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 hawaiiensis). 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 sephacyrl 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 (GGGGCGGGG) and GT/CACC-boxes (GGTGTGGGG) [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
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 Danio rerio [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 ecctognathous 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 pseudoobscura 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 (milkwed 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 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 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. 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 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 one 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. 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. 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. hawaiensis* 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 mandibulary segment and later abuts the cephalic furrow (Fig. 3f). Later a metameric (segmentally repeated) pattern emerges that might be
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 mandibulary 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/mandibulary 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. 6h). 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...
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. hawaiensis. 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 alveoli. 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 (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. 4.
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 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 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-i 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.
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 un assembled 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 T. adhaerens. This suggests 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 cnidian 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, “eumetazoa 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 fluorescence stereomicroscope using SYTOX Green 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.
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. hawaiensis* (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 prouch 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. hawaiensis* 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_GRATCDCPNC (GGC MGG CGI ACI TGY GAY TGY CCI AAY TG), Fw_CRCRCPPC (MGI TGY MGI TGY CCI AAY TG), Fw_CHV/IPGCCGK (TGY CAY RTI CCI GGI TGY GGI AA), Rev_RSDELQHR (TGI CKY TGI ARY TCR TCI SWI C), Rev_KRFMRSDHL (ARR TGR TCI SWI CKT ATR AAY 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_0078545], 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/6)) [71], N. vitripennis [72], D. melanogaster (release 5.10) [23], D. pseudoobscura (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 (GI-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. floridana (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 PhyloWidge [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 Wohlfart et al. [85], O. fasciatus in situ hybridization was done according to Liu and Kaufman [64], P. hawaiensis in situ was performed as reported by 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 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. vectens: Nematomastella vectensis v1.0. The data for the N. vectensis Hox genes can be found in the references given in the table. 

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 specific gene pairs using the genome sequence information. For D. fasciatus, T. domestica, F. candida, and P. haowaensis 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 Sp-9 resulted from priming of the given primer pair. 

Abbreviations

The abbreviations used to denominate the different species are as follows: Ag: Anoplophora gambiae; Ar: Asellus italicus; Bm: Bombyx mori; Dm: Drosophila melanogaster; Dp: Daphnia pulex; Dps: Drosophila pseudoobscura; Dr: Dania riai; Fc: Falsomia candida; Fr: Fugu rubripes; Gg: Gallus gallus; Hs: Homo sapiens; Mm: Mus musculus; Nav: Nasonia vitripennis; Nv: Nematomastella vectensis; Ox: Oncopeltus fasciatus; Ph: Parhyale hawaensis; Sp: Strongylacentrus puppuratus; Tc: Trichoplax adhaerens; Td: Tribolium castaneum; Ts: Thermobia domestica.

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 1199/1-1 to NWP).

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, Philippsen P: 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, 33:79-87.
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, Maisak FR, Tijhan 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. Philippsen 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, Pumell BA, Jäckle H: buttonhead and D-Sp1: a novel Drosophila gene pair. Mech Dev 1996, 59:S3-62.
9. Marin M, Katis A, Visher P, Gruss P, Philippsen S. Transcription factor Sp1 is essential for early embryonic development but dispensable for cell growth and differentiation. Cell 1997, 89:619-627.
10. Black AR, Jensen D, Lin SY, Azkakan-Clifford J: Growth/Cell Cycle Regulation of Sp1 Phosphorylation. Journ of Bioi Chem 1999, 274:1207-1215.
11. 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, 107:175-179.
12. Treichel D, Schöck F, Jäckle H, Gruss P, Mansouri A. m88 is required to maintain signaling during murine limb development. Genes Dev 2003, 17:2630-2635.
13. Kawakami Y, Esteban CR, Matsus T, Rodriguez-León J, Kato S, Izsúna 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.
14. Nakamura T, Unita F, De-Vega S, Wala A, Fukumoto S, Yamada KM, Yamada Y: The Krüppel-like factor epiporphin is expressed by epithelium of developing teeth, hair follicles, and limb buds and promotes cell proliferation. J Bioi Chem 2004, 279:6626-634.
15. Safe S, Abdellahim M. Sp transcription factor family and its role in cancer. Eur J Cancer 2005, 41:2388-2448.
16. Chen Y, Guo Y, Ge X, Itoh H, Watanabe A, Fujitara W, Kodama T, Aburatani H. Elevated expression and potential roles of human Sp5, a member of Sp transcription factor family, in human cancers. Biochem Biophy Res Commun 2006, 340:758-766.
17. Cohen SM, Jurgens G. Mediation of Drosophila head development by gap-like segmentation genes. Nature 1990, 346:483-485.
18. Wimmer EA, Jäckle H, Pfeifle C, Cohen SM. A Drosophila a homologue of human Sp1 is a head-specific segmentation gene. Nature 1993, 366:690-694.
19. Schöck F, Pumell BA, Wimmer EA, Jäckle H: Common and diverged functions of the Drosophila gene pair D-Sp1 and buttonhead. Mech Dev 1999, 89:125-132.
20. Aparicio S, Chapman J, Stupka E, Putnam N, Chia JM, Dehal P, Christoffels A, Pfeiffer BD, Wan KH, Doyle C, Baxter EG, Helt G, Nelson CR, et al. Whole-genome shotgun assembly and analysis of the genome of Fugu rubripes. Science 2002, 297:1301-1310.
21. Zebradish Sequencing Project. [http://www.sanger.ac.uk/Projects/Z_rerio/]
22. 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, BRac RC, Rogers YH, Blazek R, Chumpe 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.
23. Beermann A, Aranda M, Schroder R: The Sp zinc-finger transcription factor is involved in allometric growth of the limbs in the beetle Tribolium castaneum. Development 2004, 131:733-742.
24. Tallofuss F, Wilm HF, Crozier D, Mepfert P, Wassef M, Bally-Cuif L. The zebrafish buttonhead-transcription factor Bst1 is an early regulator of par2 expression during mid-hindbrain development. Development 2001, 128:4021-4034.
25. Knaufenb F: Insecta (Hexapoda), Insekten. Spezielle Zoologie. Teil 1: Einteller und wirbellose Tiere. Munich: Spektrum Akademischer Verlag, ElsevierWestheide W, Rieger R, Second 2007, 638-654.
Schaeper ND, Prpic NM, Wimmer EA: Are the insects terrestrial crustaceans? A discussion of some Myriapod-insect relationships as opposed to an insect crucasean sister group relationship. Arthropod Relationships London: Chapman & HallFortey RA, Thomas RH 1997: 305-315.

Dohle W: Myriapod-insect relationships as opposed to an insect crucasean sister group relationship. Arthropod Relationships London: Chapman & HallFortey RA, Thomas RH 1997: 305-315.

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 Crucasea plus Hexapoda. Ann Soc Entomol France NS 2001, 37:85-103.

Budd GE, Telford MJ: The origin and evolution of arthropods. Nature 2009, 457:812-817.

Bouwman P, Philipsen S: Regulation of the activity of Sp1-related transcription factors. Mol Cell Endocrinol 2002, 52:27-38.

Schinko JB, Kreuser N, Offen N, Posien 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.

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

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

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

Gollner H, Dari C, Phillips B, Philipsen S, Suske G: Impaired ossification in mice lacking the transcription factor Sp3. Mech Dev 2001, 107:78-83.

Schierwater B, Dellaporta SL, Rokhsar DS: The early ANTP gene repertoire: insights from the placozoan genome. BMC Biol 2003, 30:929-954.

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

Harrison SM, Houzellestein D, Dunwoodie SL, Bedington 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.

Schaeper ND, Pripp NC, Wimmer EA: A conserved function of the zinc finger transcription factor Sp6/9 in allometric appendage growth in the milkweed bug Oncopeltus fasciatus. Dev Genes Evol 2009, 219:427-439.

Schoy S, Gabant P, Van Reeth T, Hertvelt V, Drzez PL, Van Vooren P, Rivelle M, Szpirer J, Szpirer C: Identification of Klfl3 and Klfl4 (Sp6b), novel members of the Sp6/klfl transcription factor family. Genomics 2000, 70:93-101.

Hertvelt V, Louyan S, van Reeth T, Drzez 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.

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.

Milona MA, Gough JE, Edgar AJ: Expression of alternatively spliced isoforms of human Sp5 in osteoblast-like cells. BMC Genomics 2003, 4:43.

Kalback LA, Sounig do Y, Nair A, Smith N, Schwarz EM, O'Keefe RJ, Dinsie H: Osterix/Sp5 regulates mesenchymal stem cell mediated endothelial ossification. J Cell Physiol 2008, 214:173-182.

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

Griesel G, Treczeli P, Collombat P, Kavli J, Zembrzuk A, Akker van den WM, Guiss P, Simeone A, Mansouri A: Sp8 controls the anteroposterior patterning at the midbrain-hindbrain border. Development 2006, 133:1779-1787.

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

Putnam NH, Srivastava M, Hellisten U, Dirks B, Chapman J, Salamon A, Terry A, Shapiro H, Lindquist E, Kapitomov W, Jukka J, Genrikhov G, Grigoriev N, Lucas SM, Steele RE, Finnerty JR, Techau U, Martindale MQ, Rokhsar DS: Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. Science 2007, 317:86-94.

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

Shemild SM: C2H2 zinc finger gene of the GlI, Zf, KL, 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.

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.

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.

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, 218:576-590.

Holland LZ, Albaltar R, Asumi K, Benito-Gutierrez E, Blow MJ, Bronner-Fraser M, Brunet F, Butts T, Candiani S, Dishaw LJ, Ferrier DE, Garcia-Fernandez J, Gibson-Brown JJ, Giss C, Godzik A, Halboekk F, Hirose D, Hormesoki C, Ikuta T, Irokio H, Kasahara M, Kasamatsu K, Kawashima T, Kimura A, Kobayashi M, Kozmik Z, Kubokawa K, Lauter V, Litman GW, McHardy AC, et al: The amphioxus genome illuminates vertebrate origins and cephalochordate biology. Genome Res 2008, 18:1100-1111.

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

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

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

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

Taylor JS, Peer Van der Y, Braaich 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.

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

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

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

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.

Browne WE, Schmid BM, Wimmer EA, Martindale MQ: Expression of otd orthologs in the amphipod crustacean, Parhyale hawaiensis. Dev Genes Evol 2006, 216:581-596.

Browne WE, Price AL, Gerberding M, Paten NH: Stages of embryonic development in the amphipod crustacean, Parhyale hawaiensis. 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, Adams PD, et al: The genome sequence of Drosophila melanogaster. Science 2000, 287:1763-74.

71. Mouse Genome Sequencing Consortium: Initial sequencing and comparative analysis of the mouse genome. Nature 2002, 420:520-523.

72. Nasonia Genome Project. [http://nasonia.hgsb.tmi.cshl.edu/project-species-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, Bussmaker HJ, van Batenburg MF, Howells SL, Scherer SE, Sodergren E, Matthews BB, Crosby MA, Schroeder AJ, Ortiz-Banetats D, Rives CM, Metzker ML, Muzny DM, Scott G, Steffen D, Wheeler DA, et al: Comparative genome sequencing of Drosophila pseudoobscura: chromosome assembly 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, 443:911-949.

75. Holt RA, Subramanian GM, Halpern A, Sutton GG, Charlab R, Nusskern DR, Wendl MC, Wincker P, Clark AG, Lander ES, Mikkelsen TS, et al: Initial splicing and transcript structure analysis of the mouse genome. Nature 2004, 428:807-815.

76. Tribolium Genome Sequencing Consortium: The genome of the model beetle and pest Tribolium castaneum. Nature 2006, 442:494-495.

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. Nature 2004, 432:695-716.

81. Putnam NH, Bluts T, Ferrier DE, Furlong RF, Heilman U, Kawashima T, Robinson-Rechavi M, Shuguchi I, Terry A, Yu JK, Benito-Gutiérrez E, 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, Satou Y, Schaeper U, Schmutz J, Shin-I T, et al: The amphioxus genome and the evolution of the chordate karyotype. Nature 2008, 453:1064-1071.

82. Blass C, Bolanos R, Boscus D, Barnstead M, Atkinson PW, Baden H, de Berardinis V, Baldwin D, Benes V, Biedler J, Blass C, Bolanos R, Boscus D, Barnstead M, et al: The amphioxus genome and the evolution of the chordate karyotype. Nature 2008, 453:1064-1071.

83. Hultén ER, Westergren TH, Jonsson G: The classification of the Crustacea, volume 1. Almqvist & Wiksell, Stockholm; 1951.

84. 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.