

RESEARCH ARTICLE

Open Access

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

Nina D Schaeper, Nikola-Michael Prpic, Ernst A Wimmer\*

## 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. The purpose of the presented study was 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 a holometabolous insect (*Tribolium castaneum*), a hemimetabolous insect (*Oncopeltus fasciatus*), primitively wingless hexapods (*Folsomia candida* and *Thermobia domestica*), and an amphipod crustacean (*Parhyale hawaiensis*). 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, gene expression, and chromosomal location. 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 synteny 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 not the result of a recent gene duplication, but directly traces back to 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. [1]). 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 [2-4].

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. [1,5]). 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 (GGGGCGGG) and GT/CACC-boxes (GGTGTGGG) [4]. These binding sites are present in many control regions of both tissue-specific and ubiquitously expressed genes [6,7] indicating that Sp-family transcription factors potentially regulate a large number of target genes. Indeed, it was

\* Correspondence: ewimmer@gwdg.de  
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

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. [5,8-17]). 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 [8,18-20].

The number of Sp-family genes present in the genome varies in the Metazoa. Humans and mice, for example, have nine Sp-family genes [5], and some teleost fishes have even more (11 in the pufferfish *Fugu rubripes* [21], 13 in the zebrafish *Danio rerio* [22]). From *D. melanogaster* two Sp-family genes have been reported, *btd* and *D-Sp1* [8], but a third one is present in the fully sequenced genome sequence [23]. This variable complement of Sp-family genes and their evolutionary diversification made 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. For example, *D. melanogaster D-Sp1* is not most closely related to human *Sp1* but to *Sp8* [24] and the gene originally termed mouse *mBtd* is in fact orthologous to *Sp8* [13].

Especially the origin and orthology of the *D. melanogaster* head gap gene *btd* has been debated. Previous studies discovered functional similarities between *btd* and some vertebrate Sp genes, but could not confidently identify a genuine *btd* orthologue in vertebrates [13,15,25], 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 in *D. melanogaster* [8,20]. 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 [8,20]. 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 hexapods, 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, spatio-temporal mRNA expression analysis, as well as genomic localisation 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 synteny 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.

## 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* [8,19]. 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 [8,20]. In order to identify when such a gene duplication event might have occurred, 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* (silkworm), 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 [26]. 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 primitive hexapods (Entognatha) [26]. 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 could 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. [27-30]). The waterflea *Daphnia pulex* is a member of the Branchiopoda. 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 hawaiiensis* (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. hawaiiensis*.

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 lineage *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 potentially have originated independently.

#### Phylogenetic analysis of Sp-family genes supports three large clades

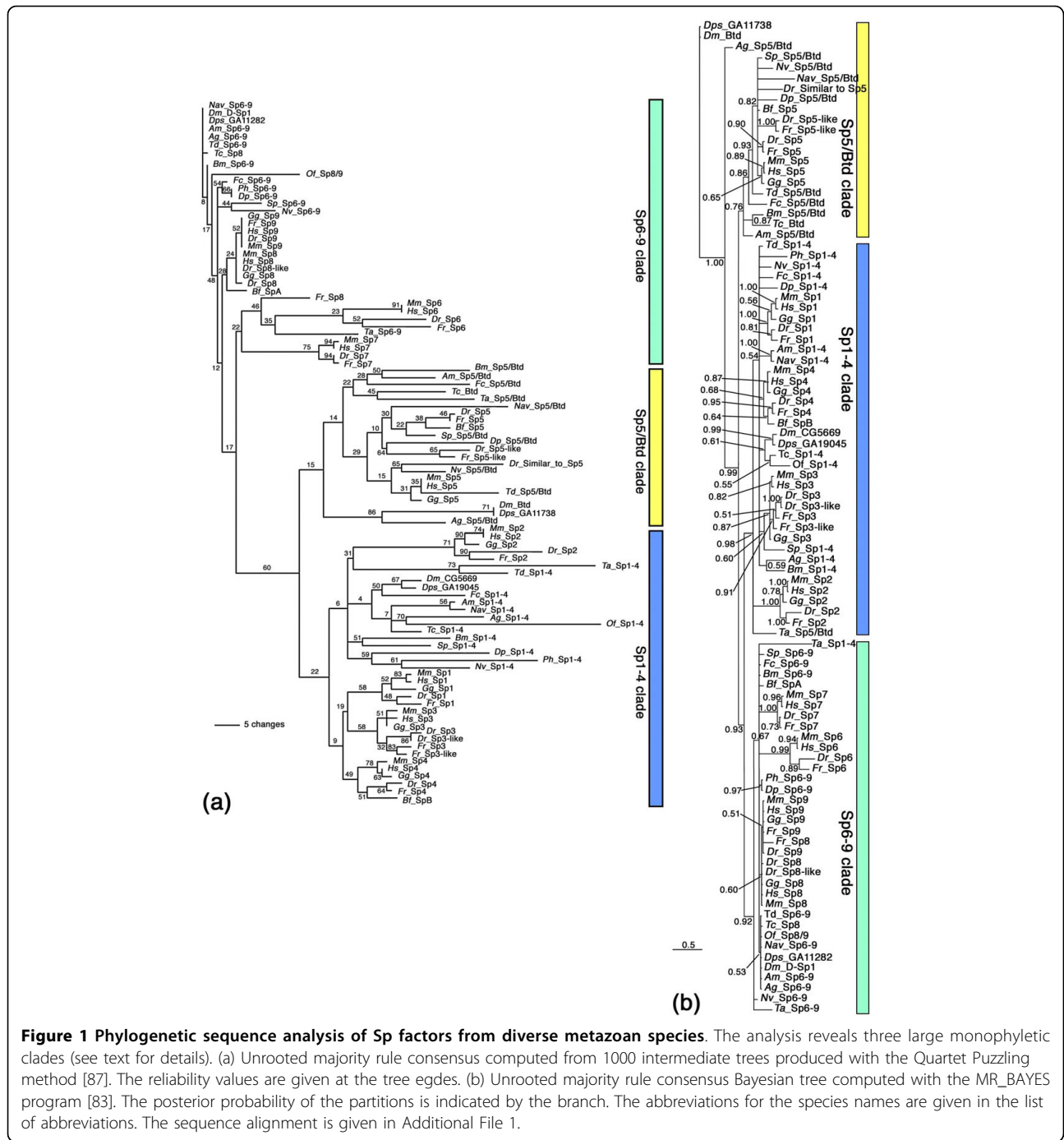
In order to distinguish between a possible ancestral set of three Sp-family genes and the alternative possibility of several independent duplication events, we reconstructed the evolutionary history of identified Sp-family factors and assigned orthology by phylogenetic sequence analysis. We used the 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. 1a and the alignment is shown in Additional File 1. 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. hawaiiensis* 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 vertebrate 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 this 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 further duplications in the vertebrate lineage due to the whole genome duplications that occurred early in vertebrate evolution (discussed below).

We have in addition performed a Bayesian analysis of the same dataset. The resulting unrooted tree is shown in Fig. 1b. The Bayesian tree differs from the quartet puzzling tree in several places, but the only marked differences are Sp5/Btd and Sp1-4 from *T. adhaerens*, which are not included in the Sp5/Btd and Sp1-4 clade, respectively. The inconsistent placement of these *T. adhaerens* sequences in the two analyses might be explained by the phylogenetically old age of this lineage. Importantly, the three monophyletic clades Sp1-4, Sp5/Btd, and Sp6-9 are also recovered with this method and all three clades have very high support values. Thus, this additional analysis over all supports the quartet puzzling analysis.

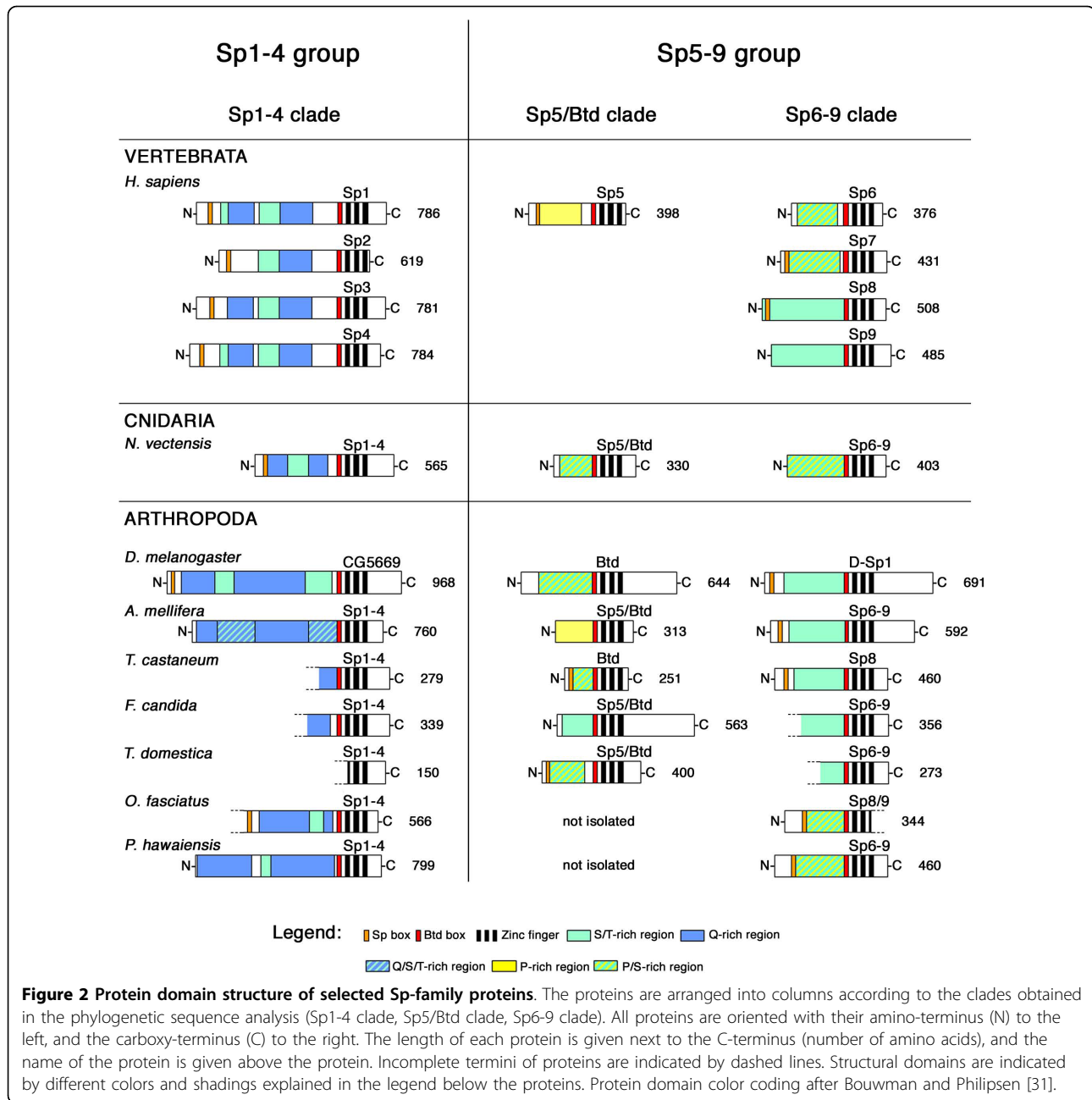
#### Protein structure supports the existence of two large groups of Sp factors

It has been noted previously that the Sp proteins contain additional structural domains besides the zinc fingers and Btd box (e.g. [31]). 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 latter 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.

**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 an additional 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 orthologs, but show different patterns if they originated through unrelated duplication events. In the following we compare the expression data from insects, the crustacean *P. hawaiiensis* and published data from vertebrates arranged according to the three Sp-gene clades.

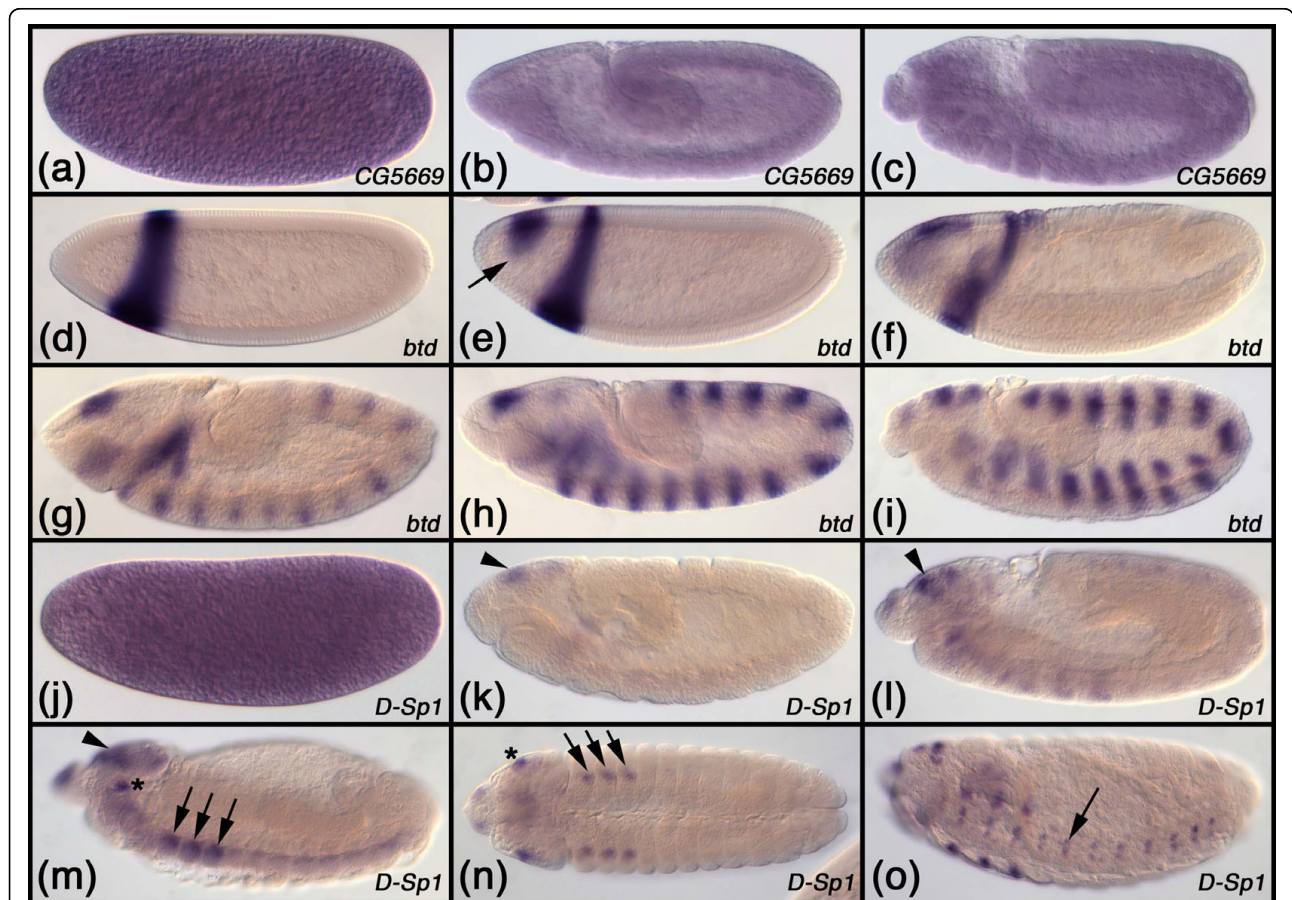
**The genes of the Sp1-4 clade**

*CG5669*, which is the *D. melanogaster* representative of this clade, is maternally contributed (Fig. 3a) and then expressed ubiquitously throughout development (Fig. 3b, c). In *T. castaneum* the *Sp1-4* gene (in [32] previously termed *Tc-SP1234*) is expressed ubiquitously throughout development as well (Fig. 4a-c). The same is true for the *Sp1-4* gene of *O. fasciatus* (Fig. 5a-c), *T. domestica* (Fig. 6a-c) and *F. candida* (Fig. 7a, b). In the crustacean *P. hawaiiensis* the *Sp1-4* gene is also expressed ubiquitously throughout all studied developmental stages (Fig. 8a-c).

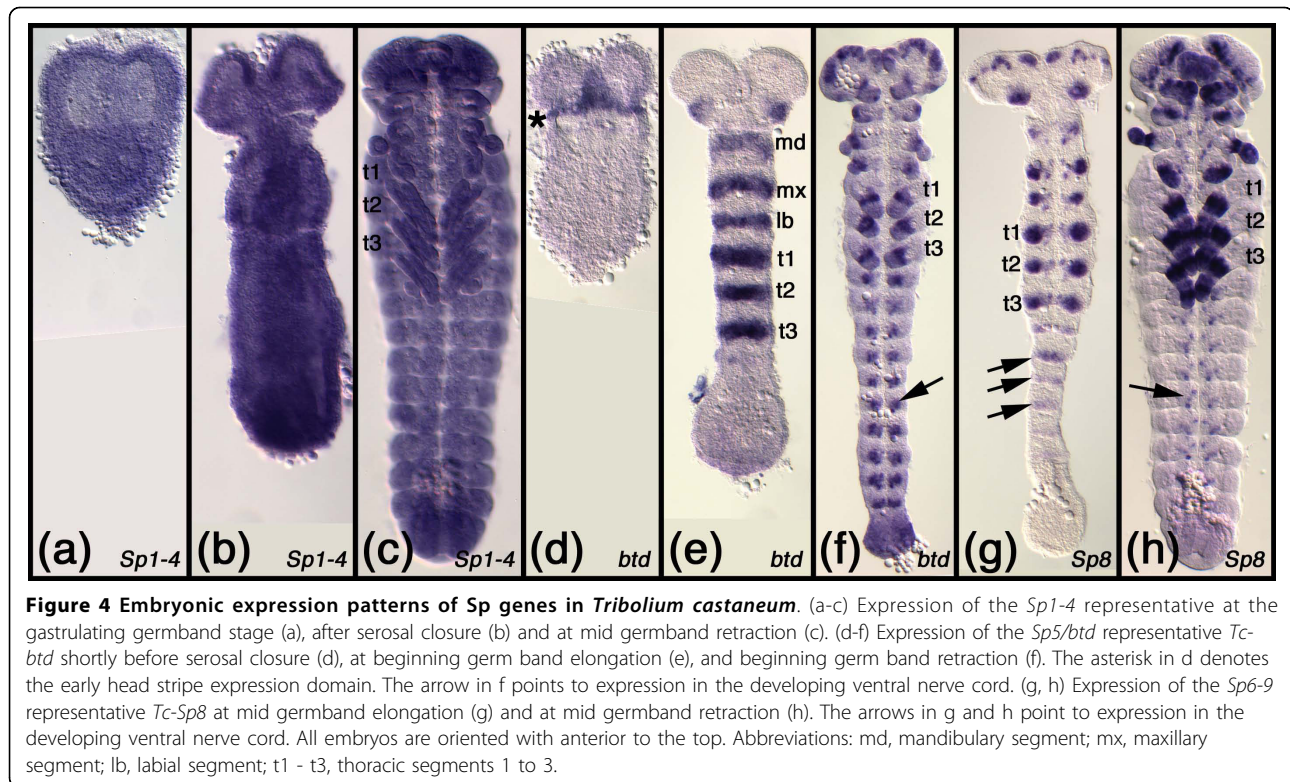
The members of this clade from the mouse have not all been studied as to their embryonic expression pattern, but data are available for murine *Sp1*, *Sp3* and *Sp4* [33-37]. All three genes are expressed ubiquitously during development. Taken together, these data show that all analyzed members of this 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 [8,19]. The gene is first expressed in an anterior head stripe (Fig. 3d) and a dorsal spot appears slightly later (Fig. 3e). The head stripe is roughly located in the area of the intercalary and mandibular segment and later abuts the cephalic furrow (Fig. 3f). Later a metameric (segmentally repeated) pattern emerges that might be



**Figure 3 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 4 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, mandibular segment; mx, maxillary segment; lb, labial segment; t1 - t3, thoracic segments 1 to 3.

correlated with segment formation and peripheral nervous system development (Fig. 3g-i) [8,20]. Furthermore, *Dm btd* is expressed in the imaginal discs of legs and antennae [8,38]. The expression of the *T. castaneum btd* gene has been published before [32] 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 mandibular segment (Fig. 4d) and later a metameric pattern emerges (Fig. 4e). In older stages the gene is also expressed in the appendages and in the nervous system (Fig. 4f). 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. 6d), that lies in the intercalary/mandibular area in slightly more advanced germ band stage embryos (Fig. 6e). Later a metameric pattern emerges (Fig. 6f, g) and in older stages expression in the nervous system and, weakly, in the appendages is detected (Fig. 6 h). 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. 7c, d), a weak expression in the appendages (Fig. 7d), and expression in the nervous system (Fig. 7e).

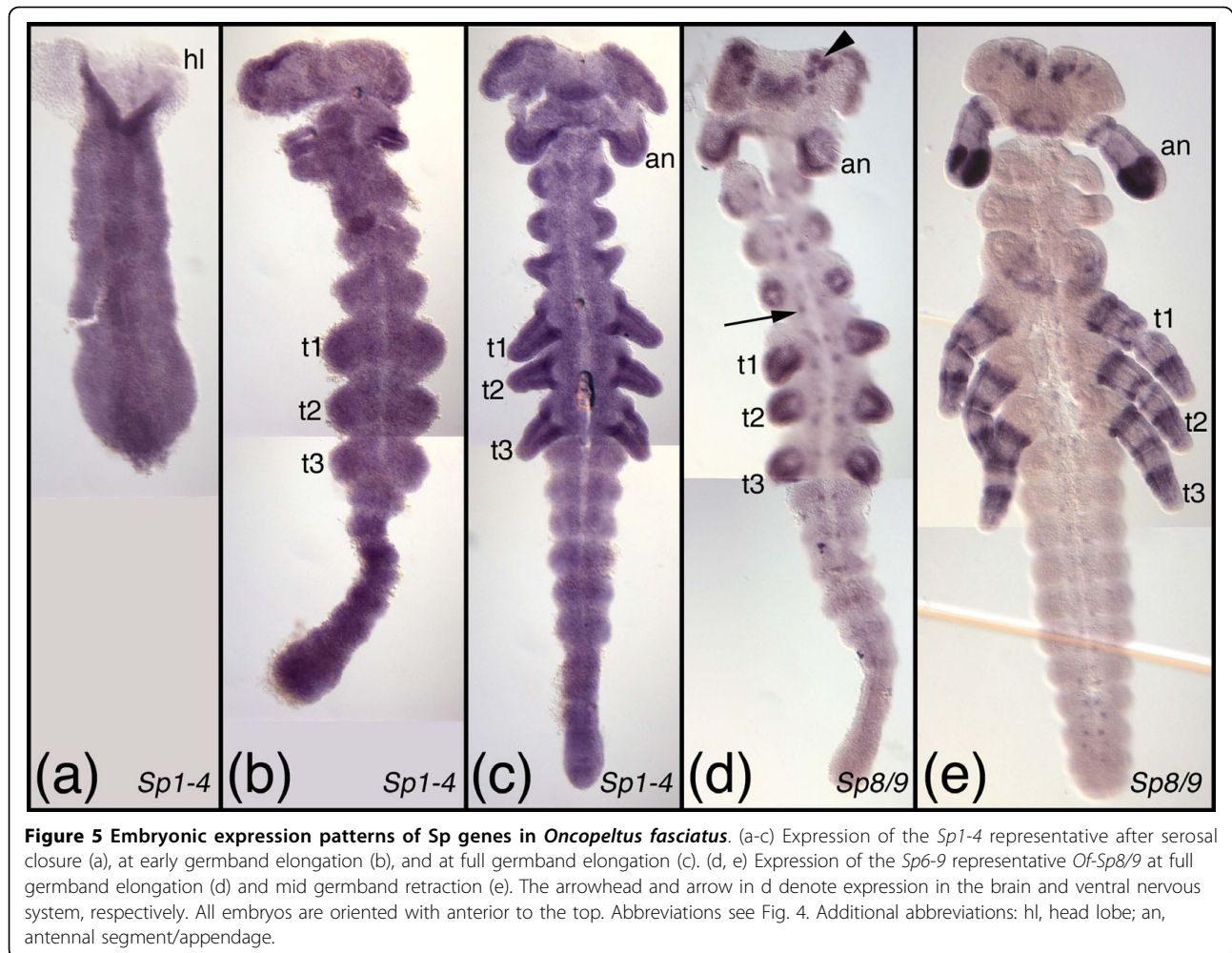
There are 3 genes related to *Sp5* in the zebrafish genome. *Sp5* (also known as *bts1*) [25], *Sp5-like* (also

known as *spr2*) [39] 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 [25]. Zebrafish *Sp5-like* expression is partially overlapping the *Sp5* expression in ectodermal and mesodermal tissue, the brain, trunk neural crest cells, and somites [39]. 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 [12,40]. 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 [8,20]. The gene is maternally contributed (Fig. 3j), and earliest embryonic expression is seen in the brain (Fig. 3k, l). Later, strong expression is seen in the limb primordia of the antennae and legs (Fig. 3m, n) and in a punctate pattern in the ventral nerve cord (Fig. 3o). The expression of the *T. castaneum Sp8* gene has been reported earlier [24]. 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. 4g, h). In the growing legs the gene is expressed in a pattern





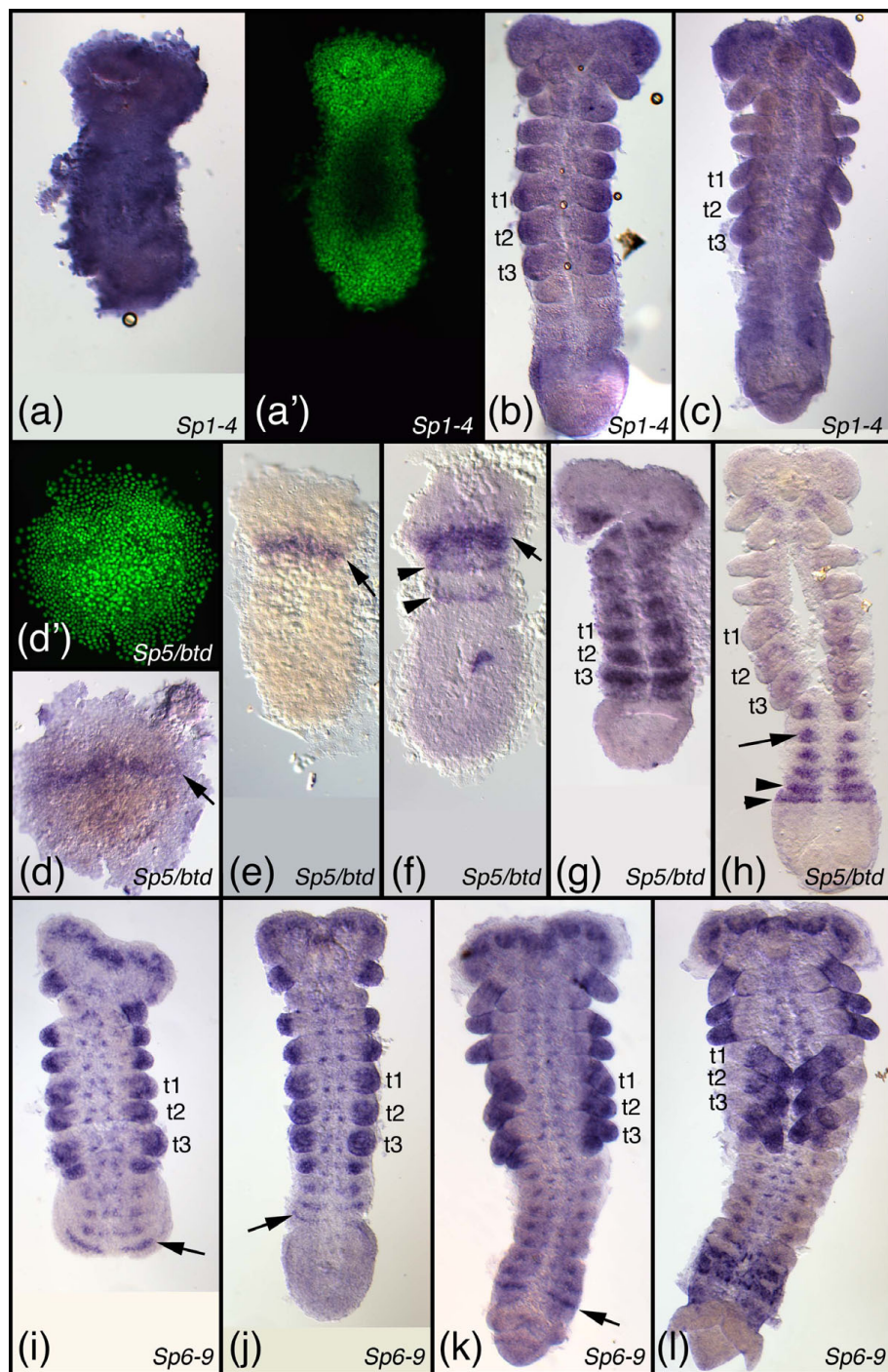
**Figure 5 Embryonic expression patterns of Sp genes in *Oncopeltus fasciatus*.** (a-c) Expression of the *Sp1-4* representative after serosal closure (a), at early germband elongation (b), and at full germband elongation (c). (d, e) Expression of the *Sp6-9* representative *Of-Sp8/9* at full germband 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. 4. Additional abbreviations: hl, head lobe; an, antennal segment/appendage.

comprising several rings (Fig. 4h) [24]. The gene *Sp8/9* from *O. fasciatus* has been published recently [41]. *Sp8/9* is expressed in the brain, in a punctate pattern in the ventral nerve cord and in the limbs (Fig. 5d). Similar to the legs in older *T. castaneum* embryos, the *O. fasciatus Sp8/9* gene is expressed in several rings in the legs (Fig. 5e). The *Sp6-9* gene from *T. domestica* is expressed in the limb buds (Fig. 6i, j) and later in at least two rings in the legs (Fig. 6k, 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. 7f-h). The gene is also expressed in the limb buds (Fig. 7f-h) and at later stages in two separate rings in the legs (Fig. 7i). 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*

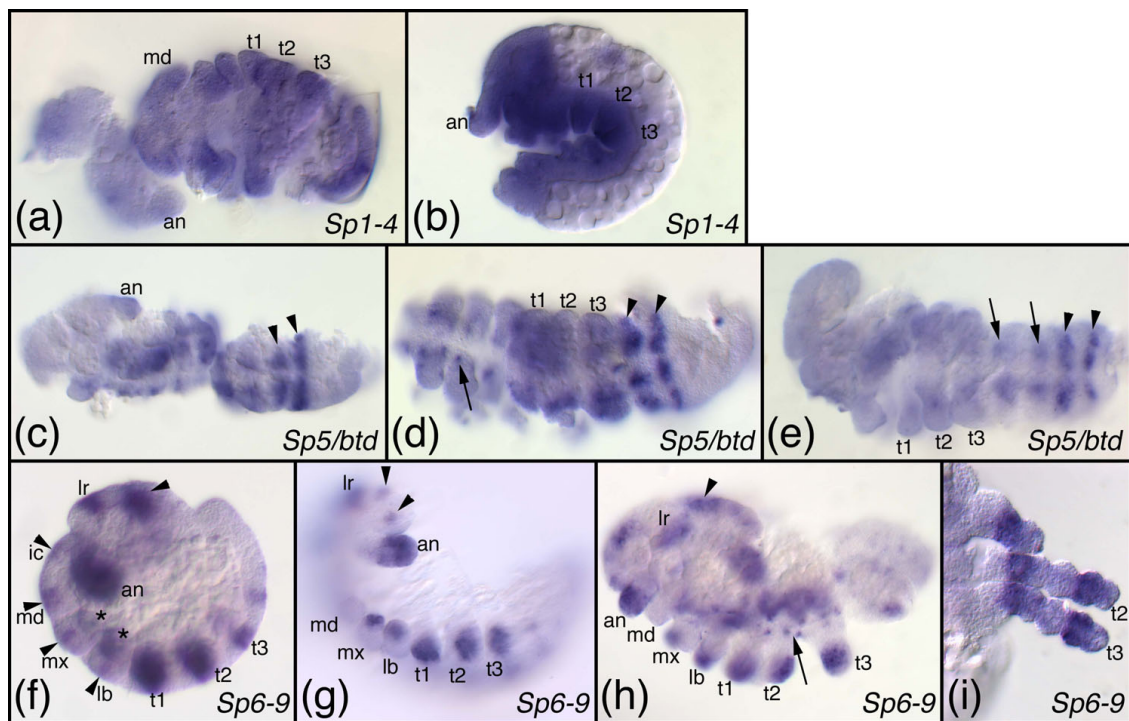
expression in *P. hawaiiensis*. In this species the gene is expressed in the limb buds (Fig. 8d, e) and at later stages in the peraeopods and in the two branches of the pleopods and the first two pairs of uropods (Fig. 8f). In addition, there is a punctate expression pattern in the ventral nerve cord (Fig. 8f).

Expression data for the members of this clade are also available from vertebrates. Initial RT-PCR analysis of mouse *Sp6* expression suggested expression in all tissues studied [42], but later studies showed a specific expression pattern in hair follicles and the apical ectodermal ridge (AER) of the developing limbs [15,43]. 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 [44-46]. 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





**Figure 6 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 (a). (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. 4.



**Figure 7 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 metamer 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. 4. Additional abbreviations: an, antennal segment/appendage; lr, labrum; ic, intercalary segment.

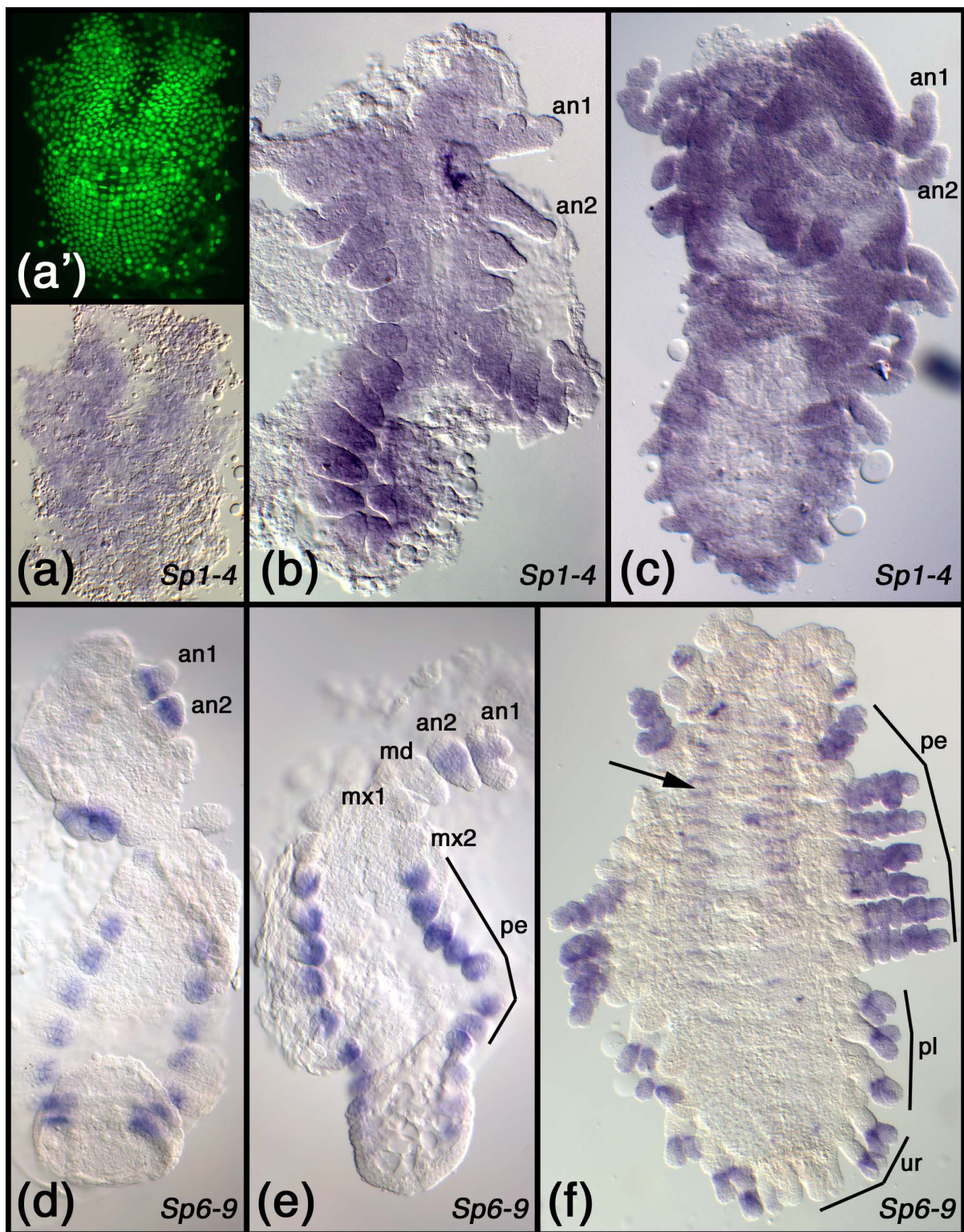
for limb and fin outgrowth [13,14,47,48]. In summary, the expression patterns of the genes in this clade are strikingly similar in the insects and crustaceans and very similar expression patterns also exist from some vertebrate representatives of this clade, again supporting the orthology of the genes 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 (the *Sp5-9* group).

### 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. 9 and the exact locations are given in Additional File 2. 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. [49] and Putnam et al. [50] 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





**Figure 8 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. [68]. (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.

*N. vectensis* Hox genes on the hypothetical ancestral linkage group PAL A (Fig. 9, top) [50]. 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 conserved their close linkage in Cnidaria and vertebrates, and to a lesser extent in arthropods (Fig. 9). The Sp genes are located close to the Hox gene cluster in other animals as well (see also [31,51]. 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. 9, 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. 9, 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 to the assembled chromosome. The genomes of *S. purpuratus*, *B. floridae* and *T. adhaerens* are not assembled to the chromosome or linkage group level, 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 [52]) and *T. adhaerens* the *Sp5/Btd* and *Sp6-9* genes are located on the same scaffold (see also [53]). Whether the Sp-family genes are also linked to the Hox genes in *S. purpuratus* (see [54]), *B. floridae* (see [55,56]), or *T. adhaerens* (see [57]) has to await the full assembly of the scaffolds.

## 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 since it has only one Hox-like gene (*Trox-2*) along with one further, very derived gene with potential Hox-like affinities [57,58]. But it is

yet unclear whether the single *T. adhaerens* Hox-like gene is physically close to the Sp-family genes.

The eumetazoan ancestor already possessed a triplet cluster of Sp-family genes (Fig. 10, “eumetazoan 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. 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 [49,59]. This cluster apparently has broken apart during cnidarian evolution leading to the dispersed set of 8 Hox genes in *N. vectensis* [49]. 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” linkage group [50]. 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 also had an impact on 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* genes (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 in tetrapods [60], 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 [61,62] 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 gene duplication, but traces back to an ancient *Sp5/Btd* gene already present in basal metazoans.

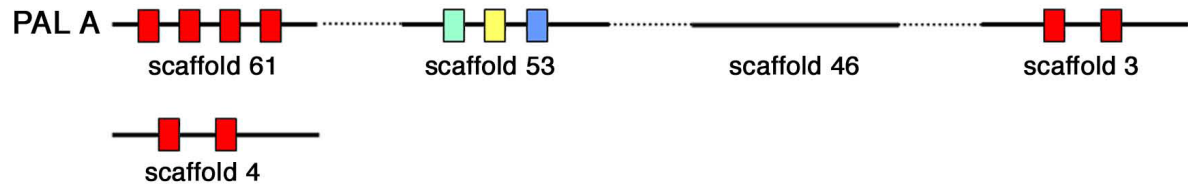
## Materials and methods

### Arthropod husbandry, embryo collection and fixation

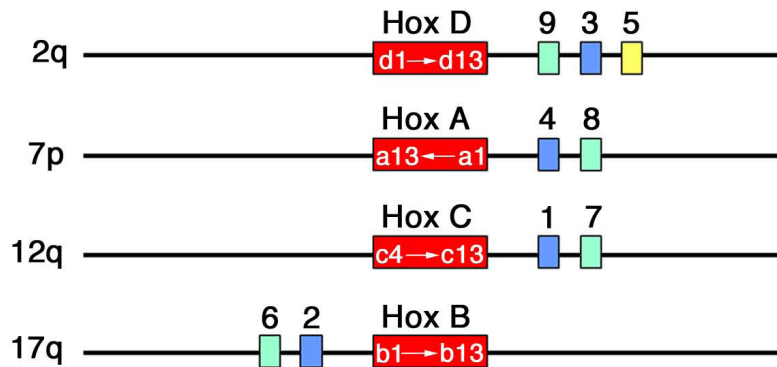
The *O. fasciatus* (milkweed bug) culture was kept as described in Hughes and Kaufman [63]. Embryos of all stages were fixed as reported previously [64]. Dissections of milkweed bug embryos were performed under a fluorescence stereomicroscope using SYTOX Green



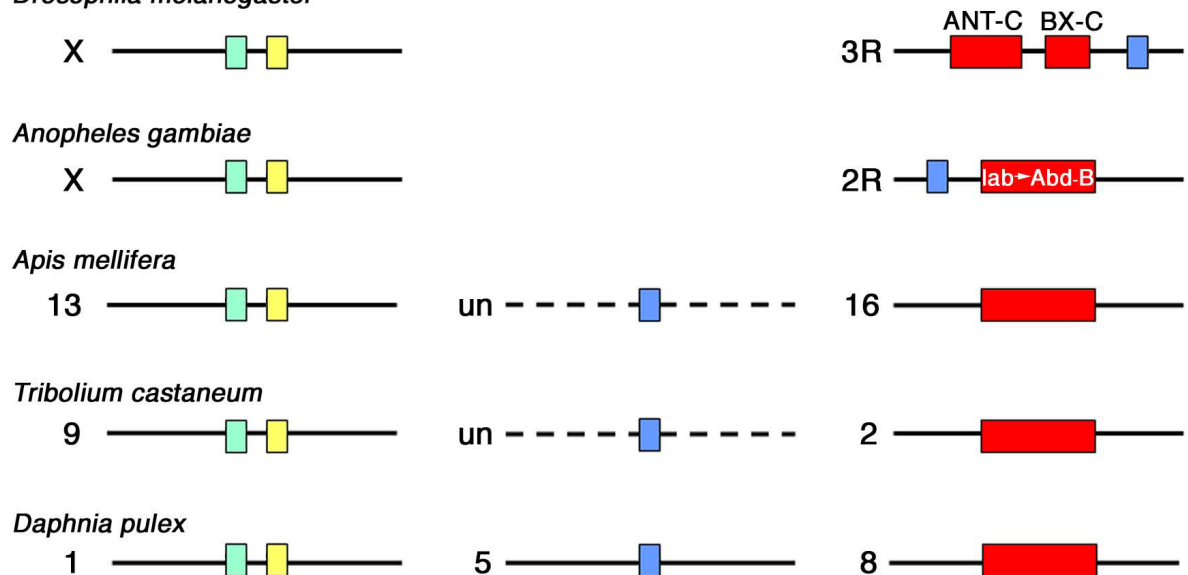
*Nematostella vectensis*



*Homo sapiens*



*Drosophila melanogaster*



*Anopheles gambiae*



*Apis mellifera*



*Tribolium castaneum*

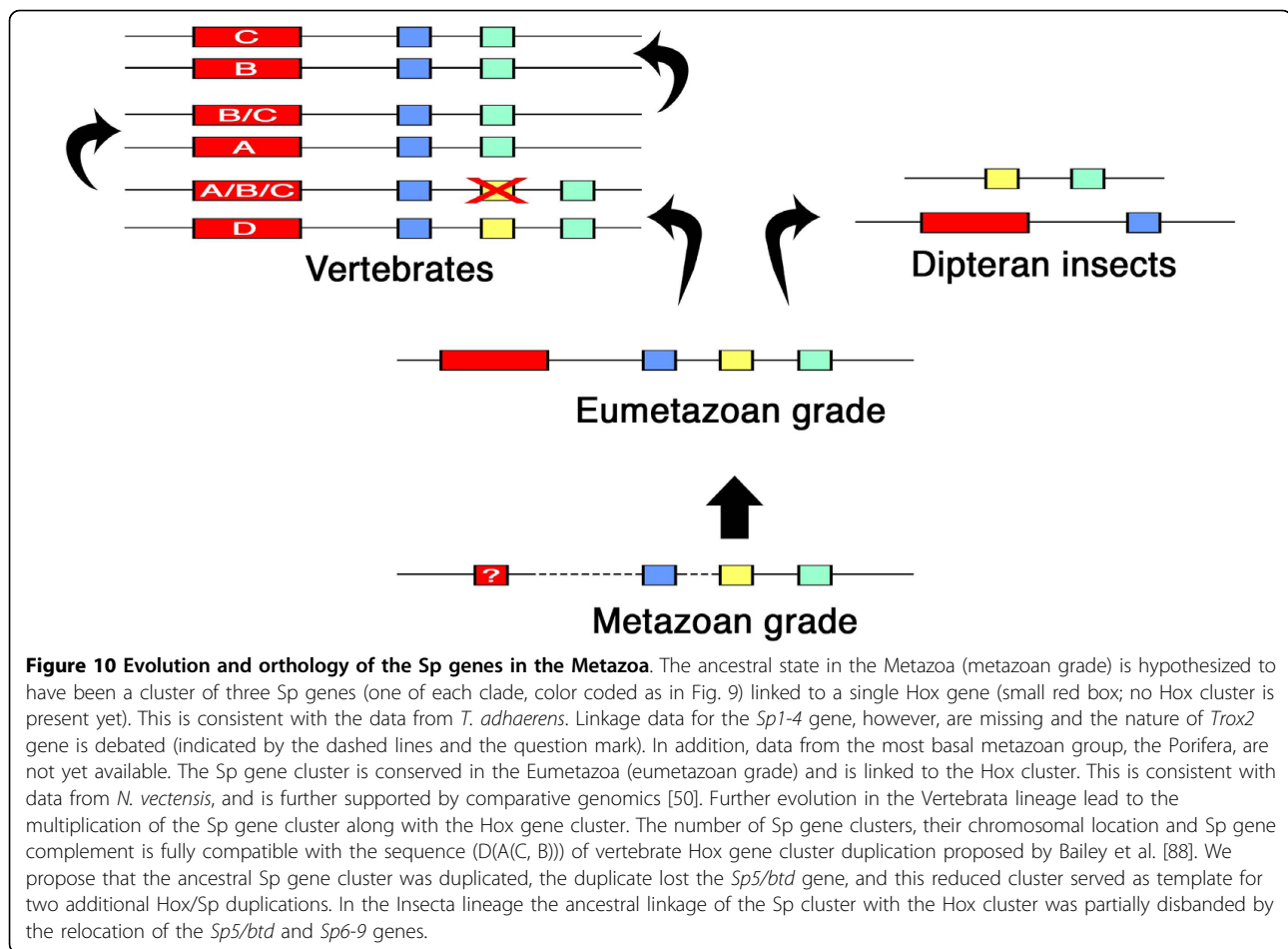


*Daphnia pulex*



Legend: ■ Sp1-4 clade ■ Sp5/Btd clade ■ Sp6-9 clade ■ Hox

**Figure 9 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 [50]. The hypothetical linkage between these scaffolds is indicated by the dotted lines.



nucleic acid stain (Invitrogen) before in situ staining [65]. *T. domestica* (firebrat) were cultured as described in Rogers et al. [66] 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* [65]. *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, embryos 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. hawaiiensis*

(amphipod beachhopper) were cultured in shallow plastic boxes at 26°C filled with a thin layer of crushed coral substrate and artificial seawater (30 g/l of synthetic sea salt) and fed with dry fish flakes twice a week. Membrane pumps ventilated the water. Gravid amphipod females were anaesthetized with clove oil (10 µl per 50 ml seawater) and embryos were collected out of the brood pouch with forceps. Dissection and fixation was performed as described in Browne et al. [67].

#### 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. hawaiiensis* embryos of all described stages [68], 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 different 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\_GRATDCPNC (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 gene specific primers designed on the basis of the results of the degenerate primers PCR. RACE primer sequences are given in Additional File 3. PCR fragments were cloned into the pCR-II (Invitrogen) and sequenced. 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 (Sp1-4)* [GenBank: NM\_142975] from *D. melanogaster* cDNA. The sequences of these primers are given in Additional File 3. We have used the publicly available genome sequencing data for a selection of metazoan species: *H. sapiens* (Genome Reference Consortium Human Build 37 (GRCh37), Primary\_Assembly) [69,70], *M. musculus* (Reference assembly (C57BL/6J)) [71], *N. vitripennis* [72], *D. melanogaster* (release 5.10) [23], *D. pseudobscura* (release Dpse\_2.0) [73], *A. mellifera* (Amel\_4.0) [74], *A. gambiae* (AgamP3.3) [75], *T. castaneum* (Tcas\_3.0) [76], *B. mori* (version 01 BABH01000000) [77], *D. pulex* (JGI-2006-09) [78], *S. purpuratus* (Build 1.1) [79], *N. vectensis* (Nematostella vectensis v1.0) [50], *G. gallus* (Gallus\_gallus-2.1) [80], *F. rubripes* (Fourth Fugu Genome assembly) [21], *D. rerio* [22], *B. floridae* (Branchiostoma floridae v1.0) [81], and *T. adhaerens* (Trichoplax adhaerens Grell-BS-1999 v1.0) [58]. Phylogenetic analysis of different Sp transcription factor sequences with the Quartet Puzzling method was performed as described in Prpic et al. [82]. Additional Bayesian analysis was performed using MrBayes [83] and the tree was visualized with PhyloWidget [84]. The accession numbers and the protein sequences alignment are described in Additional File 1.

## In situ hybridization

The length of the templates, the clone ID, and the RNA polymerase used for digoxigenin labeled RNA probe synthesis are given in Additional File 3. *D. melanogaster* and *T. castaneum in situ* was performed essentially as described in Wohlfrom et al. [85], *O. fasciatus in situ* hybridization was done according to Liu and Kaufman [64], *P. hawaiiensis in situ* was performed as reported in Browne et al. [67], and *in situ* hybridizations for *F. candida* and *T. domestica* were done essentially as described in Hughes et al. [86].

### Additional file 1: Sequence alignment used as basis for the phylogenetic analysis shown in Fig. 1.

CLUSTAL X (1.81) multiple sequence alignment of different Sp factors comprising the conserved region of the Btd box (in blue) and the three zinc fingers (in red). Accession numbers of used Sp proteins: *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].

**Additional file 2: Genomic locations of Sp genes and Hox genes.**

This table supplements the schematic overview given in Fig. 9. 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. The base pair values and genomic positions are based on the following genome assembly versions: *H. sapiens*: Genome Reference Consortium Human Build 37 (GRCh37), Primary\_Assembly; *D. melanogaster*: release 5.10, *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.

**Additional file 3: Sequence information about the gene specific primers used in this study.** The first column gives the species and the gene. The second column gives the primer sequences in 5' to 3' orientation. The third column gives the length of the cloned fragment resulting from the PCR with the given primers. The fourth column gives the clone ID number. The fifth column gives the polymerase used to transcribe the RNA probe used for in situ hybridizations. The primers for *D. melanogaster* and *T. castaneum* have been designed as gene specific pairs using the genome sequence information. For *O. fasciatus*, *T. domestica*, *F. candida*, and *P. hawaiiensis* we first isolated a small fragment of the genes using degenerate primers specified in the Materials and methods section. The gene specific RACE primers were designed on the basis of this sequence information and were used in conjunction with the commercial RACE adaptor primers. The cloned fragment of *Ph Sp6-9* resulted from priming of the given primer pair. Abbreviations: Fwd, forward; Rev, reverse.

**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 hawaiiensis*; *Sp*: *Strongylocentrotus purpuratus*; *Ta*: *Trichoplax adhaerens*; *Tc*: *Tribolium castaneum*; *Td*: *Thermobia domestica*.

**Acknowledgements**

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

**Authors' contributions**

NDS carried out the molecular genetic studies, genome location, protein domain analyses, 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.

Received: 13 January 2010 Accepted: 30 March 2010

Published: 30 March 2010

**References**

1. Suske G, Bruford E, Philipsen S: Mammalian SP/KLF transcription factors: bring in the family. *Genomics* 2005, **85**:551-556.
2. Dynan WS, Tjian R: The promoter-specific transcription factor Sp1 binds to upstream sequences in the SV40 early promoter. *Cell* 1983, **35**:79-87.
3. Dynan WS, Tjian R: Isolation of transcription factors that discriminate between different promoters recognized by RNA polymerase II. *Cell* 1983, **32**:669-680.

4. 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.
5. Zhao C, Meng A: Sp1-like transcription factors are regulators of embryonic development in vertebrates. *Develop Growth Differ* 2005, **47**:201-211.
6. Philipsen S, Suske G: A tale of three fingers: the family of mammalian Sp/KLF transcription factors. *Nucleic Acids Res* 1999, **27**:2991-3000.
7. Suske G: The Sp-family of transcription factors. *Gene* 1999, **238**:291-300.
8. Wimmer EA, Frommer G, Purnell BA, Jäckle H: *buttonhead* and *D-Sp1*: a novel *Drosophila* gene pair. *Mech Dev* 1996, **59**:53-62.
9. 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.
10. Black AR, Jensen D, Lin SY, Azizkhan-Clifford J: Growth/Cell Cycle Regulation of Sp1 Phosphorylation. *Journ of Biol Chem* 1999, **274**:1207-1215.
11. 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.
12. 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.
13. 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.
14. Kawakami Y, Esteban CR, Matsui T, Rodríguez-León J, Kato S, Izpisua 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.
15. 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.
16. Safe S, Abdelrahim M: Sp transcription factor family and its role in cancer. *Eur J Cancer* 2005, **41**:2438-2448.
17. 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.
18. Cohen SM, Jürgens G: Mediation of *Drosophila* head development by gap-like segmentation genes. *Nature* 1990, **346**:482-485.
19. Wimmer EA, Jäckle H, Pfeifle C, Cohen SM: A *Drosophila* homologue of human Sp1 is a head-specific segmentation gene. *Nature* 1993, **366**:690-694.
20. 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.
21. 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, Verhoeve 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, et al: Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*. *Science* 2002, **297**:1301-1310.
22. Zebrafish Sequencing Project. [http://www.sanger.ac.uk/Projects/D\_zerrio/].
23. 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, et al: The genome sequence of *Drosophila melanogaster*. *Science* 2000, **287**:2185-2195.
24. 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.
25. 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.
26. Klausnitzer B: *Insecta (Hexapoda), Insekten. Spezielle Zoologie. Teil 1: Einzeller und wirbellose Tiere* Munich: Spektrum Akademischer Verlag, ElsevierWestheide W, Rieger R, Second 2007, 638-654.



27. Friedrich M, Tautz D: Ribosomal DNA phylogeny of the major extant arthropod classes and the evolution of myriapods. *Nature* 1995, **376**:165-167.
28. Dohle W: Myriapod-insect relationships as opposed to an insect crustacean sister group relationship. *Arthropod Relationships* London: Chapman & Hall/Forley RA, Thomas RH 1997, 305-315.
29. 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.
30. Budd GE, Telford MJ: The origin and evolution of arthropods. *Nature* 2009, **457**:812-817.
31. Bouwman P, Philippen S: Regulation of the activity of Sp1-related transcription factors. *Mol Cell Endocrinol* 2002, **95**:27-38.
32. 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.
33. Saffer JD, Jackson SP, Annarella MB: Developmental expression of Sp1 in the mouse. *Mol Cell Biol* 1991, **11**:2189-2199.
34. 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.
35. Bouwman P, Gollner H, Elsässer HP, Eckhoff G, Karis A, Grosveld F, Philippen S, Suske G: Transcription factor Sp3 is essential for post-natal survival and late tooth development. *EMBO J* 2000, **19**:655-661.
36. Gollner H, Dani C, Phillips B, Philippen S, Suske G: Impaired ossification in mice lacking the transcription factor Sp3. *Mech Dev* 2001, **106**:77-83.
37. Gollner H, Bouwman P, Mangold M, Karis A, Braun H, Rohner I, Del Rey A, Besedovsky HO, Meinhardt A, Broek van den M, Cutforth T, Grosveld F, Philippen S, Suske G: Complex phenotype of mice homozygous for a null mutation in the Sp4 transcription factor gene. *Genes Cells* 2001, **6**:689-697.
38. Estella C, Rieckhof G, Calleja M, Morata G: The role of *buttonhead* and Sp1 in the development of the ventral imaginal discs of *Drosophila*. *Development* 2003, **130**:5929-5941.
39. 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.
40. 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.
41. 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.
42. 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.
43. 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.
44. Nakashima K, Zhou X, Kunkel G, Zhang Z, Deng JM, Behringer RR, de Crombrughe B: The novel zinc finger-containing transcription factor Osterix is required for osteoblast differentiation and bone formation. *Cell* 2002, **108**:17-29.
45. Milona MA, Gough JE, Edgar AJ: Expression of alternatively spliced isoforms of human Sp7 in osteoblast-like cells. *BMC Genomics* 2003, **4**:43.
46. 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.
47. 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.
48. Griesel G, Treichel D, Collombat P, Krull J, Zembrzycki A, Akker van den WM, Gruss P, Simeone A, Mansouri A: Sp8 controls the anteroposterior patterning at the midbrain-hindbrain border. *Development* 2006, **133**:1779-1787.
49. 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.
50. 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.
51. Abbasi AA, Grzeschik KH: An insight into the phylogenetic history of HOX linked gene families in vertebrates. *BMC Evol Biol* 2007, **7**:239.
52. Shimeld SM: C2H2 zinc finger gene of the Gli, Zic, KLF, SP, Wilms' tumor, Hucklebein, Snail, ovo, Spalt, Odd, Blimp-1, Fez and related gene families from *Branchiostoma floridae*. *Dev Genes Evol* 2008, **218**:639-649.
53. 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.
54. 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.
55. 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.
56. 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-Fernández 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, et al: The amphioxus genome illuminates vertebrate origins and cephalochordate biology. *Genome Res* 2008, **18**:1100-1111.
57. 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.
58. 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, **427**:955-960.
59. Hejnol A, Martindale MQ: Coordinated spatial and temporal expression of Hox genes during embryogenesis in the acoel *Convolutriloba longifissura*. *BMC Biol* 2009, **7**:65.
60. 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.
61. Taylor JS, Peer Van der 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.
62. Venkatesh B: Evolution and diversity of fish genomes. *Curr Opin Genet Dev* 2003, **13**:588-592.
63. 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.
64. 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* 2004, **131**:1515-1527.
65. 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* 2004, **131**:4567-4579.
66. 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.
67. Browne WE, Schmid BGM, Wimmer EA, Martindale MQ: Expression of *otd* orthologs in the amphipod crustacean, *Parhyale hawaiiensis*. *Dev Genes Evol* 2006, **216**:581-595.
68. Browne WE, Price AL, Gerberding M, Patel NH: Stages of embryonic development in the amphipod crustacean, *Parhyale hawaiiensis*. *Genesis* 2005, **42**:124-149.

69. International Human Genome Sequencing Consortium: **Initial sequencing and analysis of the human genome.** *Nature* 2001, **409**:860-921.
70. 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, et al: **The sequence of the human genome.** *Science* 2001, **291**:1304-1351.
71. Mouse Genome Sequencing Consortium: **Initial sequencing and comparative analysis of the mouse genome.** *Nature* 2002, **420**:520-562.
72. **Nasonia Genome Project.** [http://www.hgsc.bcm.tmc.edu/project-species-i-Nasonia%20vitripennis.hgsc?pageLocation=Nasonia%20vitripennis].
73. 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, et al: **Comparative genome sequencing of *Drosophila pseudoobscura*: chromosomal, gene, and cis-element evolution.** *Genome Res* 2005, **15**:1-18.
74. Honeybee Genome Sequencing Consortium: **Insights into social insects from the genome of the honeybee *Apis mellifera*.** *Nature* 2006, **443**:931-949.
75. 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, Boscuti D, Barnstead M, et al: **The genome sequence of the malaria mosquito *Anopheles gambiae*.** *Science* 2002, **298**:129-149.
76. Tribolium Genome Sequencing Consortium: **The genome of the model beetle and pest *Tribolium castaneum*.** *Nature* 2008, **452**:949-955.
77. International Silkworm Genome Consortium: **The genome of a lepidopteran model insect, the silkworm *Bombyx mori*.** *Insect Biochem Mol Biol* 2008, **38**:1036-1045.
78. **Daphnia Genomics Consortium.** [http://daphnia.cgb.indiana.edu/].
79. Sea Urchin Genome Sequencing Consortium: **The genome of the sea urchin *Strongylocentrotus purpuratus*.** *Science* 2006, **314**:941-952.
80. International Chicken Genome Sequencing Consortium: **Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution.** *Nature* 2004, **432**:695-716.
81. 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-Fernández 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, et al: **The amphioxus genome and the evolution of the chordate karyotype.** *Nature* 2008, **453**:1064-1071.
82. 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.
83. Huelsenbeck JP, Ronquist F: **MRBAYES: Bayesian inference of phylogenetic trees.** *Bioinformatics* 2001, **17**:754-755.
84. Jordan GE, Piel WH: **PhyloWidget: web-based visualizations for the tree of life.** *Bioinformatics* 2008, **24**:1641-1642.
85. Wohlfrom H, Schinko JB, Klingler M, Bucher G: **Maintenance of segment and appendage primordia by the *Tribolium* gene *knödel*.** *Mech Dev* 2006, **123**:430-439.
86. 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.
87. Strimmer K, von Haeseler A: **Quartet puzzling: a quartet maximum likelihood method for reconstructing tree topologies.** *Mol Biol Evol* 1996, **13**:964-969.
88. Bailey WJ, Kim J, Wagner GP, Ruddle FH: **Phylogenetic reconstruction of vertebrate hox cluster duplications.** *Mol Biol Evol* 1997, **14**:843-853.

doi:10.1186/1471-2148-10-88

**Cite this article as:** Schaeper et al.: A clustered set of three Sp-family genes is ancestral in the Metazoa: evidence from sequence analysis, protein domain structure, developmental expression patterns and chromosomal location. *BMC Evolutionary Biology* 2010 **10**:88.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
www.biomedcentral.com/submit

