Expression of homeobox genes shows chelicerate arthropods retain their deutocerebral segment. *Proc. Natl. Acad. Sci. USA* 1998, 95:10671-10675.

Treichel D, Becker MB, Gruss P. The novel transcription factor gene *Sp5* exhibits a dynamic and highly restricted expression pattern during mouse embryogenesis. *Mech Dev* 2001, 101:175-179.

Treichel D, Schöck F, Jäckle H, Gruss P, Mansouri A. *mBtd* is required to maintain signaling during murine limb development. *Genes Dev* 2003, 17:2630-2635.

Vervoort M, Crozatier M, Valle D, Vincent A. The COE transcription factor Collier is a mediator of short-range Hedghog-induced patterning of the *Drosophila* wing. *Curr Biol* 1999, 9:632-639.

Wimmer EA, Jäckle H, Pfeifle C, Cohen SM. A *Drosophil*a homologue of human Sp1 is a head-specific segmentation gene. *Nature* 1993, 366:690-694.

Wimmer EA, Frommer G, Purnell BA, Jäckle H. *buttonhead* and *D-Sp1*: a novel *Drosophila* gene pair. *Mech Dev* 1996, 59:53-62.

Wimmer EA, Cohen SM, Jäckle H, Desplan C. *buttonhead* does not contribute to a combinatorial code proposed for *Drosophila* head development. *Development* 1997, 124:1509-1517.

Zhao J, Cao Y, Zhao C, Postlethwait J, Meng A. An SP1-like transcription factor Spr2 acts downstream of Fgf signaling to mediate mesoderm induction. *EMBO J* 2003, 22:6078-6088.

Zhao C, Meng A. Sp1-like transcription factors are regulators of embryonic development in vertebrates. *Develop Growth Differ* 2005, 47:201–211.

6 Appendix

6.1 Abbreviations

A. gambiae (*Ag*)	*Anopheles gambiae*
A. mellifera (*Am*)	*Apis mellifera*
D. melanogaster (*Dm*)	*Drosophila melanogaster*
D. pulex (*Dp*)	*Daphnia pulex*
D. rerio (*Dr*)	*Danio rerio*
F. candida (*Fc*)	*Folsomia candida*
N. vectensis (*Nv*)	*Nematostella vectensis*
O. fasciatus (*Of*)	*Oncopeltus fasciatus*
P. hawaiensis (*Ph*)	*Parhyale hawaiensis*
T. castaneum (*Tc*)	*Tribolium castaneum*
T. adhaerens (*Ta*)	*Trichoplax adhaerens*
T. domestica (*Td*)	*Thermobia domestica*

dsRNA	double stranded RNA
Fig.	Figure
RNAi	RNA interference
Tab.	Table
WT	wild type

6.2 Sequences

All sequences are shown in EMBL format.

```
ID   FN562984; SV 1; linear; mRNA; STD; INV; 1800 BP.
XX
ST * private 31-DEC-2011
XX
AC   FN562984;
XX
DT   02-NOV-2009 (Rel. 102, Created)
DT   02-NOV-2009 (Rel. 102, Last updated, Version 0)
XX
DE   Oncopeltus fasciatus partial mRNA for Sp1-4 protein
XX
KW   .
XX
OS   Oncopeltus fasciatus (milkweed bug)
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera;
OC   Paraneoptera; Hemiptera; Euhemiptera; Heteroptera; Panheteroptera;
OC   Pentatomomorpha; Lygaeoidea; Lygaeidae; Lygaeinae; Oncopeltus.
XX
RN   [1]
RP   1-1800
RA   Schaeper N.D.;
RT   ;
RL   Submitted (02-NOV-2009) to the EMBL/GenBank/DDBJ databases.
RL   Schaeper N.D., Department of Developmental Biology, GZMB Ernst Caspari
RL   Haus, University of Goettingen, Blumenbach Institute, Justus-von-Liebig Weg
RL   11, Goettingen, 37077, GERMANY.
XX
RN   [2]
RA   Schaeper N.D., Prpic N.M., Wimmer E.A.;
RT   "A clustered set of three Sp-family genes is ancestral in the Metazoa:
RT   evidence from sequence analysis, protein domain structure, chromosomal
RT   location and developmental mRNA expression";
RL   Unpublished.
XX
FH   Key             Location/Qualifiers
FH
FT   source          1..1800
FT                   /organism="Oncopeltus fasciatus"
FT                   /mol_type="mRNA"
FT                   /dev_stage="0-4 days"
FT                   /tissue_type="embryo"
FT                   /db_xref="taxon:7536"
FT   CDS             <1..1701
FT                   /codon_start=1
FT                   /gene="sp1-4"
FT                   /product="Sp1-4 protein"
FT                   /protein_id="CBH30973.1"
FT                   /translation="VRLFFQERKGLPSPLAILAATCIKIEEQRISEESDSSEGLVANAE
FT                   THDPLSLSPPPQQPQVITLQQFQNLLTLQQVNGNEKYVPQATATTPTVSVQGLPGQFIQ
FT                   QSGGGGGGVTYSVLGGSQGGGMQAVTIDGQEAVFIPSNAGLMSPLRVLPTGVLPSSPLT
FT                   TAPTTAIFPGQNVSVRPDTQVLQVQQSIPVQVPISTSNGQTVYQTVHFPIQSFTSLPNI
FT                   LQTTVIPQISQVGTTQMAQMVTPSGQVQQVQVVTPASQQQTLVVQQQPTSSSASSTSTT
FT                   TTDSILVPHQQFLLGGVGGNQQQVTVIPVSSLANLTNGLTLPASSSVPQSVQQPIIQTI
FT                   PQGTIQIISSGSPGGGQDSSSGERWQVVGVQPAGSPPLEEEDQLDSKPRMRRVACSCPN
FT                   CIEGHAGCSRLRAEEATTSLEHVRTINKATWGGIRKKQHVCHVAGCNKVYGKTSHLRAH
FT                   LRWHTGERPFICSWEFCGKCFTRSDELQRHRRTHTGEKRFMCTQCHKKFMRSDHLSKHI
FT                   KTHQKARNDVTTSTQGDEETSGDEDKMLITLQAENEPSELSIVDTIDSM"
```

```
FT   3'UTR           1702..1800
FT                   /gene="sp1-4"
XX
SQ   Sequence 1800 BP; 515 A; 435 C; 394 G; 456 T; 0 other;
     gtccgcctct ttttccagga gcggaaaggc ttgccctccc ctttggctat tctagcggcg        60
     acctgcatca agatagagga gcagaggata tcagaggaga gtgattcaag tgagggactt       120
     gtggctaatg ctgagacaca tgacccccctg tctctgagcc caccgccaca acaacctcag      180
     gtaatcactc tccaacagtt tcaaaatctt ctaacccttc agcaagtgaa tggaaatgag       240
     aagtatgtgc ctcaagcaac tgccactacc ccaactgttt ctgtccaggg tctgcctgga       300
     cagtttatcc agcagtctgg tgggggtggc ggcggagtca cttactcagt tttgggtggc       360
     tcccaaggag gagggatgca ggcagtcacc attgatggac aggaagctgt tttcattccc       420
     tctaatgccg ggttaatgtc gccacttagg gtcctcccta ctggtgtcct cccttcttcg       480
     ccactcacta cagcacctac aactgctatc ttccctggcc aaaatgtttc tgttcggcca       540
     gacacccagg tgctacaggt gcaacaaagc attccagttc aagtgcccat ctccacttca       600
     aatgggcaga ctgtctatca gacggtccat tttcccattc agtctttcac aagtctcccc       660
     aatattttac aaacaactgt gataccccaa atctcccagg ttggaacaac acagatggcc       720
     caaatggtca cgccatcagg acaggttcag caagtccaag tagtaacacc agccagccag       780
     cagcaaactc ttgtcgtgca acaacagcct actagcagca gtgcctcctc cacttctacc       840
     acaacgactg atagtatatt agtgcctcat cagcaatttc tcctaggtgg tgttggtggc       900
     aaccaacaac aagtaactgt gatacctgtt tcaagcttag ctaacttgac caatggatta       960
     acgttaccag cgtcaagctc agtacctcaa tctgttcagc agcccatcat tcagactatt      1020
     cctcaaggaa ctatccagat aatttcaagt ggaagccctg gtggtggtca agacagtagt      1080
     agtggagaga gatggcaggt tgttggagtt cagcctgcag ggtccccacc attagaagag      1140
     gaggatcagc tggattccaa acctcgcatg aggagagtag cctgttcctg tcctaactgc      1200
     attgaaggcc atgctggttg ttctcgtctt cgggctgaag aagctacaac tagtcttgag      1260
     catgtgcgaa ccatcaataa agcaacttgg ggtggtataa gaaagaaaca gcatgtctgt      1320
     catgtagcgg gttgcaacaa agtctatgga aaaacatctc atctacgagc ccatcttaga      1380
     tggcacactg gtgagaggcc atttatttgc tcctgggaat tttgtgggaa gtgtttcacc      1440
     cgctctgatg aactacagag gcatcgtaga acacatacag gagagaaacg ttttatgtgt      1500
     actcagtgcc ataaaaaatt catgaggagt gatcaccttt ccaagcatat taaaactcat      1560
     cagaaagcta gaaatgatgt aacaacttca acgcaagggg acgaggaaac ttcaggggat      1620
     gaagataaaa tgctcataac cctccaagca gaaaatgaac cttcagagct ctccattgta      1680
     gatacaattg actctatgta gatttgtaat ttgttctatt ttttttttt ttagaaacta      1740
     ttcttgtaaa tattactggt aatctttgta ttaaaaaaaa aaaaaaaaaa aaaaaaaaaa      1800
//
```

```
ID   FN396612; SV 1; linear; mRNA; STD; INV; 1213 BP.
XX
AC   FN396612;
XX
DT   06-NOV-2009 (Rel. 102, Created)
DT   06-NOV-2009 (Rel. 102, Last updated, Version 1)
XX
DE   Oncopeltus fasciatus partial mRNA for Sp8/9 protein
XX
KW   .
XX
OS   Oncopeltus fasciatus (milkweed bug)
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera;
OC   Paraneoptera; Hemiptera; Euhemiptera; Heteroptera; Panheteroptera;
OC   Pentatomomorpha; Lygaeoidea; Lygaeidae; Lygaeinae; Oncopeltus.
XX
RN   [1]
RP   1-1213
RA   Schaeper N.D.;
RT   ;
RL   Submitted (29-MAY-2009) to the EMBL/GenBank/DDBJ databases.
RL   Schaeper N.D., Department of Developmental Biology, GZMB Ernst Caspari
RL   Haus, University of Goettingen, Blumenbach Institute, Justus-von-Liebig Weg
RL   11, Goettingen, 37077, GERMANY.
XX
RN   [2]
RA   Schaeper N.D., Prpic N.M., Wimmer E.A.;
RT   "A conserved function of the zinc finger transcription factor Sp8/9 in
RT   allometric appendage growth in the milkweed bug Oncopeltus fasciatus";
RL   Dev. Genes Evol. 219:427-435(2009).
XX
FH   Key             Location/Qualifiers
FH
FT   source          1..1213
FT                   /organism="Oncopeltus fasciatus"
FT                   /mol_type="mRNA"
FT                   /tissue_type="embryos"
FT                   /db_xref="taxon:7536"
FT   CDS             182..>1213
FT                   /gene="sp8/9"
FT                   /product="Sp8/9 protein"
FT                   /protein_id="CAZ39568.1"
FT                   /translation="MLTDVTSRVPAGPLSQVTKPMVGAEKYHLDDHLRKEHFLCFKVAE
FT                   LNSPSSRKTCASDTWPSEQEIMEHPGLRDTPLAMLAAQCNKLSSKSPPPLADAAVGKGF
FT                   HPWKKSPQSRDPPTSSYSRPPVSSCSYTSSGELWTGTADIAKVETVYSRHPYDGWGLPT
FT                   HHVKPENAGWWEVPGPGGWLDMAGHGQVQYGPEYGLGLASNSLLQGQHLLQDYKMLPGQ
FT                   TGAFLHAPPAPAPAPQAPSPRSQRRYTGRATCDCPNCQEAERLGPAGVHLRKKNIHSCH
FT                   IPGCGKVYGKTSHLKAHLRWHTGERPFVCNWLFCGKRFTRSDELQRHLRTHTGEKRFAC
FT                   PVCN"
```

```
XX
SQ   Sequence 1213 BP; 278 A; 382 C; 322 G; 231 T; 0 other;
     acgcggggag agtgtgatct tcccagtccg cgacagccgg ccttcctata ctcaggatag        60
     tctgtctcct actgccaaca ttcgcccatg tagatcacta acaggctgtt cgcacttttt       120
     ttttccccct ttctgcgata tcttcacttc gctcggccag attcagtgac caactcaaga       180
     aatgttgacg gacgtcacgt cgagagtccc agccgggcca ctatcacaag tcaccaaacc       240
     gatggttggt gcagaaaaat atcatcttga tgatcattta cgtaaagaac atttcctctg       300
     cttcaaggtg gcagaactga actcgccaag ttccagaaaa acatgtgcct ccgacacctg       360
     gccctcagaa caagaaatca tggagcaccc aggtctacga gacacaccat tagcgatgct       420
     agccgcgcag tgcaacaagt tgtcttctaa gtcaccacct cccctagcgg atgctgcagt       480
     gggcaaagga tttcaccctt ggaagaaaag tccacagtcc cgcgatccgc cgacctcgag       540
     ttacagcagg cctccggtca gcagttgctc ttacacctca tcgggggagc tgtggacggg       600
     gaccgccgac atagctaaag tggagacggt ctacagcagg cacccttacg atggctgggg       660
     gttgccgaca catcacgtaa agccggaaaa cgccggctgg tgggaagtgc ccggaccggg       720
     cggctggctg gatatggcgg gccacgggca ggtgcaatac gggcccgaat acggactggg       780
     cctcgcttcg aacagcctcc tccaagggca gcacctcctt caggactaca aaatgctgcc       840
     cggacagacg ggggccttcc ttcacgctcc tccggccccg gccccagcgc cccaggcccc       900
     gtctccccgc tctcaacggc gatacacggg ccgcgcgaca tgcgactgcc ctaactgcca       960
     ggaggccgag cgactcggtc ccgcaggagt ccatctccgc aagaagaaca tccatagctg      1020
     ccacatacct ggctgcggta aggtctacgg taagacctcg cacctcaagg ctcacctgag      1080
     gtggcacact ggagagcggc cgttcgtctg caactggctg ttctgcggga aacggttcac      1140
     gaggtctgat gaactgcagc gccaccttcg tacccataca ggagagaaaa ggttcgcctg      1200
     tcccgtatgc aac                                                        1213
//
```

```
ID   FN562985; SV 1; linear; mRNA; STD; INV; 1365 BP.
XX
ST * private 31-DEC-2011
XX
AC   FN562985;
XX
DT   02-NOV-2009 (Rel. 102, Created)
DT   02-NOV-2009 (Rel. 102, Last updated, Version 0)
XX
DE   Folsomia candida partial mRNA for Sp1-4 protein
XX
KW   .
XX
OS   Folsomia candida
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Collembola; Arthropleona;
OC   Entomobryoidea; Isotomidae; Proisotominae; Folsomia.
XX
RN   [1]
RP   1-1365
RA   Schaeper N.D.;
RT   ;
RL   Submitted (02-NOV-2009) to the EMBL/GenBank/DDBJ databases.
RL   Schaeper N.D., Department of Developmental Biology, GZMB Ernst Caspari
RL   Haus, University of Goettingen, Blumenbach Institute, Justus-von-Liebig Weg
RL   11, Goettingen, 37077, GERMANY.
XX
RN   [2]
RA   Schaeper N.D., Prpic N.M., Wimmer E.A.;
RT   "A clustered set of three Sp-family genes is ancestral in the Metazoa:
RT   evidence from sequence analysis, protein domain structure, chromosomal
RT   location and developmental mRNA expression";
RL   Unpublished.
XX
FH   Key             Location/Qualifiers
FH
FT   source          1..1365
FT                   /organism="Folsomia candida"
FT                   /mol_type="mRNA"
FT                   /dev_stage="0-5 days"
FT                   /tissue_type="embryo"
FT                   /db_xref="taxon:158441"
FT   CDS             <1..1020
FT                   /codon_start=1
FT                   /gene="sp1-4"
FT                   /product="Sp1-4 protein"
FT                   /protein_id="CBH30974.1"
FT                   /translation="WYQRRVRGINQAALGQSFGQFQNIQNIPGLGNVPILPAGTITLGQ
FT                   GQQNVTIQQQSSPNQQIIQAVQAPQIITGLGNLGNLGQMGGAQLQQDPLDPTKWQVINQ
FT                   APLMVTTSLGNAGTVTAIKTELDKPKTRLRRVACTCPNCTDIDASSNNSRNNGEKKKQH
FT                   LCHIPGCGKVYGKTSHLRAHLRWHAGERPFVCQWLFCGKRFTRSDELQRHRKTHTGEKR
FT                   FQCPECLKKFMRSDHLSKHIKTHSKNRSVTGADGSLINSTKLELDDDDEKLIKTEVETS
FT                   DAMCSDDEDDTGEMGDDGEVDGDGDLDDQENAEDMMITINTEGEQPELISNDSIENQA"
```

```
FT   3'UTR           1021..1365
FT                   /gene="sp1-4"
XX
SQ   Sequence 1365 BP; 482 A; 254 C; 300 G; 329 T; 0 other;
     tggtatcaac gcagagtacg cgggattaac caagctgcac ttggccagtc gttcggacaa        60
     ttccaaaaca ttcaaaacat tcccggattg ggaaacgttc cgattttacc agccggaaca       120
     atcacattag gtcaaggaca acaaaatgtg accattcagc aacaaagttc tccaaatcag       180
     caaattatcc aagctgttca agcaccacag attatcactg gcctaggaaa tctagggaac       240
     ctgggacaaa tgggtggagc tcaacttcaa caggatccac ttgatcccac gaaatggcaa       300
     gtaataaacc aagctcccct tatggttacc acgtcgttgg gaaatgcagg gactgttact       360
     gcaattaaaa cagaactcga taagccaaag acaagactaa ggagagttgc atgcacctgt       420
     ccaaactgta ccgatattga tgccagttcg aataattcca gaaataatgg agaaaagaaa       480
     aagcagcatt tgtgtcatat tcctggttgt ggtaaagttt atggtaaaac atcacatcta       540
     cgcgctcacc ttagatggca tgcaggagaa aggccatttg tctgtcaatg gttgttttgt       600
     gggaaacggt ttacacgatc cgatgaactc caaaggcata gaaagacgca cacgggagag       660
     aaacgattcc agtgtcctga atgtctcaaa aagtttatga ggagcgatca cctttctaaa       720
     catataaaaa ctcattcaaa gaatagaagc gtgacaggag cggacggatc tcttattaat       780
     tcgacaaagt tggaactaga cgatgatgac gaaaaactta ttaaaacgga agttgaaact       840
     tctgacgcga tgtgtagtga tgatgaggac gatactggtg aaatgggaga cgatggggaa       900
     gttgatgggg atggtgacct cgatgaccaa gaaaatgctg aggacatgat gatcacaatt       960
     aacacggaag gagaacagcc ggaactaata agtaatgact cgattgaaaa tcaagcatga      1020
     aatcatctgc gaaaaaggag gaggggagga gcttcattga attattgaat tcaagaccac      1080
     ccatcatcat ccgaataaag caaaatatcc acaaacaaag cgtcgccgga cctcactgac      1140
     aaatttatgc aaatgagttt ctttcaagag tacaagctgg acagacgcag agaaaggtct      1200
     agtttcataa taggatataa atgatagaat atagttttga ttttttccta atttcgtgtc      1260
     gactagaaat atagtagttt tattaagctg taaatttacg tatcaacttt aatacaacaa      1320
     taaataatat taatacagtt aagagagtca gtctgtatta gatct                      1365
//
```

```
ID   FN562986; SV 1; linear; mRNA; STD; INV; 2339 BP.
XX
ST * private 31-DEC-2011
XX
AC   FN562986;
XX
DT   02-NOV-2009 (Rel. 102, Created)
DT   02-NOV-2009 (Rel. 102, Last updated, Version 0)
XX
DE   Folsomia candida mRNA for Sp5/Btd protein
XX
KW   .
XX
OS   Folsomia candida
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Collembola; Arthropleona;
OC   Entomobryoidea; Isotomidae; Proisotominae; Folsomia.
XX
RN   [1]
RP   1-2339
RA   Schaeper N.D.;
RT   ;
RL   Submitted (02-NOV-2009) to the EMBL/GenBank/DDBJ databases.
RL   Schaeper N.D., Department of Developmental Biology, GZMB Ernst Caspari
RL   Haus, University of Goettingen, Blumenbach Institute, Justus-von-Liebig Weg
RL   11, Goettingen, 37077, GERMANY.
XX
RN   [2]
RA   Schaeper N.D., Prpic N.M., Wimmer E.A.;
RT   "A clustered set of three Sp-family genes is ancestral in the Metazoa:
RT   evidence from sequence analysis, protein domain structure, chromosomal
RT   location and developmental mRNA expression";
RL   Unpublished.
XX
FH   Key             Location/Qualifiers
FH
FT   source          1..2339
FT                   /organism="Folsomia candida"
FT                   /mol_type="mRNA"
FT                   /dev_stage="0-5 days"
FT                   /tissue_type="embryo"
FT                   /db_xref="taxon:158441"
FT   5'UTR           1..243
FT                   /gene="sp5/btd"
FT   CDS             244..1935
FT                   /gene="sp5/btd"
FT                   /product="Sp5/Btd protein"
FT                   /protein_id="CBH30975.1"
FT                   /translation="MMEGIFVNGFHHHQYGSLQSHQFNHLPSYPGSAYHPQHHSHHPSF
FT                   CNMNIPMSMSTLQQHQGGIGGRSLESNFHSTNVGGIGGNGGVTPKDHYPWWSTTASPSQ
FT                   SSQSQLQQSPPMSTSGNHSSFTPSMFFSPSPSSVAAAVASMLPTSSQRKSRRCRCPNCL
FT                   AGIQPTQVPGEKKKKRQHICHMIGCGKIYGKTSHLKAHLRGHTGERPFVCNWLFCGKRF
FT                   TRSDELQRHLRTHTGEKRFICNVCQKRFMRSDHLAKHVKTHNNPSSSRNKGQQSSKMSD
FT                   SQKSKSVNNTTTATFDSILPEVTSTTRTPPLLSSSSSSSPLSTASSAGEHKDLSGLLYY
FT                   DRQPGKGTFAGRFAMGNSIQSDIKNSMGILAENNNGNSGSRSRVVDTGENIHYSLLSPS
FT                   SSTLSPASSSSSTCSSTRSSSSPSSKRTIVTPSSSPSSLELTPVKSPSSPQAQAQQTSS
FT                   FHSNQFPLHPFQPHFPASYYDNNGVVNQSVCNPGNNMSNGSNNAHFTSYPPLHHGSYSS
FT                   SVAVKACSPVNSGFQGTNVNFQNNFSSHYQLHQQHLHPGMNSLFGM"
```

```
FT   3'UTR           1936..2339
FT                   /gene="sp5/btd"
XX
SQ   Sequence 2339 BP; 710 A; 575 C; 416 G; 638 T; 0 other;
     gaatcgtttt tctcagctcg ttagaccagc ttctagtact tactacttgg cttggttgac        60
     aacttctcat taattacagc aaaatctcac ctctaaagag cagttttaca aaataataaa       120
     agtttcttgc caagtattta atacttacgt tgagcacatg aaccagttat aaagttcatt       180
     ctcaagccga agtaaagtgc gtgtacgttg attgattgga ataattgtcg atacattccg       240
     tccatgatgg aagggatctt cgtgaatggg tttcatcatc accagtacgg atcattgcaa       300
     tctcatcagt ttaaccatct gccctcgtac cctgggagtg cataccaccc acaacatcat       360
     tcccaccatc catcattttg caatatgaat attccaatgt ccatgagcac ccttcagcaa       420
     catcaaggcg gcattggtgg cagaagcctg gagtccaact ttcactccac caacgttggg       480
     ggaattggtg gcaatggtgg cgtcacaccc aaggaccatt acccttggtg gtcgacgact       540
     gcctcccccc cacagtcgtc acaatcccaa ttgcagcagt caccaccaat gtcaacttcc       600
     gggaaccatt cgtccttcac cccctccatg ttcttctccc cgtctccctc gtccgtggcg       660
     gctgccgtgg cctccatgct ccccacctcc agtcagagaa agtcaaggag atgtcgatgt       720
     ccaaattgcc tcgcaggaat tcaacctact caagttccgg gtgaaaagaa aaagaagcga       780
     cagcacattt gtcacatgat tggatgtgga aagatctatg ggaaaacatc tcatcttaaa       840
     gcccacttac gcggacacac gggagagcgt ccgttcgttt gcaattggct gttttgtgga       900
     aaacgcttca cacgttccga tgaactgcag cggcacttga ggacgcacac gggcgaaaag       960
     agattcatat gcaatgtatg ccaaaaacgg ttcatgcgca gtgatcatct cgcaaaacat      1020
     gtcaaaaccc acaacaaccc ctcctcatct cggaacaaag ggcagcaaag tagcaagatg      1080
     agtgacagcc agaaatcaaa gtcagtcaat aataccacaa ccgcaacatt tgactcaatt      1140
     ttgcccgaag taacgtcaac aacaagaaca ccaccattat tgtcatcctc ttcgtcatca      1200
     tccccattat ctacagcctc atctgcaggt gagcacaaag acctgtccgg actgctctac      1260
     tacgacagac agcctggaaa gggaactttt gctggaagat ttgcgatggg caacagcatt      1320
     cagtcagata ttaaaaattc aatgggcatc ttagcagaga ataataatgg caacagtggc      1380
     agcagaagcc gagtggttga cactggtgaa aatattcact attcactcct gtcaccgtct      1440
     tcttccacct tgtcacctgc atcgtcttcc tcatcaacct gttcttccac acggtcgtcc      1500
     tcaagtcctt catcaaagcg gactattgtt accccgtcat cttctccgtc atcattggag      1560
     ttgactcctg tcaaatctcc gtcttctcct caggcacaag cacagcagac gtcgtccttc      1620
     cactcaaatc aatttcctct tcatcccttt caacctcatt tccctgcctc gtattacgat      1680
     aacaatggtg tcgttaacca aagcgtttgt aatcccggta ataatatgag taatggaagc      1740
     aataatgcac actttacatc ctacccacca cttcatcacg ggtcctattc ctcatccgtc      1800
     gcagtgaaag cttgttcccc agttaatagt ggcttccaag gaacaaatgt caactttcag      1860
     aacaatttct cctcccatta tcaacttcat cagcaacatt tgcatccggg aatgaatagc      1920
     ttatttggga tgtagaatgc ataaaagatg agaattgaga cagatacata cattcttttg      1980
     atgcgcatgg tagcgaaaga tcagattgaa tcacacactt gccatatttt tattattatt      2040
     attcaaaaag attgattttc agctatgcat atattgaaat taatgttaat aatatttgtg      2100
     gtttttgaaa catgctcaat catagtggct gcaagtaaaa acaaattatt taagatttta      2160
     tgtatattac attactagaa ttatacatat tttgagcatt attttcagta atagccttcg      2220
     taaataacta tatactgcta gcagttttat caatatattt ttagtgagta taaaaaaata      2280
     attaatgaca atccataaac acaagtaatc aaaaaaaaaa aaaaaaaaaa aaaaaaaaa       2339
//
```

```
ID   FN562987; SV 1; linear; mRNA; STD; INV; 1342 BP.
XX
ST * private 31-DEC-2011
XX
AC   FN562987;
XX
DT   02-NOV-2009 (Rel. 102, Created)
DT   02-NOV-2009 (Rel. 102, Last updated, Version 0)
XX
DE   Folsomia candida partial mRNA for Sp6-9 protein
XX
KW   .
XX
OS   Folsomia candida
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Collembola; Arthropleona;
OC   Entomobryoidea; Isotomidae; Proisotominae; Folsomia.
XX
RN   [1]
RP   1-1342
RA   Schaeper N.D.;
RT   ;
RL   Submitted (02-NOV-2009) to the EMBL/GenBank/DDBJ databases.
RL   Schaeper N.D., Department of Developmental Biology, GZMB Ernst Caspari
RL   Haus, University of Goettingen, Blumenbach Institute, Justus-von-Liebig Weg
RL   11, Goettingen, 37077, GERMANY.
XX
RN   [2]
RA   Schaeper N.D., Prpic N.M., Wimmer E.A.;
RT   "A clustered set of three Sp-family genes is ancestral in the Metazoa:
RT   evidence from sequence analysis, protein domain structure, chromosomal
RT   location and developmental mRNA expression";
RL   Unpublished.
XX
FH   Key             Location/Qualifiers
FH
FT   source          1..1342
FT                   /organism="Folsomia candida"
FT                   /mol_type="mRNA"
FT                   /dev_stage="0-5 days"
FT                   /tissue_type="embryo"
FT                   /db_xref="taxon:158441"
FT   CDS             <1..1086
FT                   /codon_start=1
FT                   /gene="sp6-9"
FT                   /product="Sp6-9 protein"
FT                   /protein_id="CBH30976.1"
FT                   /translation="GGNGVMTNPMTHSSLLSAAQKSLSSSVCVTSSNHNAYSRTPVSSS
FT                   CGATPTYGADLYFHGSSGGNPCGDQSPVGSVGVYSRVPHDSWPFNAMSGAHAGIKPDVS
FT                   GMNGSSWWDAVHGATGGSWLDVGPGSAISQTAGMHHPHHPHHQMSQYNDYSSLTHTLSN
FT                   PHILTSGQHLLRVPSPRSQRRYTGRATCDCPNCQEAERLGPAGAHLRKKNIHSCHIPGC
FT                   GKVYGKTSHLKAHLRWHTGERPFVCNWLFCGKRFTRSDELQRHLRTHTGEKRFACPICN
FT                   KRFMRSDHLSKHVKTHNNSSGSSTGNKRSGSEDDDEEDNSQSEANSNHSPGSSVTTPPL
FT                   PPHAMSPSVHGLTQQDIKPQI"
```

```
FT   3'UTR           1087..1342
FT                   /gene="sp6-9"
XX
SQ   Sequence 1342 BP; 341 A; 395 C; 298 G; 308 T; 0 other;
     gggggaaacg gtgtgatgac caaccccatg acccacagct cccttctcag cgctgctcaa        60
     aagtctctct cgtcctccgt ctgtgtcacg tcctccaatc acaacgccta cagcagaaca       120
     cctgtctcct cttcctgcgg ggccacaccc acgtacgggg ctgatctcta ctttcacggc       180
     tcaagtggtg gcaacccatg cggggaccaa agtccggtgg gctccgtggg cgtctattca       240
     agagttcctc atgattcctg gcccttcaat gcaatgtctg gggcacatgc aggaattaag       300
     cctgatgtgt cagggatgaa tggttcatcg tggtgggatg ccgttcacgg ggctacgggt       360
     ggcagctggt tggacgttgg accggggagt gcaatctcac aaacggcagg gatgcaccac       420
     ccacatcatc ctcaccacca aatgtcccaa tacaatgact actcctccct cactcacacc       480
     ctctccaacc ctcacatctt gacctctggt cagcaccttc tgcgggtgcc ttccccccgt       540
     tcgcaaagga ggtacaccgg tagagcaact tgtgattgtc caaattgtca agaggctgag       600
     cggctgggcc cagcaggagc tcaccttagg aaaaagaata ttcactcctg tcacatccct       660
     ggctgtggaa aggtatacgg gaaaacatca catcttaagg cccacctcag atggcacacc       720
     ggtgagagac cttttgtatg caattggctg ttttgtggaa aacgcttcac acgctccgat       780
     gaactgcaac gacacttgcg gacgcatact ggagagaaga gattcgcgtg cccaatctgt       840
     aataagcgct ttatgaggtc ggatcaccta tcaaaacatg tcaaaactca caacaactcg       900
     tcggggagca gtacgggcaa caagcggagt gggagtgaag atgacgatga ggaggataat       960
     agtcaatccg aggcgaattc gaaccacagt cctggctcct ctgtgaccac tccacctctt      1020
     cctcctcacg ccatgtcacc ctcagtgcat gggttgacac aacaagacat caaaccccaa      1080
     atatgaacct gtctccgtac ttaccacccc gagtgaaagt cgtcctcctt gcatctgatc      1140
     tctccaccac caccgcgaca catgctgaag tagacgttgt actctaatcg taattctcaa      1200
     acacatttta accacacttc tattttaata cactatcaaa catttcgcat ccaatcaacc      1260
     aacccaataa tagctcacta acaattaaat gggatccttc atcacgtttc atggtcactt      1320
     ttctacttgt gcgatctagg ta                                              1342
//
```

```
ID   FN562988; SV 1; linear; mRNA; STD; INV; 697 BP.
XX
ST * private 31-DEC-2011
XX
AC   FN562988;
XX
DT   02-NOV-2009 (Rel. 102, Created)
DT   02-NOV-2009 (Rel. 102, Last updated, Version 0)
XX
DE   Thermobia domestica partial mRNA for Sp1-4 protein
XX
KW   .
XX
OS   Thermobia domestica (firebrat)
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Thysanura; Lepismatidae;
OC   Thermobia.
XX
RN   [1]
RP   1-697
RA   Schaeper N.D.;
RT   ;
RL   Submitted (02-NOV-2009) to the EMBL/GenBank/DDBJ databases.
RL   Schaeper N.D., Department of Developmental Biology, GZMB Ernst Caspari
RL   Haus, University of Goettingen, Blumenbach Institute, Justus-von-Liebig Weg
RL   11, Goettingen, 37077, GERMANY.
XX
RN   [2]
RA   Schaeper N.D., Prpic N.M., Wimmer E.A.;
RT   "A clustered set of three Sp-family genes is ancestral in the Metazoa:
RT   evidence from sequence analysis, protein domain structure, chromosomal
RT   location and developmental mRNA expression";
RL   Unpublished.
XX
FH   Key             Location/Qualifiers
FH
FT   source          1..697
FT                   /organism="Thermobia domestica"
FT                   /mol_type="mRNA"
FT                   /dev_stage="0-4 days"
FT                   /tissue_type="embryo"
FT                   /db_xref="taxon:89055"
FT   CDS             <1..384
FT                   /codon_start=1
FT                   /gene="sp1-4"
FT                   /product="Sp1-4 protein"
FT                   /protein_id="CBH30977.1"
FT                   /translation="NKVYGKTSHLRAHLRWHTGERPFVCTWMFCGKRFTRSDELQRHRR
FT                   THTGEKRFQCPECSKKFMRSDHLSKHIKTHQKQRIMETATSTHSGSSDASNSTDDKMMI
FT                   TIQTEADQSDLAINETLESSLSN"
FT   3'UTR           385..697
FT                   /gene="sp1-4"
XX
SQ   Sequence 697 BP; 249 A; 116 C; 135 G; 197 T; 0 other;
     aataaggtat atggaaagac atcacatttg agagctcact tacggtggca cacgggagaa        60
     cgacctttg tctgtacctg gatgttttgc gggaaacgat tcacacgatc agatgaattg       120
     cagcggcaca gaagaacaca tactggcgag aaaaggttcc agtgtcctga atgtagcaag       180
     aaatttatga gaagcgacca cttatctaaa catatcaaaa cgcatcagaa acagaggata       240
     atggaaactg ctacttcgac gcattctgga tcgtctgatg ccagcaattc aacagacgat       300
     aaaatgatga tcacaattca aaccgaagca gatcaatccg accttgctat caatgaaaca       360
     ttggaatcat cactgtccaa ctgagtgaaa tagaatcaga acaaacagaa ggaggtttgt       420
     aggcgggata aattattttt atatacttac tgagaaagct tgctccagcc agtgttcaaa       480
     aatttaaata ctaatagtat ggtactagtt tttctgttgc catatgtact aaagtagtac       540
     ggtgaacaca aacgtacatg catttactat ttggttttgt ttttatactt acgaccattt       600
     agtttagtga tgtttggttt gcttttcatt ctacaaattt tataattatt gatttatgtt       660
     gtcgtatgga aaaaaaaaaa aaaaaaaaaa aaaaaa                                 697
//
```

```
ID   FN562989; SV 1; linear; mRNA; STD; INV; 1453 BP.
XX
ST * private 31-DEC-2011
XX
AC   FN562989;
XX
DT   02-NOV-2009 (Rel. 102, Created)
DT   02-NOV-2009 (Rel. 102, Last updated, Version 0)
XX
DE   Thermobia domestica mRNA for Sp5/Btd protein
XX
KW   .
XX
OS   Thermobia domestica (firebrat)
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Thysanura; Lepismatidae;
OC   Thermobia.
XX
RN   [1]
RP   1-1453
RA   Schaeper N.D.;
RT   ;
RL   Submitted (02-NOV-2009) to the EMBL/GenBank/DDBJ databases.
RL   Schaeper N.D., Department of Developmental Biology, GZMB Ernst Caspari
RL   Haus, University of Goettingen, Blumenbach Institute, Justus-von-Liebig Weg
RL   11, Goettingen, 37077, GERMANY.
XX
RN   [2]
RA   Schaeper N.D., Prpic N.M., Wimmer E.A.;
RT   "A clustered set of three Sp-family genes is ancestral in the Metazoa:
RT   evidence from sequence analysis, protein domain structure, chromosomal
RT   location and developmental mRNA expression";
RL   Unpublished.
XX
FH   Key             Location/Qualifiers
FH
FT   source          1..1453
FT                   /organism="Thermobia domestica"
FT                   /mol_type="mRNA"
FT                   /dev_stage="0-4 days"
FT                   /tissue_type="embryo"
FT                   /db_xref="taxon:89055"
FT   5'UTR           1..82
FT                   /gene="sp5/btd"
FT   CDS             83..1285
FT                   /gene="sp5/btd"
FT                   /product="Sp5/Btd protein"
FT                   /protein_id="CBH30978.1"
FT                   /translation="MASLLPNPGIAYHQNQSNPLTMLAATCSRLGSQVPTSTDFPRTST
FT                   SHFQTRNYHNQQTYPEVTLQSIGFQQNSSGFHHDTALSASPQVPTGNTTNSAPYNLSIH
FT                   QYHSTSAVAPTSAQSTAGILPSSPVIAQSPPPLEASDQQQQIQDRNSRMAWWFHQASTA
FT                   KQPVATSSPPAFTNSYFFAHHQAAQTDQTLYHQQRFAAAAAMLKSSAAVASRRCSRCRC
FT                   PNCQNGGNTAGQQPQRKRQHVCHVPGCSKTYGKTSHLKAHLRWHTGERPFVCNWLFCGK
FT                   SFTRSDELQRHLRTHTGEKRFACPVCNKRFMRSDHLAKHVKTHDANRKLSGGRGFKSPS
FT                   NKGNKESGEGKNKIDDFVTSTDSFSEDAFVTTEDSVTSSTKENYYHQLHFHQLCSMYGH
FT                   V"
```

```
FT   3'UTR           1286..1453
FT                   /gene="sp5/btd"
XX
SQ   Sequence 1453 BP; 443 A; 358 C; 280 G; 372 T; 0 other;
     gcttccgttg aggttgttgt cctgctgtat taaccctcac taaagggaaa gcagtggtat        60
     caacgcagag tacgcgggcc gtatggcatc gctactcccg aatccaggga ttgcttatca       120
     tcaaaaccag agtaatcctt tgacaatgtt ggcagcaaca tgcagtcgtc taggatctca       180
     ggtgccaacc tcaacagatt ttcccagaac gtccacatct catttccaga cacggaacta       240
     ccataatcaa cagacatatc ctgaagtcac gttacaaagc attggtttcc agcagaactc       300
     atcaggtttc caccatgata cggctctttc agctagtcca caagtcccta ctggtaatac       360
     cacaaactct gcaccgtata atttatcaat acatcaatat cattccacat ctgctgttgc       420
     cccaactagt gctcaatcca ctgctggaat tcttccgtca tctccggtga tagcacaatc       480
     acctccacca ctggaggctt ctgatcagca acagcaaatt caggaccgta attccaggat       540
     ggcttggtgg tttcatcagg catcaactgc aaaacaaccg gttgccactt cttcgccgcc       600
     agctttcacc aactcgtatt tctttgctca tcatcaggca gctcaaacag accaaacctt       660
     gtatcaccaa cagcggtttg ctgctgcagc cgcaatgctc aagtcttctg cagcagttgc       720
     atcaaggcgc tgcagtcgtt gtcgatgtcc aaactgtcaa aatggaggca atacagcagg       780
     acaacaacct caacggaagc gtcaacatgt ctgccacgtc cctgggtgct caaagacgta       840
     cggaaaaaca tcacatctga aagctcattt gcgctggcac acgggagaac gtccatttgt       900
     ttgcaattgg ctttttgtg gtaaaagttt cacaagatca gacgaattac aacgccattt       960
     gcggactcac actggtgaga agcgatttgc ttgtcccgtt tgcaacaaaa ggttcatgcg      1020
     atcggatcat ctcgcgaaac atgtcaaaac tcacgacgca aatagaaaat tgtccggcgg      1080
     tcggggtttt aagtccccca gcaataaggg taacaaagaa agtggtgaag gtaaaaataa      1140
     gattgacgat tttgtgacct caacagactc cttttcagaa gatgcttttg tgaccacaga      1200
     agacagcgta acatcctcaa ccaaagaaaa ctattaccat cagctgcatt ttcatcaatt      1260
     gtgttctatg tacggtcatg tgtagaaatt ttaaaactca tatcaatttc caaagattag      1320
     tttttaaata cagaaagacc atggaattat ttcatcaatt tctaatcttg tgccccggtt      1380
     tgctatgatg tacaagatat aacacccgaa attcatttaa gtgtatgaaa aaaaaaaaaa      1440
     aaaaaaaaaa aaa                                                         1453
//
```

```
ID   FN562990; SV 1; linear; mRNA; STD; INV; 1210 BP.
XX
ST * private 31-DEC-2011
XX
AC   FN562990;
XX
DT   02-NOV-2009 (Rel. 102, Created)
DT   02-NOV-2009 (Rel. 102, Last updated, Version 0)
XX
DE   Thermobia domestica partial mRNA for Sp6-9 protein
XX
KW   .
XX
OS   Thermobia domestica (firebrat)
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Thysanura; Lepismatidae;
OC   Thermobia.
XX
RN   [1]
RP   1-1210
RA   Schaeper N.D.;
RT   ;
RL   Submitted (02-NOV-2009) to the EMBL/GenBank/DDBJ databases.
RL   Schaeper N.D., Department of Developmental Biology, GZMB Ernst Caspari
RL   Haus, University of Goettingen, Blumenbach Institute, Justus-von-Liebig Weg
RL   11, Goettingen, 37077, GERMANY.
XX
RN   [2]
RA   Schaeper N.D., Prpic N.M., Wimmer E.A.;
RT   "A clustered set of three Sp-family genes is ancestral in the Metazoa:
RT   evidence from sequence analysis, protein domain structure, chromosomal
RT   location and developmental mRNA expression";
RL   Unpublished.
XX
FH   Key             Location/Qualifiers
FH
FT   source          1..1210
FT                   /organism="Thermobia domestica"
FT                   /mol_type="mRNA"
FT                   /dev_stage="0-4 days"
FT                   /tissue_type="embryo"
FT                   /db_xref="taxon:89055"
FT   CDS             <1..853
FT                   /codon_start=2
FT                   /gene="sp6-9"
FT                   /product="Sp6-9 protein"
FT                   /protein_id="CBH30979.1"
FT                   /translation="VNSLNTTSWWDMHSAGAGSWLDMSGATASGLSGMTGMHATQMANY
FT                   VGSDYSTLSHSLAASNTAHLLSSGQHLLQDTYKSMLPSQGVGASVASFGLSQPSLPQVP
FT                   SPRSQRRYTGRATCDCPNCQEAERLGPAGVHLRKKNIHSCHIPGCGKVYGKTSHLKAHL
FT                   RWHTGERPFVCNWLFCGKRFTRSDELQRHLRTHTGEKRFACPVCNKRFMRSDHLAKHVK
FT                   THNGNGSKKGSSESCSDSENSQSEPPSVSSPASVNTPPLAQGTPLSDASVASSPVDVKP
FT                   VV"
```

```
FT   3'UTR           854..1210
FT                   /gene="sp6-9"
XX
SQ   Sequence 1210 BP; 320 A; 282 C; 297 G; 311 T; 0 other;
     agttaatagt ctaaatacaa cgtcgtggtg ggacatgcac tcagccggcg ctggttcatg        60
     gctcgatatg agtggagcca cggcgtcggg tctttccgga atgacaggaa tgcacgctac       120
     tcaaatggct aactacgttg gttcggatta ttcgacgctc agccattctt tagcagcatc       180
     taatacagcc cacttactgt cttcaggtca gcatctgttg caagatactt acaagtcaat       240
     gttaccatcg cagggagtcg gtgcgtcagt ggcgtctttt ggtctcagcc agccttctct       300
     ccctcaggta ccgtctccga ggtcccagag gaggtataca ggaagagcta cttgtgattg       360
     tcccaattgt caggaagcag aacgcttagg gcctgctggt gtccatctta gaaagaagaa       420
     tatccatagt tgtcacatac caggttgtgg taaggtatac gggaagacat cacatttaaa       480
     agctcatctt cgttggcaca caggagaacg tccattcgtt tgtaattggc tattctgtgg       540
     taagcgtttc acacgatccg acgaattaca acgtcacttg cgaacacaca ccggggagaa       600
     gaggttcgct tgtcccgttt gtaacaaacg gttcatgcgt tccgatcatt tggcaaaaca       660
     cgtcaagaca cataatggga acggcagcaa aaaagggagc tctgagagct gcagcgactc       720
     tgaaaacagc caatcagagc ctcccagtgt gagctcccca gcgtccgtta atacgccgcc       780
     cctcgctcaa gggactccat tgtccgacgc ctccgttgcg agctctccag tcgacgtgaa       840
     gcccgtcgtc tgaggaaggt ggaccccggc gcttgcggca ggggtaccga gtttgaacct       900
     tttctttcat ccgtcacttg gtcaagtaag atccgaccag caagcgacag atggacattg       960
     atggttctct ttcacgggaa cagaagaaaa ggaagggtca tattcgctcg tcggggatcg      1020
     aaccgcgatc ctcgagtcac ggtgctctcg tcaagtaaag tgtttaactt tcatagataa      1080
     tattgagtta ttaacaaaat atgtgaatgt agcaggtata aagtttctac gtcaatcaaa      1140
     atttggttaa tagtttatgg taaaatttat tcagaatata aagtatcatg atctgcaaaa      1200
     atatattttt                                                            1210
//
```

```
ID   FN562991; SV 1; linear; mRNA; STD; INV; 3018 BP.
XX
ST * private 31-DEC-2011
XX
AC   FN562991;
XX
DT   02-NOV-2009 (Rel. 102, Created)
DT   02-NOV-2009 (Rel. 102, Last updated, Version 0)
XX
DE   Parhyale hawaiensis mRNA for Sp1-4 protein
XX
KW   .
XX
OS   Parhyale hawaiensis
OC   Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca; Eumalacostraca;
OC   Peracarida; Amphipoda; Gammaridea; Talitroidea; Hyalidae; Parhyale.
XX
RN   [1]
RP   1-3018
RA   Schaeper N.D.;
RT   ;
RL   Submitted (02-NOV-2009) to the EMBL/GenBank/DDBJ databases.
RL   Schaeper N.D., Department of Developmental Biology, GZMB Ernst Caspari
RL   Haus, University of Goettingen, Blumenbach Institute, Justus-von-Liebig Weg
RL   11, Goettingen, 37077, GERMANY.
XX
RN   [2]
RA   Schaeper N.D., Prpic N.M., Wimmer E.A.;
RT   "A clustered set of three Sp-family genes is ancestral in the Metazoa:
RT   evidence from sequence analysis, protein domain structure, chromosomal
RT   location and developmental mRNA expression";
RL   Unpublished.
XX
FH   Key             Location/Qualifiers
FH
FT   source          1..3018
FT                   /organism="Parhyale hawaiensis"
FT                   /mol_type="mRNA"
FT                   /dev_stage="0-12 days"
FT                   /tissue_type="embryo"
FT                   /db_xref="taxon:317513"
FT   5'UTR           1..127
FT                   /gene="sp1-4"
FT   CDS             128..2527
FT                   /gene="sp1-4"
FT                   /product="Sp1-4 protein"
FT                   /protein_id="CBH30980.1"
FT                   /translation="MKQENTVQQPQAVSQAAPISVSAGDVTPGQIIAANGQTYSIMPQA
FT                   AVQMQTITVDGQEILYIPSGVPAAQAQQQMAVQLGGATQAVLTPSGQLVRAASVPSLLQ
FT                   GTAAIQQQQQQQQQQQQQQQQQQQQQLQQQQLQQLQQQLQLQLQQQQQQTTGAAGQLG
FT                   SQIIQLGMGGMGQTIPVQVPVSGSNGSSVLQTIPLQLPFVQNMMQAAGTGAATAAQGIQ
FT                   VIPQIAQMPAQPQLAHIVLPNGQLQQVQIMAAAAPAGLAAAGGGGVVSLAQLAGGSGAV
FT                   ITAGSGVVSGTTGGVTWTTTATTPSLTVPGNPTASPPQEDKQHVEAAVAAANGVQQQQQ
FT                   IISAASVATHAQVNNSNGAAVSSQASQQQQQVAAAVARNTTNSNQQQQQQQQQQQQQQQ
FT                   QQQQQQAAALAAQQQVAAAAAAAAAGGSIQMPSLQVMQPGQLFHHIPGIGPVQVVNATN
FT                   LSSLAAASMPISLATIQQTAAAAAAAAGAAPGTILAAQPQTTATAGAAAAAAPLQGHLL
FT                   PGGAQIISLPGDGVDAASKWGIFSPSAQIVSPMPAVASSPSTTANAQQAQQQQQQQQQQ
FT                   QQQQQQQQQSNQLMVTNVATALAGVPATADPLQLDDDLAEPRTRLKRVACTCPNCREGG
FT                   VNNERGEPSSSNGIGGSTKRRQHICHIPGCNKTYGKTSHLRAHLRCDAGDRPFACNWMF
FT                   CTKRFTRSDELQRHKRTHTGEKRFHCPDCDKKFMRSDHLSKHVKTHIAKKTTVPSVVES
FT                   SDSSGHSSDNGGDSKMLITLPSEATSDSPTIVTSAVTPTEESTTTQ"
```

```
FT   3'UTR           2528..3018
FT                   /gene="sp1-4"
XX
SQ   Sequence 3018 BP; 768 A; 849 C; 752 G; 649 T; 0 other;
     gaacgggcga cgggagcgca tgtatttaca cacgagcagt ttcgttttac gttattattg        60
     tgtggcaagt gtttcgttca gctaaatgct gtgaaataac aacttgttgc aaggagcata       120
     tttcacaatg aagcaagaga acacagttca gcagcctcaa gcagttagtc aggctgctcc       180
     catctctgtc tcagctggag atgtcacacc aggccagatt attgcggcca atggccagac       240
     ctactccatc atgccacagg cagctgttca gatgcagacc attacagttg atggccagga       300
     aattctgtac ataccttcag gtgtgccagc agctcaagcc cagcaacaga tggctgtgca       360
     acttggagga gcaacacagg ctgttctgac tcccagtggt cagttggtac gggctgcatc       420
     agtgccatcc ctcttgcagg gtacagctgc tattcagcag cagcaacaac agcaacaaca       480
     acaacagcag cagcagcaac agcaacaaca gcagcaacag ttgcaacaac agcagcttca       540
     acagttacag cagcagttgc agttgcaact ccagcaacaa cagcagcaaa caactggagc       600
     tgctggacag ttgggatcgc agatcataca gctcggcatg ggtggcatgg ggcagacaat       660
     acccgtgcaa gttcctgtgt cgggtagcaa tggtagcagt gtgctgcaga ccatccccct       720
     gcagctacct tttgtgcaga acatgatgca ggccgcgggt actggtgctg ctactgctgc       780
     tcaaggcatt caggttatcc cacagattgc ccagatgcct gcacagcccc agctggccca       840
     tattgtgttg cctaacggcc agctgcagca agtgcagatc atggctgctg cagcgccagc       900
     tgggctcgct gcagcaggag gaggtggtgt ggtgtcccta gcacagctcg caggcggcag       960
     tggtgcagtg atcacagcag gcagcggtgt agtgagtggc accactggag gtgtcacgtg      1020
     gaccacaacc gccaccacac cgtctctcac tgtgccgggc aaccctactg cctcaccgcc      1080
     tcaagaagac aagcagcacg tggaggcggc tgtggctgct gccaatggtg tgcagcagca      1140
     acagcagatc atttctgcag catctgtagc cactcatgct caggtgaaca atagtaatgg      1200
     agcagcggtg agcagtcagg cgtctcaaca gcagcagcag gtagctgccg ctgtcgcgag      1260
     aaacaccacc aacagcaatc agcagcaaca acagcagcag cagcaacaac aacaacaaca      1320
     gcagcagcag caacaacaac aagcagcagc attagctgcg caacaacagg ttgcggcagc      1380
     ggctgcggca gctgcagcag gaggcagtat acagatgccc tcgttacaag ttatgcaacc      1440
     tggtcaacta ttccatcaca tccccggcat cggtcccgtg caagtggtaa atgcaacgaa      1500
     tctatcgtcc cttgcggcgg ccagcatgcc catatcgctc gccaccatac aacagacggc      1560
     agcagctgca gctgccgctg cgggggctgc accgggcacc atactggccg cccagccaca      1620
     gacaacggct actgcaggag ctgctgccgc tgcagcaccc ctgcagggcc accttctgcc      1680
     tggtggagcc caaattatca gtttacccgg tgatggtgtc gacgctgcca gcaagtgggg      1740
     cattttctct ccgtcagctc agatcgtgtc acccatgcct gctgtcgcca gctctccctc      1800
     caccaccgcc aatgctcaac aagctcagca acagcagcaa caacaacagc agcagcaaca      1860
     acagcaacag cagcaacaac aatctaacca gttaatggta actaatgttg ctacggccct      1920
     cgcgggtgtg cctgccactg ccgatccctt acagctggac gacgaccttg cagagcctcg      1980
     caccagacta aagagagtcg cgtgcacgtg tcccaactgt cgggagggtg gagtcaataa      2040
     cgagagaggc gagcctagca gcagtaatgg cattggtggc tctactaagc ggcgtcagca      2100
     catttgtcac attcctggtt gcaataagac ctacggcaaa acttctcatc tgcgggcgca      2160
     cttgagatgc gatgctggtg acagaccttt tgcgtgcaat tggatgtttt gcacaaagcg      2220
     ctttacacga tctgacgagt tacagagaca caagagaact cacacgggcg agaagcgttt      2280
     ccactgtcct gactgcgaca aaaaattcat gcggtcggat catctttcta agcacgtcaa      2340
     gacgcatatc gccaagaaaa ctacggtgcc gtcggtggta gagagctcag acagcagcgg      2400
     ccactcttct gacaatggcg gagatagcaa gatgctcatt accctgccca gtgaagccac      2460
     cagcgactcc cccaccatcg tcaccagcgc cgtcactcct acggaagagt ccaccaccac      2520
     gcagtagaag atcctccctt ctacttctat ccaattctag tgctacaccc tgcctcctgg      2580
     ttgtgatcgt atgtccgtgg gcctctaggt ttcagggaac tgtcgggcag tatgtagtac      2640
     ctagtatatg tattatgtgt aataatgtca tgcctatcga ttacaagctt atcaagaagc      2700
     gtctgtatta catgtttcca gtgttgcccc ccccggtctc agcactattt tattttcgcg      2760
     atctcgcttc aaatcaaatt gcatgaatta tgtttgtgtt tttcttgttg ttacacgcta      2820
     ctggcagctc agcggaaata tcccgatttt attttcggca ttcactgaac aatggacatc      2880
     gtttggaaaa cgtgcctcac tgcctagtac catttaattt tatatcaatg ctaataattt      2940
     ggttcttcat cgatctttat ttttacaaaa tgatgattcg ccagcgaact cttcgactgt      3000
     cgtcttattt aaaaggat                                                   3018
//
```

```
ID   FN562992; SV 1; linear; mRNA; STD; INV; 1502 BP.
XX
ST * private 31-DEC-2011
XX
AC   FN562992;
XX
DT   02-NOV-2009 (Rel. 102, Created)
DT   02-NOV-2009 (Rel. 102, Last updated, Version 0)
XX
DE   Parhyale hawaiensis mRNA for Sp6-9 protein
XX
KW   .
XX
OS   Parhyale hawaiensis
OC   Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca; Eumalacostraca;
OC   Peracarida; Amphipoda; Gammaridea; Talitroidea; Hyalidae; Parhyale.
XX
RN   [1]
RP   1-1502
RA   Schaeper N.D.;
RT   ;
RL   Submitted (02-NOV-2009) to the EMBL/GenBank/DDBJ databases.
RL   Schaeper N.D., Department of Developmental Biology, GZMB Ernst Caspari
RL   Haus, University of Goettingen, Blumenbach Institute, Justus-von-Liebig Weg
RL   11, Goettingen, 37077, GERMANY.
XX
RN   [2]
RA   Schaeper N.D., Prpic N.M., Wimmer E.A.;
RT   "A clustered set of three Sp-family genes is ancestral in the Metazoa:
RT   evidence from sequence analysis, protein domain structure, chromosomal
RT   location and developmental mRNA expression";
RL   Unpublished.
XX
FH   Key             Location/Qualifiers
FH
FT   source          1..1502
FT                   /organism="Parhyale hawaiensis"
FT                   /mol_type="mRNA"
FT                   /dev_stage="0-12 days"
FT                   /tissue_type="embryo"
FT                   /db_xref="taxon:317513"
FT   5'UTR           1..48
FT                   /gene="sp6-9"
FT   CDS             49..1431
FT                   /gene="sp6-9"
FT                   /product="Sp6-9 protein"
FT                   /protein_id="CBH30981.1"
FT                   /translation="MMLTDLNLAGRTELYPTDKGNSLIHSAADPQHITEHPSLRGTPLA
FT                   MLAAQCNRITTKSPPPLADAAVGKGFHPWKKSPGSTTPVPASPTTAPQRPPSAASTNST
FT                   TANYQRTPTTTTAACSYPQPSDFYLTPATTSGGQDLLTKTELPAMSSMYSRVHPYDSWP
FT                   FNSVASAASGIKPEVSAVNTFNTASSLWDVHSAAGSWLDMSGAAASGWSGVAGMSQMNY
FT                   PAGDYSSLSHTLAASNTAHLLSSGQHLFQDTYKSMLPGQGMGGGMGAPFTLGQPSLPQV
FT                   PSPRSQRRYTGRATCDCPNCQEADRLGPAGAHLRKRNIHSCHIPGCGKVYGKTSHLKAH
FT                   LRWHTGERPFVCNWLFCGKRFTRSDELQRHLRTHTGEKRFACPVCNKRFMRSDHLSKHV
FT                   KTHNACGGGGGGQGKKSSDGSEDSDNSHSEAASASSPASISSPPAPPTPTSPSTPDLKP
FT                   LV"
```

```
FT   3'UTR           1432..1502
FT                   /gene="sp6-9"
XX
SQ   Sequence 1502 BP; 369 A; 439 C; 365 G; 329 T; 0 other;
     cgagaagaac gaacggatcc agtcaaagaa gagagttgga actgaccgat gatgttgacg        60
     gatctgaacc ttgctggccg gacggagctc tacccgaccg ataaagggaa ctcccttatc       120
     catagcgcag cagaccctca acatataacg gaacacccat ctcttcgcgg cacccctcta       180
     gccatgctcg ccgcccagtg taacagaata accaccaaaa gtccgcctcc tttggcggac       240
     gctgctgtcg gcaaaggctt ccatccatgg aagaagagtc caggttccac tactcctgtt       300
     ccagccagcc ccacgacggc gcctcaaagg ccgccttctg cagcctcgac aaattccaca       360
     acagcaaatt atcaacggac tccgactaca accacagcgg cttgcagtta tccccagccg       420
     tcagattttt atctaacacc ggcgacgact tctggtggac aagacctgct caccaaaacc       480
     gaacttccag caatgtcttc tatgtattcc cgagtgcatc cctacgattc ttggcctttc       540
     aattcagtag cgtccgctgc ttctggaata aagccagaag tgtctgcggt aaatacattt       600
     aacacggcga gctcgttgtg ggacgtacac agcgctgctg gatcttggtt agatatgtct       660
     ggggcggcgg catcgggctg gtctggagta gccggaatga gtcaaatgaa ctatccggca       720
     ggtgattatt catctctatc acacacatta gctgcctcaa acacagccca tcttttatca       780
     tcgggtcagc acttatttca agacacgtac aagtctatgc tgccggggca agggatggga       840
     ggaggaatgg gtgccccatt cacactgggt cagccatcgt tacctcaagt accatcaccg       900
     aggtcgcagc gtcgatacac gggccgggcg acgtgtgact gtcccaactg tcaggaagcg       960
     gatcgattgg gtccagccgg tgcacatctg cgcaagagga atatccactc gtgtcacatt      1020
     ccgggctgcg gcaaagttta cggcaagact tcacatctca aggctcatct tcgctggcac      1080
     actggtgaga gacccttgt ttgcaactgg ctattttgtg gcaaaaggtt cacaaggtct       1140
     gatgaactac aacgacacct tcgcacgcac acaggtgaga agcggtttgc ctgcccggtg      1200
     tgcaacaagc gattcatgcg ttccgaccac ctgagcaagc atgtgaagac gcataacgca      1260
     tgtggcggcg gcggtggtgg tcagggcaag aaaagcagtg atggaagcga agacagcgac      1320
     aacagccata gcgaagctgc ttctgcttcc tcccctgcct ctatcagctc tcctcctgcc      1380
     ccccctactc ccacctctcc ttccaccccc gatctcaaac ctctggtgtg agcgatgatc      1440
     gatcgatcga tccttcgatc attcatttct atattgatat atagtcactg agcaccaaca      1500
     tg                                                                    1502
//
```

```
ID   FN557065; SV 1; linear; mRNA; STD; INV; 516 BP.
XX
ST * private 31-DEC-2011
XX
AC   FN557065;
XX
DT   21-OCT-2009 (Rel. 102, Created)
DT   21-OCT-2009 (Rel. 102, Last updated, Version 0)
XX
DE   Oncopeltus fasciatus partial mRNA for collier protein (col gene)
XX
KW   .
XX
OS   Oncopeltus fasciatus (milkweed bug)
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera;
OC   Paraneoptera; Hemiptera; Euhemiptera; Heteroptera; Panheteroptera;
OC   Pentatomomorpha; Lygaeoidea; Lygaeidae; Lygaeinae; Oncopeltus.
XX
RN   [1]
RP   1-516
RA   Schaeper N.D.;
RT   ;
RL   Submitted (19-OCT-2009) to the EMBL/GenBank/DDBJ databases.
RL   Schaeper N.D., Department of Developmental Biology, GZMB Ernst Caspari
RL    Haus, University of Goettingen, Blumenbach Institute, Justus-von-Liebig
Weg
RL   11, Goettingen, 37077, GERMANY.
XX
RN   [2]
RA   Schaeper N.D., Pechmann M., Damen W.G.M., Prpic N.M., Wimmer E.A.;
RT   "Evolution of collier function in head development in insects and other
RT   arthropods";
RL   Unpublished.
XX
FH   Key             Location/Qualifiers
FH
FT   source          1..516
FT                   /organism="Oncopeltus fasciatus"
FT                   /mol_type="mRNA"
FT                   /tissue_type="embryos aged 0-4 days"
FT                   /db_xref="taxon:7536"
FT   CDS             <1..>516
FT                   /codon_start=1
FT                   /gene="col"
FT                   /product="collier protein"
FT                   /protein_id="CBH20112.1"
FT                   /translation="PSNLRKSNFFHFVVALYDRAGQPIEIERTAFIGFIEKDQENETQK
FT                  TNNGIQYRLQLLYANGVRQEQDIFVRLIDSVTKQAIICEGQDKNPEMCRVLLTHEVMCS
FT                  RCCDKKSCGNRNETPSDPVIIDRFFLKFFLKCNQNCLKNAGNPRDMRRFQVVISTMVSV
FT                  EGPLLAISD"
XX
SQ   Sequence 516 BP; 144 A; 139 C; 123 G; 110 T; 0 other;
     ccctcgaatc tgaggaagtc caacttcttc cacttcgttg tcgccctcta cgaccgagcg        60
     ggacagccga tagagatcga gaggactgcg ttcatcggct tcatcgagaa ggaccaggaa       120
     aacgaaaccc aaaaaaccaa taacggaatc cagtaccgtc tccaactcct ttacgccaat       180
     ggtgttcgac aggagcaaga cattttgta cgtttgatag actcagtaac gaaacaggca        240
     atcatctgcg agggtcaaga caaaaatccc gagatgtgca gagtacttct gacccacgag       300
     gtgatgtgca gccgatgctg cgacaagaag agctgtggta accggaacga gactccctcc       360
     gacccggtaa ttatcgacag gttcttccta aagttcttcc ttaaatgcaa ccaaaactgc       420
     cttaagaacg ccggtaaccc cagagacatg aggaggtttc aggtggtgat ctcgaccatg       480
     gtctctgttg aaggtcctct gctcgccatc tccgac                                 516
//
```

```
ID   FN557064; SV 1; linear; mRNA; STD; INV; 1883 BP.
XX
ST * private 31-DEC-2011
XX
AC   FN557064;
XX
DT   21-OCT-2009 (Rel. 102, Created)
DT   21-OCT-2009 (Rel. 102, Last updated, Version 0)
XX
DE   Parhyale hawaiensis mRNA for collier protein (col gene)
XX
KW   .
XX
OS   Parhyale hawaiensis
OC   Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca; Eumalacostraca;
OC   Peracarida; Amphipoda; Gammaridea; Talitroidea; Hyalidae; Parhyale.
XX
RN   [1]
RP   1-1883
RA   Schaeper N.D.;
RT   ;
RL   Submitted (19-OCT-2009) to the EMBL/GenBank/DDBJ databases.
RL   Schaeper N.D., Department of Developmental Biology, GZMB Ernst Caspari
RL    Haus, University of Goettingen, Blumenbach Institute, Justus-von-Liebig
Weg
RL   11, Goettingen, 37077, GERMANY.
XX
RN   [2]
RA   Schaeper N.D., Pechmann M., Damen W.G.M., Prpic N.M., Wimmer E.A.;
RT   "Evolution of collier function in head development in insects and other
RT   arthropods";
RL   Unpublished.
XX
FH   Key             Location/Qualifiers
FH
FT   source          1..1883
FT                   /organism="Parhyale hawaiensis"
FT                   /mol_type="mRNA"
FT                   /tissue_type="embryonic tissue of all stages"
FT                   /db_xref="taxon:317513"
FT   5'UTR           1..166
FT                   /gene="col"
FT   CDS             167..1672
FT                   /codon_start=1
FT                   /gene="col"
FT                   /product="collier protein"
FT                   /protein_id="CBH20111.1"
FT                   /translation="MFNLGVDATGLHPTHGSHIPHHPHHASSIGHPHITSSLTSLKEEP
FT                   IISSAASVVTSSTTTSPSVATSTPTSISSATSRVWMQPNMLDTSAHHAHQQSPGLGRAH
FT                   YEKQPPSNLRKSNFFHFVLAFYDRTGQPIEIERTAFVGFIEKESEQDGQKTNNGIHYRL
FT                   QLLYSNGVREERDFYVRLVDASTKQAIAYEGQDKNPEMCRVLLTHEVMCSRCCDKKSCG
FT                   NRNETPSDPVIIDRFFLKFFLKCNQNCLKNAGNPRDMRRFQVVISTQVDVSGPLLAVSD
FT                   NMFVHNNSKHGRRAKRIDTADAMFGLYPPLPTPCVKAICPSEGWTSGGATVVIVGDNFF
FT                   DGLQVVFGTTLVWSELITPHAIRMQTPPRPVPGVVEITLSYKSKQFCKGSPGRFVYVSL
FT                   TDPTIEYGFHRLHKLIPRHPGDPEKLPKEIILKRAADLAEALYTMPRTNQLSLPAPRSP
FT                   ATMNQSAAAAAMTGFNSYTGQLVSVQDTAANAAAAQWGEGFFF"
FT   3'UTR           1673..1883
FT                   /gene="col"
```

```
XX
SQ   Sequence 1883 BP; 517 A; 531 C; 429 G; 406 T; 0 other;
     ggattttacc aatcacccct ttaccaagcg tttactgtga agtgtaatta caaaaatcca        60
     aagtgctttc aggtgaacca caaacttgta aatttactga gcccttatca cccgacgttc       120
     cttttagata ttgaatgaat gtacttgaaa attattcatc catgtcatgt ttaaccttgg       180
     agtagacgcc acggggcttc atccaacgca tggtagccat attccacatc atcctcacca       240
     cgcatcctca ataggtcatc cccacatcac ctcatctctc acatccctaa aagaagaacc       300
     catcatctcc tcagcagcat ctgtggtcac ctcctctacc accacatctc catctgtggc       360
     cacatccacg cctacatcca tcagttctgc cacttcgcga gtgtggatgc aaccaaacat       420
     gttggataca tccgcccatc atgcacatca acagagcccc ggcctaggac gagcgcacta       480
     cgagaaacag ccgcccagca acctgcgcaa gagcaacttc ttccactttg tcctggcgtt       540
     ttacgacaga acgggacagc caattgaaat agagaggact gccttcgttg gcttcataga       600
     aaaagagtcg gaacaggacg gtcagaaaac aaacaatggc atccattaca gattacaact       660
     gctctactca aacggtgtac gagaagagcg agatttttat gtacgattgg ttgacgcctc       720
     caccaaacag gcaattgcgt acgaaggcca ggataagaat ccggagatgt gtcgggtgct       780
     gctcacacat gaagttatgt gcagtcggtg ttgtgataag aaaagctgcg gaaaccggaa       840
     tgaaactcca tcagaccccg tcatcatcga tagattttc ctgaagttct tcttaaaatg       900
     caaccagaac tgcctcaaga acgccggcaa cccacgagat atgcgacggt ttcaggtggt       960
     aatatcaaca caagtagatg tgagcggtcc gctactcgca gtgtccgaca acatgtttgt      1020
     acacaacaac agcaagcatg gacgcagagc caagcgaatc gacaccgccg acgccatgtt      1080
     cggtctgtac cctccactac caacgccttg tgttaaggcg atctgtccgt cggaaggttg      1140
     gacttcaggt ggtgctacgg tcgtcattgt gggggataat ttcttcgatg gtctccaggt      1200
     tgtgttcggc actacactcg tgtggagtga gctcatcacg ccacacgcca tcaggatgca      1260
     aacgccccca cggcctgtgc ctggggtggt ggagatcacc ctttcataca agagtaaaca      1320
     attctgcaaa ggatcccctg gcagatttgt ttacgtctcc ctgaccgatc ccactattga      1380
     gtacggtttc caccggctac acaagcttat tcctcgccat ccaggagacc ctgagaagct      1440
     accaaaggaa attattctca agcgagcggc agacctagcg gaggcgctgt acacaatgcc      1500
     ccgcaccaac caattatcgc tgcccgcgcc gcgctcaccc gccacaatga accaatcagc      1560
     ggccgccgct gcaatgacgg gcttcaacag ctacacaggc cagctcgtga gcgtgcagga      1620
     taccgccgcc aacgctgctg cagctcagtg gggtgaaggt tttttcttct agacgacaga      1680
     ggtaacaaca acgagagcag cgggagctcg tccagcagcg gcagcagcga gctccctggc      1740
     gacgcgatcc ccatacccaa caacaataat aataacgagg agaacaatag gatcggcact      1800
     tctaataaag gacccgtcat cggcaaggat gatagtctct gcagaatatc tagtcttaaa      1860
     gaagaaatcg cggcagcgag tgc                                              1883
//
```

```
ID   FN557066; SV 1; linear; mRNA; STD; INV; 839 BP.
XX
ST * private 31-DEC-2011
XX
AC   FN557066;
XX
DT   21-OCT-2009 (Rel. 102, Created)
DT   21-OCT-2009 (Rel. 102, Last updated, Version 0)
XX
DE   Achaearanea tepidariorum partial mRNA for collier protein (col gene)
XX
KW   .
XX
OS   Parasteatoda tepidariorum (common house spider)
OC   Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
OC   Araneomorphae; Entelegynae; Araneoidea; Theridiidae; Parasteatoda.
XX
RN   [1]
RP   1-839
RA   Schaeper N.D.;
RT   ;
RL   Submitted (19-OCT-2009) to the EMBL/GenBank/DDBJ databases.
RL   Schaeper N.D., Department of Developmental Biology, GZMB Ernst Caspari
RL    Haus, University of Goettingen, Blumenbach Institute, Justus-von-Liebig
Weg
RL   11, Goettingen, 37077, GERMANY.
XX
RN   [2]
RA   Schaeper N.D., Pechmann M., Damen W.G.M., Prpic N.M., Wimmer E.A.;
RT   "Evolution of collier function in head development in insects and other
RT   arthropods";
RL   Unpublished.
XX
FH   Key             Location/Qualifiers
FH
FT   source          1..839
FT                   /organism="Achaearanea tepidariorum"
FT                   /mol_type="mRNA"
FT                   /tissue_type="embryonic tissue of all stages"
FT                   /db_xref="taxon:114398"
FT   5'UTR           1..240
FT                   /gene="col"
FT   CDS             241..>839
FT                   /codon_start=1
FT                   /gene="col"
FT                   /product="collier protein"
FT                   /protein_id="CBH20113.1"
FT                   /translation="MMFGLQDPGLQPRGPMANIKEEPLTSAQLGSVRTGWMQPTMIDQT
FT                  AGSVGLGRAHFEKQPPNNPRKSNFFHFVIALYDKAGQPVEVERTAFVGFIEKDQEEEGQ
FT                  KTNNSVHYRLQLLFANGVRQEQDLYVRLIDHQRKKAVPYEGQDKNPEMCRVLLTHEVMC
FT                  SRCCDKKSCGNRNETPSDPVIIDRFFLKFFLKCNQN"
```

```
XX
SQ   Sequence 839 BP; 258 A; 185 C; 209 G; 187 T; 0 other;
     cattgctaag ttgttgctag tggctgtaaa caataagtcg atcgccgtac acggccaaac        60
     atcgagtgga ggattgcaaa agtgcgaacc gagcaccgtg tgacgtcacc atccccaaca       120
     cgtggtccaa agctgaagag ggcgggcgct gagaggactc actaaagatc tcattgttgt       180
     tggaaagtcc caggcgctcg cttgatctat cgccgtgagg cttacgacaa acgagttaag       240
     atgatgttcg ggttgcaaga tcctggcctc cagcccaggg ggcccatggc caatataaaa       300
     gaagaacctc tcacatctgc ccaactgggc agtgttcgaa cgggatggat gcaaccgacg       360
     atgattgacc agacagccgg aagtgttggt ctaggtcgag cacattttga aaaacagccc       420
     ccaaataacc ccagaaaaag caatttttt cattttgtca ttgcattgta tgataaagct       480
     ggtcagcccg ttgaagttga aagaacagct tttgtgggtt ttatagaaaa agaccaggag       540
     gaagaaggtc aaaaaacaaa taatagtgtt cactacagac tacagctgct gttcgcaaat       600
     ggagtgcggc aggagcaaga cctatatgtt cgactaatag accatcaaag aaaaaaggcc       660
     gtcccttacg aaggccaaga caaaaacccc gaaatgtgcc gagttctttt aacacatgaa       720
     gtaatgtgca gtcggtgctg cgacaagaag agttgcggaa acaggaacga aacccccttca       780
     gatcctgtga taattgacag atttttcctc aagttcttct tgaagtgcaa tcagaattg        839
//
```

Danksagung

Zuerst möchte ich mich bei Herrn Prof. Dr. Ernst A. Wimmer dafür bedanken, dass ich in der Abteilung Entwicklungsbiologie des Johann-Friedrich-Blumenbach-Instituts für Zoologie, Anthropologie und Entwicklungsbiologie meine Doktorarbeit anfertigen konnte. Ich danke ihm für das entgegengebrachte Vertrauen absolut selbständig arbeiten zu dürfen, dass ich immer wieder neue Organismen für meine Arbeit im Labor etablieren konnte, für wissenschaftliche Anregungen, sowie für die Möglichkeit an vielen nationalen und internationalen Tagungen teil zu nehmen.

Herrn Jun. Prof. Dr. Gregor Bucher danke ich für die Übernahme des Korreferats meiner Arbeit, sowie die stete Diskussionsbereitschaft und Unterstützung in Bezug auf die *Tribolium* Arbeit.

Einen besonderen Dank möchte ich auch an Herrn Prof. Dr. Nipam H. Patel richten, der mich so offen und freundlich in seinem Labor an der UC Berkeley aufgenommen hat, sowie an alle Mitarbeiter der Patel-Arbeitsgruppe. Insbesondere möchte mich bei Dr. Paul Z. Liu bedanken, der mir während meiner Zeit in Berkeley und auch danach mit Rat und Tat in Sachen *Oncopeltus*-Arbeit zur Seite gestanden ist. Vielen Dank auch an das Education Abroad Program (EAP) und die Fulbright Kommission, die mir diesen unvergesslichen Aufenthalt in Kalifornien ermöglicht haben. Bei Herrn Prof. Dr. Ullrich Technau und seiner Arbeitsgruppe in Wien bedanke ich mich dafür, dass ich in ihrer Arbeitsgruppe die in situ Hybridisierung an der Seeanemone *Nematostella vectensis* durchführen konnte.

Meinen Kollegen der gesamten Abteilung Entwicklungsbiologie möchte ich ganz herzlich für die angenehme Arbeitsatmosphäre und stete Hilfsbereitschaft danken. Danken möchte ich insbesondere Marco Winkler für stete Unterstützung beim Käfer zusammen sammeln, Inga Karch für den unermüdlichen Einsatz bei den *Tribolium* RNA Injektionen während ihres GP2 Praktikums, sowie Anna Stief für Assistenz bei der Etablierung des *Folsomia* Fixierungs- und in situ-Protokolls. Besonderer Dank gilt auch Sara Khadjeh, Matthias Pechmann und Daniela Hüber für die vielen fachbezogenen und auch fachfremden Diskussionen und Anregungen.

Ganz besonders möchte ich mich bei meiner Familie bedanken, die mich während des gesamten Zeitraums des Studiums und der Doktorarbeit unterstützt und ermutigt haben.

Nicht zuletzt möchte ich meinem Mann Niko für jegliche Unterstützung danken. Ohne ihn wäre ich heute nicht da, wo ich bin.

Curriculum Vitae

Nina Schäper

Johann-Friedrich Blumenbach Institut for Zoology and Anthropology
Dept for Developmental Biology, GZMB Ernst-Caspari-Haus
Justus-von-Liebig-Weg 11, 37077 Göttingen
Phone: 0551/397037
Enail : nschaep@gwdg.de

Personal data

Name:	Nina Daniela Schäper
Date of Birth:	13.09.1980
Place of Birth:	Weinheim (a. d. Bergstraße)
Marital Status:	married, no children

Education

May 2000	Abitur, Gymnasium Wesermünde, Bremerhaven, Germany
2000-2005	Studies in Biology, Georg August University Göttingen Main subjects: Developmental Biology (major), Genetics, Microbiology
2004-2005	Diploma Thesis in the Department of Developmental Biology, Georg-August University Göttingen, Topic: “Isolation and Characterization of *Sp* Transcription Factor Genes in the Beach Hopper *Parhyale hawaiensis*”. Supervisor: Prof. Dr. Ernst A. Wimmer
August 2005- December 2009	Dissertation (Ph.D) in Developmental Biology, Georg-August University Göttingen, Topic: “Evolution of Sp Transcription factors in Metazoans” Supervisor: Prof. Dr. Ernst A. Wimmer
August 2006 – May 2007	visting graduate student at University of California, Berkeley Dept. for Integrative Biology, Supervisor: Prof. Dr. Nipam H. Patel

Grants

2006/2007	Fulbright Foundation Travel Grant
2006/2007	EAP (eduacation abroad program), full stipend of the university Göttingen and the University of California.

Extracurricular courses and external teaching activities

2007/2008	two term correspondence course at ‘FernUniversität Hagen’ industrial and intellectual properties (“Gewerblicher Rechtsschutz”), certificate as “Patentingeneurin” / “Patentreferentin”
July 2008	’teaching assistant’ in the EMBO and Marie Curie RTN ZOONET course "Molecular approaches to Evolution and Development" Kristineberg Marine Station, Fiskebäckskil, Sweden
August-October 2003	Internship at BASF AG in Ludwigshafen, Germany, Department of Physical Chemistry and Informatics / DNA-Analytics.

Talks and presentations

July 2008	Posterpresentation, EMBO and Marie Curie RTN ZOONET course "Molecular approaches to Evolution and Development", Fiskebäckskil (Sweden)
November 2007	Talk at the Marie Curie Research Training Network (ZOONET) conference, Barcelona (Spain)
September 2007	Talk at the 3. internat. *Tribolium* conference, Göttingen
September 2005	Talk at the internat. *Parhyale* conference, Tübingen
April 2005	Posterpresentation, GfE meeting, Münster

Publication

Schaeper N.D., Prpic N.M., Wimmer E.A. (2009). A conserved function of the zinc finger transcription factor *Sp8/9* in allometric appendage growth in the milkweed bug *Oncopeltus fasciatus*. *Development Genes and Evolution* 219:427-435.