**Int. J. Dev. Biol. 61:** 723-732 (2017)
doi: 10.1387/ijdb.170205is

THE INTERNATIONAL JOURNAL OF DEVELOPMENTAL BIOLOGY
www.intjdevbiol.com

# Amphioxus *Sp5* is a member of a conserved Specificity Protein complement and is modulated by Wnt/β-catenin signalling

SIMON C. DAILEY[1,2], IRYNA KOZMIKOVA[3] and ILDIKÓ M.L. SOMORJAI*[,1,2]

*[1]University of St Andrews, Biomedical Sciences Research Complex, North Haugh, St Andrews, UK, [2]Scottish Oceans Institute, East Sands, St Andrews, UK and [3]Institute of Molecular Genetics of the Czech Academy of Sciences, Prague, Czech Republic*

**ABSTRACT A cluster of three *Specificity Protein (Sp)* genes *(Sp1-4, Sp5* and *Sp6-9)* is thought to be ancestral in both chordates and the wider Eumetazoa. Sp5 and Sp6-9 gene groups are associated with embryonic growth zones, such as tailbuds, and are both Wnt/β-catenin signalling pathway members and targets. Currently, there are conflicting reports as to the number and identity of *Sp* genes in the cephalochordates, the sister group to the vertebrates and urochordates. We confirm the SP complement of *Branchiostoma belcheri* and *Branchiostoma lanceolatum*, as well as their genomic arrangement, protein domain structure and residue frequency. We assay *Sp5* expression in *B. lanceolatum* embryos, and determine its response to pharmacologically increased β-catenin signalling. *Branchiostoma* possesses three *Sp* genes, located on the same genomic scaffold. Phylogenetic and domain structure analyses are consistent with their identification as SP1-4, SP5 and SP6-9, although SP1-4 contains a novel glutamine-rich N-terminal region. SP5 is expressed in axial mesoderm and neurectoderm, and marks the cerebral vesicle and presumptive pharynx. Early exposure to increased β-catenin caused ubiquitous SP5 expression in late gastrula, while later treatment at gastrula stages reduced SP5 expression in the posterior growth zone during axis elongation. Amphioxus possess a typical invertebrate eumetazoan SP complement, and SP5 expression in embryos is well conserved with vertebrate homologues. Its expression in the tailbud, a posterior growth zone, is consistent with expression seen in other bilaterians. *Branchiostoma* SP5 shows a dynamic response to Wnt/β-catenin signalling.**

KEY WORDS: *Specificity protein, Brachyury, Branchiostoma, Wnt, Sp5*

## Introduction

The *Specificity protein* (SP) genes are important developmentally expressed transcription factors, found across almost all metazoan species with the exception of ctenophores (Presnell *et al.,* 2015). Although the SP complement can vary significantly among species, its members can be categorised into three distinct clades: SP1-4, SP5 and SP6-9 (Schaeper *et al.,* 2010). Expression of *Sp5* and *Sp6-9* genes is typically associated with embryonic regions of high proliferation and growth, such as tail or limb buds. For example, vertebrate *Sp5* marks the primitive streak of mice in early development, and later the tailbud for the entire period of axis elongation (Treichel *et al.,* 2001). Similarly, zebrafish *Sp5* is expressed in both the tailbud and somites (Tallafuβ *et al.,* 2001). Expression domains for *Sp6, Sp8* and *Sp9* all include the apical-ectodermal-ridge (AER) of developing limb buds (Nakamura *et al.,* 2004; Kawakami *et al.,* 2004). Vertebrate *Sp7* is perhaps atypical, as it specifically marks osteoblasts (Nakashima *et al.,* 2002), but for the remaining vertebrate *Sp5* and *Sp6-9* genes, an association with embryonic growth zones is a defining feature of these two groups. This is also true outside of the vertebrates. For example, insect *Sp6-9* genes are frequently expressed in the brain, ventral

*Abbreviations used in this paper:* 2R, two rounds of vertebrate whole genome duplication; Azk, 1-Azakenpaullone; bp, base pairs; DiMes, di-mesencephalic primordium; HyPTh, hypothalamo-prethalamic primordium; Kb, kilobase; KLF, krüppel-like Factor; Mb, megabase; SP, Specificity Protein; RhSp, rhombencephalo-spinal primordium; PFA, paraformaldehyde; WMISH, whole mount *in situ* hybridisation.

***Address correspondence to:** Ildikó M.L. Somorjai. University of St Andrews, Biomedical Sciences Research Complex, North Haugh, St Andrews, KY16 9ST, United Kingdom. Tfno: 01334 463628. E-mail: imls@st-andrews.ac.uk - web: https://synergy.st-andrews.ac.uk/cord http://orcid.org/0000-0001-5243-6664

**Supplementary Material** (six figures and one table) for this paper is available at: http://dx.doi.org/10.1387/ijdb.170205is

*Submitted:* 19 August, 2017; *Accepted:* 19 September, 2017

ISSN: Online 1696-3547, Print 0214-6282

nerve cord and the limb primordial/buds of both antennae and legs (Schaeper *et al.,* 2010), and *Sp8* is able to regulate allometric growth of limbs in *Tribolium castaneum* and *Oncopeltus fasciatus* (Schaeper *et al.,* 2009; Beermann *et al.,* 2004). Similarly, insect *Sp5* is expressed in the imaginal discs of *Drosophila*, as well as the somites and posterior growth zone of *T. castaneum* (Schaeper *et al.,* 2010). Given the functional similarity of these genes both within and outside the vertebrate lineage, it would be reasonable to hypothesise that the utilisation of *Sp5* and *Sp6-9* in embryonic growth zones is a conserved bilaterian trait.

Interactions with Wnt/β-catenin signalling appear to be crucial for the roles of *Sp5* and *Sp6-9* in these embryological growth zones. For example, SP8 is an effector of Wnt/β-catenin signalling in vertebrate limb outgrowth (Kawakami *et al.,* 2004). In zebrafish, *Sp5* and *Sp5-like* are direct targets of Wnt8/β-catenin signalling during the patterning of the neurectoderm and mesoderm (Weidinger *et al.,* 2005), and *Sp5* is upregulated in the mouse telencephalon following Wnt/β-catenin activation (Fujimura *et al.*, 2007). *Sp5-like* is also a target of Wnt8 and Wnt3a in the tailbud, and is required for somitic mesoderm and notochord formation (Thorpe *et al.,* 2005). There are numerous examples, therefore, of *Sp5* and *Sp6-9* acting as downstream targets of Wnt/β-catenin signalling in vertebrate embryological growth zones. Furthermore, SP5 and SP8 are also critical components of the canonical Wnt signalling pathway itself: both proteins are components of the TCF/LEF recruitment complex for some Wnt target genes, including *Brachyury* (Kennedy *et al.,* 2016). This relationship to *Brachyury* is important during development, as whilst knocking out mouse *Sp5* alone causes no significant embryological defects, it does significantly increase the penetrance of *Brachyury (T)* knockout. These mutants develop malformed thoracic and lumbar vertebrae, and develop significantly fewer vertebrae in the tail (Harrison *et al.,* 2000). This demonstrates the involvement of *Sp5* in facilitating Wnt/β-catenin and *Brachyury*-mediated proliferation in the vertebrate tail.

The distinction between *Sp5* and *Sp6-9* genes, despite their considerable similarities during development, arise from the prediction of a three-gene *Sp* cluster in the eumetazoan ancestor (Schaeper *et al.,* 2010). This cluster, still present in the cnidarian *Nematostella vectensis* (original data from Putnam *et al.,* 2007*)*, discussed in Schaeper *et al.,* 2010)), contains *Sp1-4, Sp5* and *Sp6-9* homologues. Although the SP complement of extant metazoan species ranges between 1 to 13 genes (Presnell *et al.,* 2015), phylogenetic analysis supports the existence of these three clades (Presnell *et al.,* 2015; Schaeper *et al.,* 2010; Pei & Grishin 2015), which take their names from the human SP complement. The vertebrate *Sp* genes are noted for their linkage to the four vertebrate Hox clusters, leading to the conclusion that an ancestral three-gene *Sp* cluster was located on the same chromosome as the vertebrate Hox cluster, immediately before the two rounds of whole genome duplication (2R) at the base of the vertebrate lineage (Schaeper *et al.,* 2010).

In order to support proposed models of *Sp* gene evolution, a number of studies have performed domain and amino acid frequency analyses on SP proteins. As a gene family, they are defined by

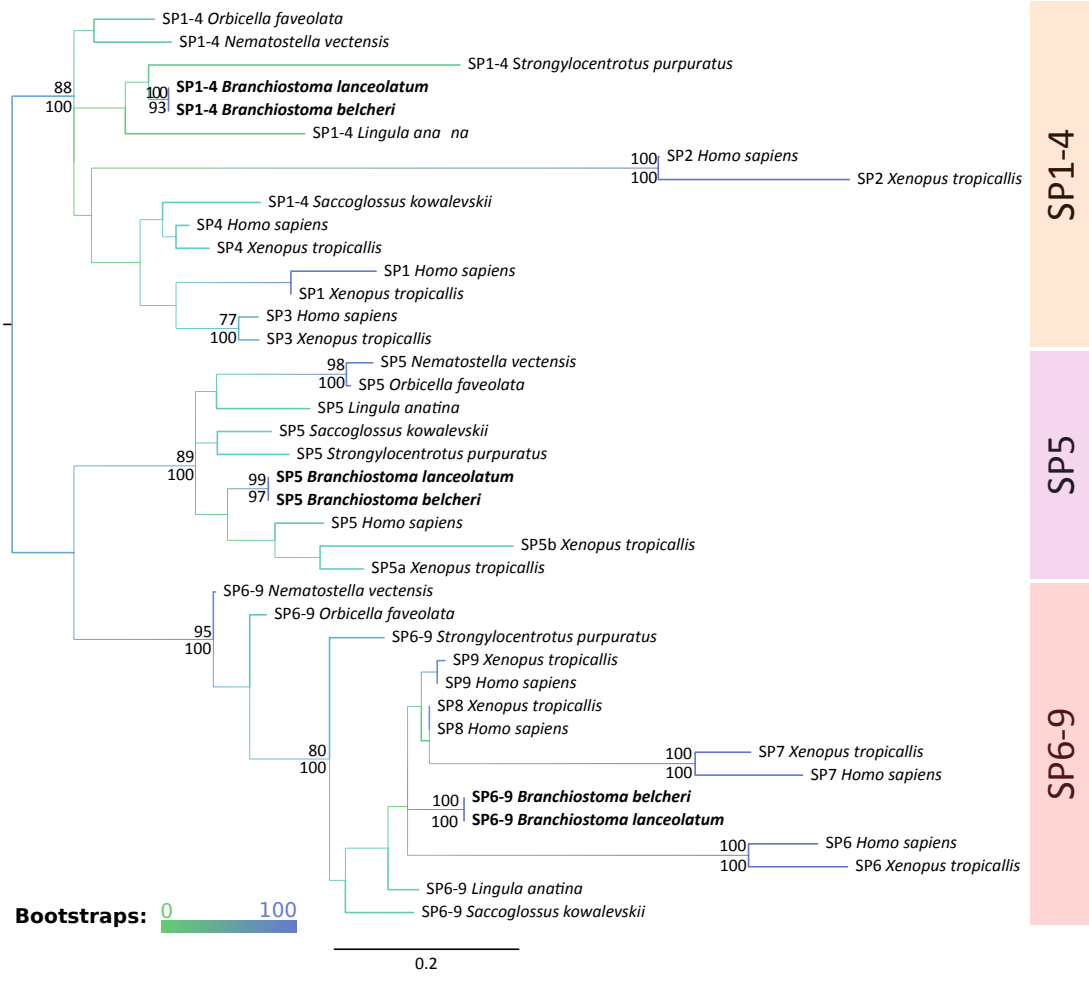


**Fig. 1. Deuterostome Specificity Protein (SP) phylogeny.** *Three distinct clades of deuterostome SP proteins were found: SP1-4, SP5 and SP6-9, which take their names from human SP proteins. The cnidarian, cephalochordate, echinoderm and hemichordate lineages shown here each possess a single SP protein per clade. Midpoint-rooted RAxML tree, labelled with percentage RAxML bootstraps (above node) and MrBayes posterior probabilities (below node) for nodes with 70% or higher support from both algorithms. Trees used the Dayhoff model of protein evolution with 4 gamma rate categories. Tree branches are coloured according to their RAxML bootstrap support (green low, blue high). Cephalochordate sequences are shown in bold.*

their possession of a Btd domain (Bouwman & Philipsen 2002; Wimmer *et al.,* 1993), which is positioned shortly before a triplet of C2H2 zinc finger domains (interspaced by conserved linker regions), which all together are located close to the protein's C-terminus (Iuchi 2001). This triplet of C2H2 domains, including their conserved linker regions, is also present in Krüppel-like factors (KLFs) and Krox/Egrs (Iuchi 2001). Outside of these conserved domains, however, SP proteins can vary substantially, both within and among species. As such, a number of previous studies have analysed the residue profiles of regions outside of the C2H2 and Btd domains to support models of *Sp* gene evolution (Schaeper *et al.,* 2010; Bouwman & Philipsen 2002; Wimmer *et al.,* 1993).

Given the considerable amount of data available for characterising SP proteins, and the convincing model for the existence of a three-gene *Sp* cluster prior to the vertebrate 2R, there is surprising inconsistency surrounding the SP complement of invertebrate chordates. Specifically, the cephalochordates, sister group to the vertebrate and urochordate clade, have been reported to possess either SPa, SPb and SP5 (Shimeld 2008), SP1-4a, SP1-4b, SP5 and SP6-9 (Presnell *et al.,* 2015) or SP1-4, SP5, SP6-9 (Pei & Grishin 2015) based on analyses in *Branchiostoma floridae*. As a chordate lineage, *Branchiostoma* is typically considered to be evolutionarily conservative, and so lineage-specific changes to their SP complement would warrant further study. Cephalochordates are also a model for chordate tail elongation, employing many of the same developmental signalling strategies with vertebrates, including the expression of Wnt ligands and *Brachyury* in the tailbud (Holland 2002). We might therefore expect *Sp5* and *Sp6-9* to interact with Wnt signalling in the amphioxus tailbud.

Previous research on *Sp* genes has used *Branchiostoma floridae* as the sole cephalochordate representative (Shimeld 2008; Presnell *et al.,* 2015; Pei & Grishin 2015). Here, we confirm the SP complement of the *B. lanceolatum* and *B. belcheri* species, through both phylogenetic and domain structure analysis. We also demonstrate the expression of *Sp5* in the *B. lanceolatum* tailbud, and document the responses of this *Sp5* and *Brachyury2* expression to pharmacological manipulation of β-catenin levels. Our data suggest that both species possess a typical invertebrate chordate SP complement, with single copies of *Sp1-4, Sp5* and *Sp6-9.* Furthermore, in both species SP1-4 contains a novel, Glutamine-rich expansion region between the C2H2 triplet and C-terminus. *Sp5* in *B. lanceolatum* neurulae and early larvae marks several tissues, including the somites, neural tube and tailbud. We demonstrate here that pharmacologically increasing Wnt/β-catenin signalling in early developmental stages caused the ectopic expression of *Sp5.* In contrast, treatment at gastrula stage led to the subsequent loss of *Sp5* expression in the posterior growth zone of neurulae and larvae.

## Results

### *The* Sp *gene complement of* Branchiostoma

Three SP-like proteins were identified in the predicted protein sets of both the *Branchiostoma lanceolatum* and *B. belcheri* genomes. These each contained the characteristic buttonhead (Btd) box and C2H2 zinc finger triplets (Supplementary Fig. S2). In order to confirm the identity of these three proteins, the C2H2 triplets from these proteins were aligned against the same domain set of metazoan SP and KLF proteins (Supplementary File 1), as well as human Wilms Tumour 1 (Wt1) as an outgroup. Phylogenetic analysis using both RAxML and MrBayes (Supplementary Fig. 3) supported their positioning among a well-supported clade of the SP proteins. However, the RAxML analysis was not able to provide strong support values for numerous well supported clades in the MrBayes analysis, including the internal topology of the SP clade. We therefore took this subset of SP proteins and aligned both the Btd domain, C2H2 triplet and short intervening sequence (Supplementary File 2). From this analysis (Fig. 1), both RAxML and MrBayes produced strong support values for the existence of three clades, which correspond to the previously observed grouping of eumetazoan SP proteins into SP1-4, SP5 and SP6-9 clades. All invertebrates examined, including the two *Branchiostoma* species included here, possessed a single representative of each clade.

Outside of the Btd and C2H2-triplet domains, the SP proteins differ significantly. Several previous studies have characterised these differences, identifying both the presence of an additional region, the SP box, which is inconsistently conserved among SP genes, as well as regions enriched for specific residues (Schaeper *et al.,* 2010; Bouwman & Philipsen 2002; Wimmer *et al.,* 1993). The same analysis was performed here in order to further support the phylogenetic gene assignations (Fig. 2). Residue frequency analysis of the regions between the N-terminus and the Btd box of each protein indicated that SP1-4 is enriched for Q and T residues, SP5 for P, S and A residues, and SP6-9 for S residues. Despite C2H2 triplets usually being observed near the C-terminus of SP and KLF proteins (Iuchi 2001), in SP1-4 they are closer to the N-terminus, a finding supported across its full length by transcriptomic data in *Branchiostoma lanceolatum* (Dailey and Somorjai, unpublished) and EST data in *B. belcheri* (data from genome website (Huang *et al.,* 2014)). This trailing region of approximately 480 residues is enriched for both Alanine and Glutamine, the latter of which comprises around 25% of the sequence. These residues are regularly spaced, with no repeat longer than $(Q)_3$ (highlighted in Supplementary Fig. 4).

Previous reports indicate that these three *Sp* genes existed in clades as distant as the cnidarian *Nematostella vectensis*, in

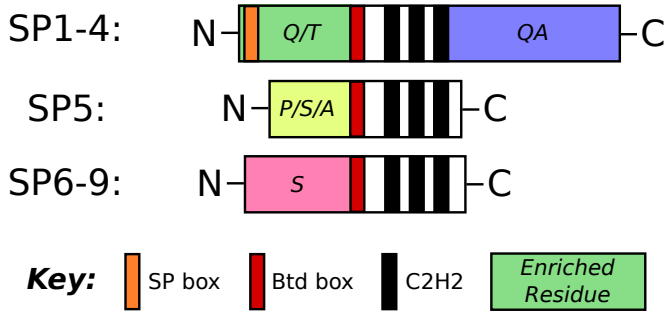


**Fig. 2. Structural comparison of *Branchiostoma* SP proteins.** *Comparison of domain presence and residue frequency in* Branchiostoma *SP proteins. Conserved SP-family domains are highlighted including the SP box, Btd box and C2H2 zinc finger domains. Sequences between the N-terminus and Btd box were analysed for residue frequency, and any enriched residues (where a single residue forms >10% of the total sequence) are indicated. The same analysis was performed for the region between the final C2H2 and C-terminus in SP1-4, but not for the remaining two SP proteins, as these regions were too short to characterise in this manner. Total protein lengths are to approximate scale.*

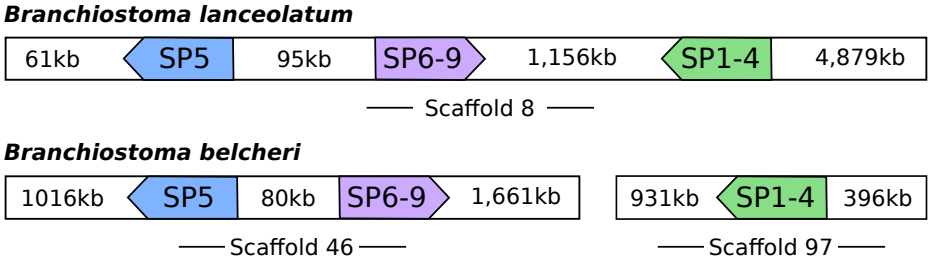


**Fig. 3. Genomic arrangement of the *Branchiostoma Sp* genes.** *Diagram showing the genomic arrangement of* sp *genes in* B. lanceolatum *and* B. belcheri. *All three* B. lanceolatum *genes are found on the same scaffold, albeit with significant distances between genes, especially in the case of* Sp1-4. *In* B. belcheri, Sp1-4 *is located on a separate scaffold to the other two* Sp *genes. Arrows depict gene loci and their orientations, and between these are genomic distances in kilobase pairs (kb). Scaffold numbers are given for the respective genome projects.*

which they are arranged in a cluster, and that these genes were still positioned on the same chromosome in the vertebrate lineage at the time of the 2R events. To investigate if such a cluster is still present in *Branchiostoma*, the genomic loci for their three *Sp* genes were determined. In *B. lanceolatum*, all of these three genes are located on the same genomic scaffold (Fig. 3), albeit with a gap of almost 1.2 megabases (Mb) between *Sp1-4* and the remaining two genes. In contrast, only *Sp5* and *Sp6-9* are positioned on the same scaffold in *B. belcheri*, which are located over 1Mb away from the end of their scaffold in both directions. *Sp1-4* was found on a separate scaffold, located 400kb from the closest scaffold end.

### *Developmental expression of* Sp5

Thus far, there are no published descriptions of *Sp* gene expression patterns in any invertebrate chordate species. We were, however, able to identify expression of all three *B. lanceolatum Sp* genes in the combined developmental and adult transcriptome available for *B. lanceolatum* (Oulion *et al.*, 2012), suggesting that they may be developmentally expressed. A 692bp region of *Sp5* was cloned from *B. lanceolatum* adult cDNA, and used to assay *B. lanceolatum* embryos via WMISH. In order to identify the earliest point of *Sp5* expression, a range of developmental stages from fertilised egg to gastrula was assayed. The earliest expression detected here was in the late blastula stage, where a hemisphere of expression is seen, centred over the macromeres (Supplementary Fig. 5). Expression continues into all later stages assayed (Fig. 4), from the G4 gastrula to L1 larvae. In G4 grastrulae (Fig. 4A), staining is found in posterior ectoderm, mesendoderm and at the dorsal blastopore lip, but is notably absent from the entire anterior. In N0 neurulae, (Fig. 4B) staining occurs in the neural plate, dorsal blastopore lip and ventral endoderm. In N1 neurulae (Fig. 4C) expression is found in the anterior tailbud, patchily throughout the notochord, in the ventral endoderm lining the archenteron, and in the anterior and posterior of the neural tube. By the N2 stage (Fig. 4D), staining is also visible in the somites. This general pattern continues throughout the remaining neurula stages, although from N3 stage onwards (Fig. 4E) expression in the ventral endoderm is only found in the posterior (marking the chordoneural hinge and hindgut) and the anterior of the embryos. In N4 stage embryos (Fig. 4F), the staining in the anterior labels the cerebral vesicle and regions within the presumptive pharynx, and expression is seen throughout the entire tailbud. By L1 (Fig. 4G) expression in the notochord is also weak in all but the anterior

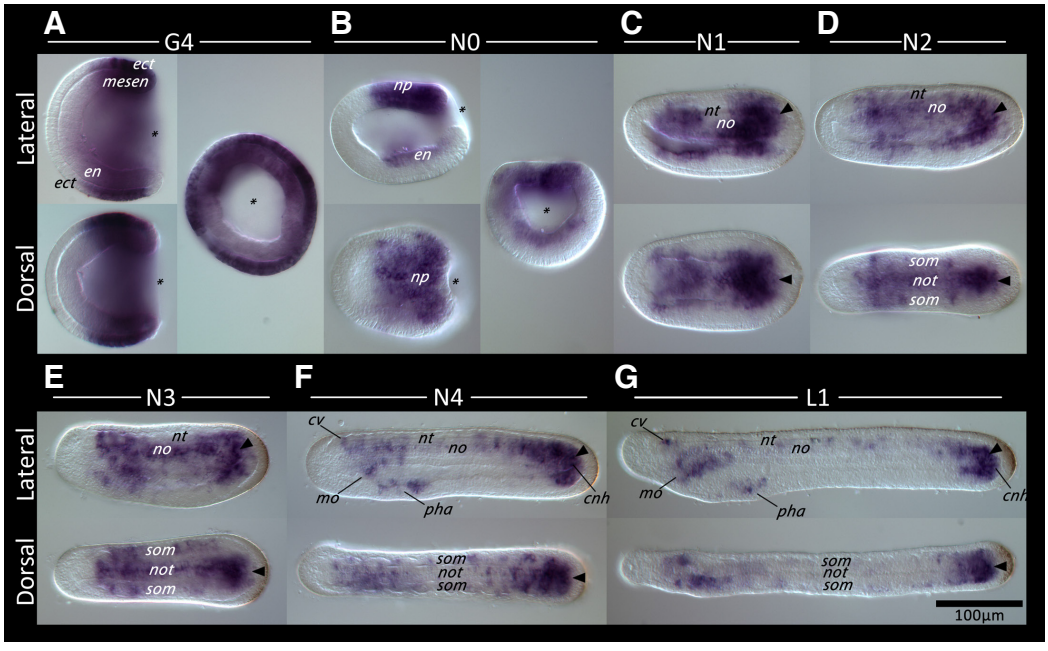


**Fig. 4. Expression of *Sp5* in *Branchiostoma lanceolatum* embryos.** *Developing* B. lanceolatum *embryos were assayed for expression of* Sp5 *between G4 stage and L1 stage by WMISH. All embryos are orientated with the anterior to the left, except for the third panels of* **(A,B)**, *which show blastoporal views with the dorsal surface to the top. Asterisks mark the blastopore, and arrowheads mark the tailbud. In G4 stage gastrulae* **(A)**, *staining is found in most of the ectoderm and the mesendoderm, with the exception of the anterior, but by N0 stage neurulae* **(B)** *ectodermal expression is found only in the neural plate. In N1-4 stage neurulae* **(C-F)**, *expression in neural tissues and ventral endoderm is found in the anterior and in the posterior, but not in between. Expression is seen throughout the somites, albeit with inconsistent penetrance, and throughout the length of the notochord. In late N4 neurulae and early L1 larvae* **(F-G)**, *expression strongly marks the tailbud and posterior-most somitic, notochord and neural tissues as well as posterior ventral endoderm. Anteriorly,* Sp5 *is expressed in the cerebral vesicle and presumptive pharynx. Abbreviations used: cnh, chordoneural hinge; cv, cerebral vesicle; ect, ectoderm; en, endoderm; mesen, mesendoderm; mo, mouth; no, notochord; np, neural plate; nt, neural tube; pha, pharynx; som, somites.*

and posterior, and somite expression is relatively weak with variable penetrance. Throughout all neurula stages and L1 larval stages, a strong domain of expression is seen overlapping the tailbud, as well as the neural and mesodermal tissues immediately anterior to this structure, and ventral posterior endoderm.

### *Modulation of β-catenin levels regulates* Sp5 *expression*

The identification of *Sp5* in the *B. lanceolatum* tailbud matches observations of this gene in the vertebrate tailbud. As reviewed in the introduction, vertebrate *Sp5* acts as a component of the Wnt signalling pathway for certain Wnt/β-catenin downstream targets, such as *Brachyury/T,* and appears to be strongly linked to its ability to drive proliferation in the vertebrate tailbud. Amphioxus also utilises a homologous tailbud structure, which progressively extends axial tissues of the embryo during axis elongation. Interestingly, several of the *Sp5* expression domains observed here, including the tailbud, overlap with previously described expression patterns for numerous Wnt ligands, as well as both *Brachyury* genes in *Branchiostoma* (Holland *et al.,* 1995; Holland 2002; Bertrand *et al.,* 2017). It is possible, therefore, that the *Sp5* and *Brachyury* genes of amphioxus also interact with each other, or with Wnt/β-catenin signalling, in the tailbud. To test this, *B. lanceolatum* embryos were treated with 1-Azakenpaullone (Azk), an inhibitor of GSK3-β that acts to increase the nuclear accumulation of β-catenin (Kunick *et al.,* 2004), mimicking increased Wnt/β-catenin signalling. Embryos were treated with 10μM Azk at the G4 stage, well before the formation of the tailbud, and then assayed using WMISH for *Sp5* and *Brachyury2* expression at N1 and L1 stages. At N1 stage, a characteristic horseshoe-shaped domain of *Brachyury2* expression can be seen overlapping the tailbud; this domain was significantly expanded as a result of treatment (Fig. 5A). Most domains of *Sp5* expression (Fig. 5B) were not obviously affected by Azk treatment. However expression overlapping the *Brachyury2* horseshoe-shaped domain is strongly reduced in treated embryos. In L1 embryos, *Brachyury2* expression exhibits a similar response to treatment (Fig. 5C), with the posterior expression domain overlapping the tailbud expanding significantly. *Sp5* again acts conversely (Fig. 5D), with expression decreasing in this posterior expression domain.

The response of *Sp5* expression to increased Wnt/β-catenin signalling was surprising, especially in comparison to the response of *Brachyury2* expression. To confirm if these contrasting responses were reproducible under other treatment conditions, embryos were treated with 1-Azakenpaullone at blastula stage, earlier than the previous treatment. These were then assayed for expression of these two same genes at N1 stage (Fig. 6). At this stage, *Brachyury2* is usually absent from the ectoderm (Fig. 6A and Terazawa and Satoh 1997), with the exception of a ring of expression around the blastopore lip. 1-Azakenpaullone treatment however caused ectopic expression of *Brachyury2* throughout the entire ectoderm. A similar response to treatment was seen for *Sp5* expression (Fig. 6B): *Sp5* is detected throughout the anterior-most ectoderm and endoderm of treated embryos, whereas this region is consistently free of expression in control and wildtype embryos at N0 and G4 stage embryos, respectively.

In light of the now contrasting responses of *Sp5* to GSK3-β inhibition with 1-Azakenpaullone, a third treatment was performed with an alternate GSK3-β inhibitor (CHIR99021) from 8-cell stage to G4 gastrula stage (Supplementary Fig. 6A). This treatment again resulted in ubiquitous *Sp5* expression (Supplementary Fig. 6B), in contrast to control embryos where no expression is detected in the anterior of the embryo. Quantitative RT-PCR of these embryos showed a marked increase in *Sp5* expression following treatment (Supplementary Fig. 6C), indicating that GSK3-β inhibition increases *Sp5* expression under these conditions.

## Discussion

### *Amphioxus possesses a typical invertebrate chordate SP complement*

The current hypothesis is that of a eumetazoan ancestor possessing a clustered set of three SP proteins: SP1-4, SP5 and SP6-9 (see introduction). Central to these evolutionary models is the assumption that such a cluster existed in the vertebrate ancestor, and by extension the chordate ancestor, with subsequent expansions in various vertebrate lineages (Schaeper *et al.,* 2010). However, genomic data from *Branchiostoma floridae* resulted in conflicting conclusions surrounding the number and identity of these genes

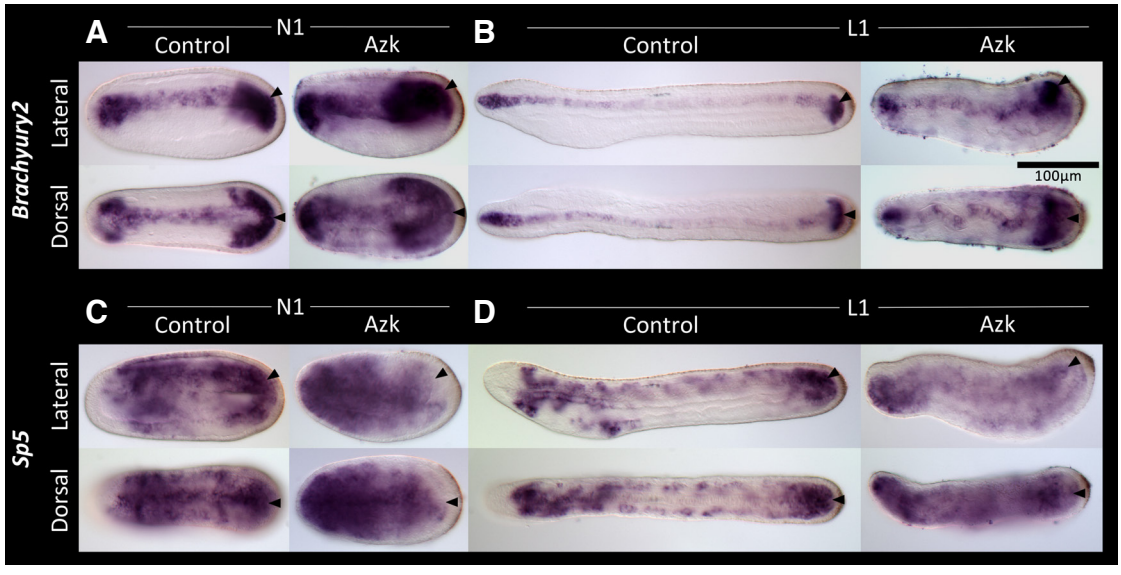


**Fig. 5. 1-Azakenpaullone treatment expands the *Brachyury2* expression domain overlapping the tailbud, but reduces that of *Sp5*.** B. lanceolatum *embryos were treated with 10μM 1-Azakenpaullone (Azk) at the G4 stage of development.* Brachyury2 *expression reveals that the tailbud (marked by the black arrowhead) of treated embryos is considerably larger at both N1* **(A)** *and L1* **(B)** *stages of development. However,* Sp5 *staining is strongly reduced in the tailbud in both instances* **(C,D)**. *All embryos are orientated with the anterior to the left and dorsal up, and all images are to scale.*

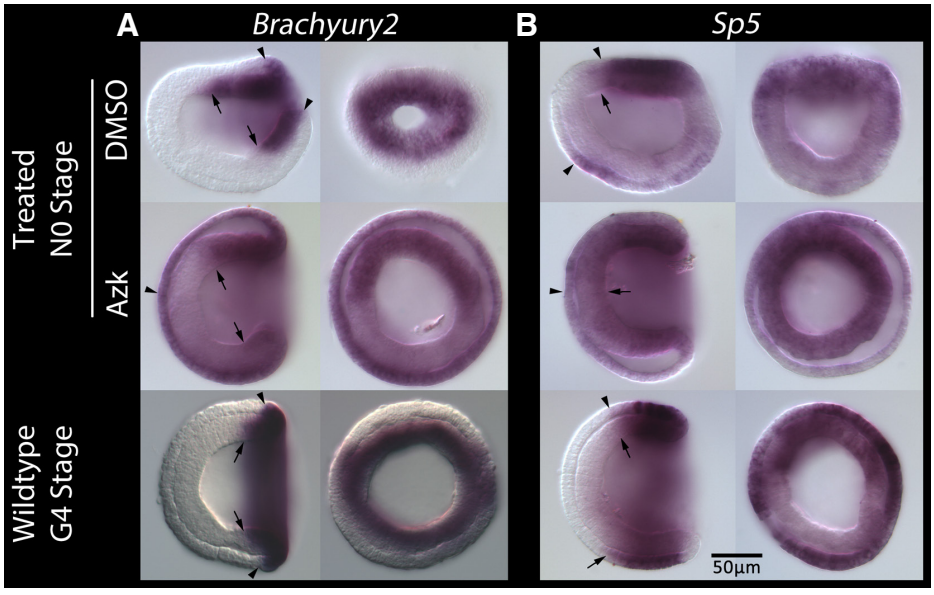


**Fig. 6. 1-Azakenpaullone treatment causes ectopic expression of both *Brachyury2* and *Sp5* anteriorly.** *Embryos treated from blastula stage with 1-Azakenpaullone (Azk) were assayed for the expression of* brachyury2 *and* Sp5 *at N0 stage. The gross morphology of treated embryos resembles that of wildtype G4 embryos, shown below for comparison. The anterior-most boundaries of dorsal and ventral expression are marked with arrowheads in the ectoderm and arrows in the endoderm.* **(A)** Brachyury2 *expression is detected throughout the entire ectoderm of treated embryos, in contrast to control embryos where ectodermal expression is only seen at the very boundary of the blastopore lips.* **(B)** Sp5 *expression in treated embryos is seen throughout the entire mesoderm and ectoderm, whereas in control embryos expression is absent in the anterior-most ectoderm or endoderm. All embryos are to the same scale.*

in the basal chordate lineage, the cephalochordates. We therefore analysed genomic datasets from two additional species, *B. belcheri* and *B. lanceolatum,* in order to resolve this confusion. Our results indicate the presence of three *Branchiostoma* SP-like proteins, which grouped strongly along with other eumetazoan SP proteins in a KLF/SP phylogenetic analysis. Further support for this assignation was found when considering domain presence, as all three possessed the Btd and C2H2 domain triplets that are defining of this protein family (Iuchi 2001). It appears therefore that the typical *Branchiostoma* condition is the possession of three *Sp-like* genes.

Phylogenetic analysis of deuterostome SP proteins provided initial support for the identity of the three *Branchiostoma* SP proteins as SP1-4, SP5 and SP6-9. To confirm this finding, we examined the residue frequency of these proteins outside of the highly conserved Btd and C2H2 domains. In the regions between the N-terminus and Btd domain (the 'pre-Btd region'), we find that SP1-4 is enriched for Q and T, which is consistent with vertebrate, cnidarian and arthropod states (Schaeper *et al.,* 2010). The same is true of our observation of P and A residues for *Branchiostoma* SP5 and SP6-9*:* although variable, they are frequently observed in these two groups of genes among the same set of species (Schaeper *et al.,* 2010). Together, these findings indicate that *Branchiostoma* possess a typical eumetazoan three *Sp*-gene complement.

Our domain structure analysis of *Branchiostoma* SP proteins also identified the inconsistently conserved SP box in *Branchiostoma* SP1-4. This domain is present in all three SP groups in humans (with only SP6 and SP9 lacking this domain) (Schaeper *et al.,* 2010), and thus the absence of this domain in both *Branchiostoma* SP5 and SP6-9 must represent lineage specific losses. The function of this domain has never been experimentally proven. However, the SP box possesses an endoproteolytic cleavage site, located near the N-terminal region in SP1, which is necessary for proteasome-dependent degradation *in vitro* (Su *et al.,* 1999), and as such it is considered likely to be involved in SP turnover (Bouwman & Philipsen 2002). Given that SP boxes are lost in some vertebrate SP proteins (Schaeper *et al.,* 2010), their independent loss in several amphioxus SP proteins is not unprecedented.

A second, more surprising observation was the discovery of a novel expansion in *Branchiostoma* SP1-4. Typically, the C2H2 triplet of SP and KLF proteins is found close to the C-terminus of these proteins, which is not the case for amphioxus SP1-4. When comparing the length of this C2H2 'trailing region' across the entire pan-metazoan KLF/SP dataset of Presnell *et al.* (2015), in almost all cases these are less than 100 residues in length, compared to almost 500 in *Branchiostoma* SP1-4. Unusually, 25% of the residues in this region are Glutamine. Whilst it is not uncommon in other contexts to see poly-Glutamine repeats occur through DNA replication slippage, in this instance there are no repeats longer than $(Q)_3$. This region of low-complexity overlaps with the assembly artefacts seen in the *B. lanceolatum Sp1-4* loci, where 100% identical introns and exons were repeated multiple times. Low-complexity regions such as there are relatively difficult for assembly software to reconstruct, providing a potential explanation for these artefacts. We also speculate, therefore, that the additional "SP1-4b" protein (which contains only the region from the Btd box to final C2H2 domain) from Presnell *et al.* (2015) may be the product of a similar computational error. The purpose of the Glutamine-rich trailing region of *Branchiostoma* SP1-4 remains unclear. However, the Glutamine rich pre-Btd region of SP1 allows the cross-linking of SP1 monomers, permitting both proximally- and distally-bound SP1 molecules to synergistically act upon a single target gene (Courey *et al.,* 1989). Given that amphioxus SP1-4 still retains a Glutamine rich pre-Btd region, it is possible that this trailing region facilitates novel interactions in *Branchiostoma.*

Our extended analyses also give further insight into the genomic arrangement of *Branchiostoma Sp* genes, and evolution of the cluster. The current model of *Sp* gene evolution predicts a eumetazoan gene cluster containing a single *Sp1-4, Sp5* and *Sp6-9,* based primarily upon the existence of such a cluster in the cnidarian *Nematostella vectensis* (Schaeper *et al.,* 2010). This model also points to the non-teleost vertebrate condition, where *Sp* genes are found on four separate chromosomes in an arrangement congruent with an original three *Sp*-gene cluster expanding during the 2R event (Schaeper *et al.,* 2010). We were able to identify *Sp* gene arrangements from a number of invertebrate species that support this model (Fig. 7). These included a second cnidarian

species, *Orbicella faveolata,* as well as the hemichordate *Saccoglossus kowalevskii.* Finally, of course, we were able to identify all three *B. lanceolatum Sp* genes on a single genomic scaffold. The location of *B. belcheri Sp1-4* on a separate scaffold, whilst not conclusive, does not preclude it from being located on the same chromosome as the other *B. belcheri* genes. Therefore, all of the additional data collated here support the existence of *Sp1-4, Sp4* and *Sp6-9* existing on the same chromosome in both the deuterostome and chordate ancestor. Interestingly, however, whilst previous studies refer to this arrangement as a 'cluster' (Schaeper *et al.,* 2010), the intragenic distances for the species observed here (Fig. 7) are considerable. Even in the cnidarian *O. faveolata,* these distances are double (50kb) those in *N. vectensis* (25-29kb). In *B. lanceolatum, Sp1-4* is 1.1Mb from the other *Sp* genes, which we feel makes it unlikely that there is meaningful "clustering" of this gene with the others.

### *Amphioxus* Sp5 *expression is dynamic*

To date, expression of *Sp* genes has not been characterised in any invertebrate chordate species. The combined developmental and adult transcriptome for *Branchiostoma lanceolatum* (Oulion *et al.,* 2012) contains fragments matching all three *Sp* genes from this species, and so it is possible that all three are expressed during amphioxus development. Here, we cloned the *B. lanceolatum Sp5* gene; as the only *Sp* not to have duplicated during the vertebrate 2R events, this allows for more direct comparison among the Metazoa, and an insight into its role in the common chordate ancestor. Between the onset of neurulation (N0) and early larva (L1), *Sp5* expression marks recently specified neural, notochord and somitic tissue, which is consistent with observations in other species for roles in specifying neurectoderm and axial mesoderm (Weidinger *et al.,* 2005). As axis elongation occurs, and the more anterior and central regions of these tissues begin to differentiate, *Sp5* expression in most neural and notochord tissue is lost, and somitic expression is more sporadic. As a result of this, most *Sp5* expression is found in the newly formed regions of these tissues just proximal of the tailbud, as well as the tailbud itself. This is therefore consistent with *Sp5's* conserved role in embryological growth zones in the Bilateria.

Several anterior domains of *Sp5* expression persist during axis elongation. Expression in the presumptive pharynx is matched by that of mouse *Sp5*, which is expressed in the first three branchial pouches and clefts (Treichel *et al.,* 2001). Expression also remains in some anterior neural tissues such as the cerebral vesicle. In the mouse, *Sp5* is reportedly expressed both in the forebrain, in regions such as the dorsal posterior telencephalon, and at the midbrain-hindbrain boundary (Treichel *et al.,* 2001; Sahara *et al.,* 2007). Classically, the cerebral vesicle in amphioxus was considered to be homologous to the vertebrate forebrain (Lacalli *et al.,* 1994). However, recent studies indicate that by early neurula stages (N2), the amphioxus neural tube can be subdivided into three regions: the hypothalamo-prethalamic primordium (HyPTh); the Di-Mesencephalic primordium (DiMes) and Rhombencephalo-Spinal primordium (RhSp) regions, which are homologous to the vertebrate midbrain, thalamus, and pretectum; and the hindbrain and spine (Albuixech-Crespo *et al.,* 2017). It would therefore be interesting to determine if the neural expression of *Sp5* overlaps the DiMes and RhSp boundary in amphioxus, or rather the more anterior HyPth, via double-labelling experiments. Despite this outstanding question, these findings indicate a high degree of conservation between amphioxus and vertebrate *Sp5* expression. Amphioxus may therefore be a useful model system in which to study the role of *Sp5* in contexts such as chordate posterior elongation or neural development.

### *Contrasting responses of* Sp5 *expression to increased Wnt/β-catenin signalling*

The role of SP5 as both a downstream effector and component of the Wnt/β-catenin pathway is well documented in vertebrates (see introduction), particularly its relationship with *Brachyury/T* in promoting proliferation in the tailbud. Cephalochordates also possess an embryonic tailbud structure, which expresses multiple Wnt ligands and two *Brachyury* genes derived from a lineage-specific

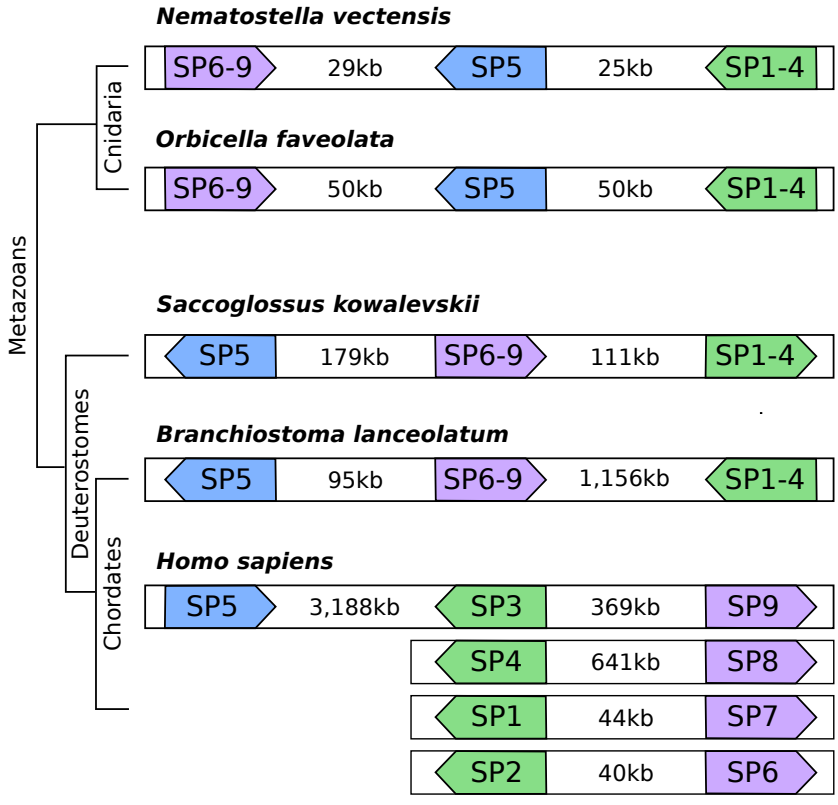


**Fig. 7. Comparison of *Specificity Protein (Sp)* gene arrangement between selected cnidarians and deuterostomes.** *We identify further support for a cnidarian three-gene* Sp *cluster from* O. faveolata*, although the intergenic distances almost double those in* N. vectensis. *Genomic scaffold data from both* B. lanceolatum *and the hemichordate* S. kowalevskii *show that these genes remained on the same chromosomes in both the deuterostome and chordate common ancestor. Whilst vertebrates have experienced two rounds of whole genome duplication, and exact* Sp *complements vary among different lineages, the chromosomes of* H. sapiens *show similar two and three-gene complements, consistent with such an ancestral arrangement existing before the 2R events.*

duplication event (Holland *et al.*, 1995; Terazawa and Satoh., 1997; Holland 2002; Bertrand *et al.*, 2017). The two amphioxus *Brachyury* genes are virtually identical within the coding region, and *in situ* hybridisation that utilises 3' UTR sequence as a probe is unsuccessful, such that the expression pattern may be a conflation of both genes (Holland *et al.,* 1995; Inoue *et al.,* 2017). Regardless, we show that *Sp5* expression overlaps with both the tailbud and *Brachyury2* expression in *B. lanceolatum*, raising the question as to whether these two genes also share a relationship with Wnt/β-catenin in the amphioxus tailbud.

When *B. lanceolatum* embryos were treated here with 1-azakenpaullone (Azk) to increase Wnt/β-catenin signalling, *Brachyury2* expression expanded significantly. This finding is generally consistent with β-catenin acting to promote *Brachyury2* expression, similar to the relationship of these two factors in vertebrates. In contrast, *Sp5* expression decreased in the *Brachyury2*$^+$ region of the tailbud. Taken in isolation, this result could be interpreted as an inhibition of *Sp5* expression, resulting directly from increased β-catenin levels. We felt however that this was unlikely, given that SP5 is an effector of Wnt/β-catenin signalling in vertebrates, and both Wnt ligands and *Sp5* are co-expressed in domains overlapping the amphioxus tailbud. In order to determine if this response was consistent, we performed additional treatments at an earlier developmental stage with two different GSK3-β inhibitors. In contrast to prolonged exposure to the drug from the gastrula stage, treatments at either the blastula or 8-cell stage resulted in upregulated expression of *Sp5* by the G4 stage, as assessed by both WMISH and qPCR. Importantly, both *Sp5* and *Brachyury2* show similar responses, becoming ectopically expressed in the ectoderm. This finding is therefore consistent with both *Sp5* and *Brachyury2* acting as downstream targets of Wnt/β-catenin signalling in amphioxus, similarly to vertebrates such as mouse (Fujimura *et al.*, 2007).

Taking all of these results together, the response of *Sp5* to ectopically increased β-catenin levels appears to be variable, for which there are a number of explanations. In the simplest scenario, it may simply be that these responses are specific to certain tissues, such as *Brachyury2*$^+$ tailbud cells. Alternatively, the direction of response may be dependent on the developmental stages at which the embryos were treated. However, we also note that the gastrula-stage treatment window was significantly longer, and that this was the only instance in which GSK3-β inhibition resulted in decreased *Sp5* expression. This could suggest that there is some compensatory regulatory mechanism that acts to combat ectopic increases in β-catenin, which is only detected after prolonged exposure to treatment. To resolve this question, qPCR time courses for *Sp5* expression following both shorter and longer treatments would be informative, and it would be important to determine if these responses differ between early and later development. It may also be possible to ectopically express *Sp5* mRNA in *B. lanceolatum* embryos. This would bypass any hypothetical transcriptional negative feedback loop, and we may expect to find a similar (or indeed additive) tailbud phenotype to that seen following pharmacological Wnt/β-catenin activation.

## Conclusions

The protein structure, phylogenetic signal and genomic position of the three *Branchiostoma Sp-like* genes found in *B. lanceolatum* and *B. belcheri* are consistent with this genus possessing a three-gene *Sp* cluster, as is typical for invertebrate eumetazoans. During embryonic development in *B. lanceolatum, Sp5* is specifically expressed in a number of tissues of the embryo, including the cerebral vesicle and tailbud, which is broadly consistent with its roles in other bilaterians. We demonstrate that expression of *Sp5* is modulated by pharmacological increases in Wnt/β-catenin signalling, although its response varies depending upon treatment conditions. We hypothesise that the decrease of *Sp5* expression seen only in long treatments is indicative of a negative-feedback loop in the Wnt/β-catenin-*Sp5* signalling pathway.

## Materials and Methods

### *Identification of* Sp-like *genes*

Human SP proteins were BLAST searched against the predicted protein set of the genome projects for *Branchiostoma belcheri* (Huang *et al.,* 2014) and *B. lanceolatum* (Unpublished data, kindly made available by the *B. lanceolatum* genome consortium). Matches were then reciprocally BLAST searched against the NCBI nr protein database, and apparent SP-like proteins for each species were taken for further analysis. The locus for each gene was confirmed using tBLASTn searches of the predicted protein against the genome scaffold set. Two apparent assembly artefacts are present in the scaffold over one gene in *B. lanceolatum* (shown in Supplementary Figure 1), which lead to the repetition of several exons and intronic regions with 100% identity, flanked by patches of unidentified nucleotides (Ns). These repeat-artefacts were however not present in the predicted protein.

In order to confirm the identity of these three *Branchiostoma* SP-like proteins, two sets of alignments and trees were generated. The first contained the C2H2 domains of a selection of metazoan SP and KLF proteins, as well as the human Wilms Tumour 1 (Wt1) as an outgroup. The subset of these proteins identified as genuine SP proteins in this analysis were then used to create an alignment of Btd and C2H2 domains in order to confirm internal SP clades. Sequences for *Saccoglossus kowalevskii, Lingula anatina* and *Orbicella faveolata* were retrieved through BLASTp searches of the NCBI nr database, and the remaining sequences (and their associated identifications) were taken from Presnell *et al.* (2015) (all sequences and accessions provided in Supplementary File 1). Both alignments were created using MAFFT on default settings, and then analysed using model_generator (Keane *et al.,* 2006) to determine the most suitable model of protein evolution for subsequent phylogenetic trees (given in the appropriate figure legends). The trees shown were created using RAxML (Stamatakis 2014), with additional support values from MrBayes (Ronquist *et al.*, 2012), and all trees were generated on the CIPRES Science Gateway (Miller *et al.,* 2010). The Btd and C2H2 domain alignments of *Branchiostoma* SP proteins were also used to identify the sequence between the N-terminus and the Btd box (referred to here as the 'pre-Btd region') and the sequence following the final C2H2 domain (referred to here as the 'trailing region'). These were then analysed for residue frequency.

### *Embryo collection and treatments*

Adult *Branchiostoma lanceolatum* with mature gonads were collected from Argelès-sur Mer (France), and heat-induced to spawn, as per the protocol detailed in (Fuentes *et al.,* 2007). Embryos used in treatment experiments were kept in a known volume of seawater, allowing the addition of stock solutions of either 1-Azakenpaullone (Azk) or CHIR99021 (both prepared in DMSO) to be added at the specified treatment time. Final concentrations for treatments were 25μM and 10μM Azk when treating at blastula and gastrula stages, respectively, and 10μM CHIR99021 at cleavage (8 cell) stages. Control embryos were treated with an equal volume of DMSO. Once embryos had developed to the desired stage, they were fixed using 4% PFA in MOPS salts (0.1M MOPS, 2mM MgSO4, 1mM EGTA, and 0.5M NaCL) and stored in 70% EtOH. Embryo staging was performed according to Hirakow & Kajita (1994) and Hirakow & Kajita (1991), with modification as per (Zhang *et al.,* 2013).

***Cloning and probe template preparation***

RNA was extracted from a whole unripe *B. lanceolatum* adult using Trisure (Bioline) and phenol chloroform extraction; cDNA was prepared using the Tetro cDNA Synthesis Kit (Bioline). A partial length gene fragments for *Sp5* was obtained through PCR (primer sequences: forward: 5' GTCCTCGTCTTCTTTTCTCACG 3', reverse: 5' TTTGTCAGTATCCCGCCCAG 3'). This fragment was ligated into pGEM-T Easy (Promega) and then transformed into XL10-Gold competent *E. coli* using standard protocols. Plasmids were harvested using peqGOLD (VWR) and sequenced to confirm successful ligation, after which the insert sequence was submitted to GenBank (accession: MF162314.1). For *Brachyury2,* the pBluescript II SK+ plasmid originating from (Somorjai *et al.,* 2008) (accession: EU685284.1) was used.

***Transcriptional evidence and Whole Mount* In situ *Hybridisation (WMISH)***

*B. lanceolatum* SP proteins were used to tBLASTn search publicly available transcript databases for this species to identify evidence for their expression. Matches of over 90% identity (for the full length of the transcript) were then confirmed by reciprocally BLASTx searching them against the predicted protein set of their respective species, before being counted (accessions given in Supplementary Table 1).

*Sp5* and *Brachyury2* containing plasmids were used as PCR templates for standard M13F (5' GTAAACGACGGCCAGT 3') and M13R (5' AACAGCTATGACCATG 3') primers. These products were purified with Bioline Isolate II gel purification kits, and used as templates for the production of DIG-labelled RNA probes, using T7 RNA polymerase with standard protocols. Probes were purified using Roche miniQuick spin columns or via precipitation with sodium acetate (3M, pH 5.2) and ethanol.

WMISH was performed as previously described (Dailey *et al.,* 2016), with minor modifications. Specifically, embryos were permeabilised using proteinase K at 1.125U/ml, for empirically-tested periods based upon embryo stage and enzyme batch, and post-fixed for 60 minutes in PFA. The pre-hybridisation wash, hybridisation wash and initial post-hybridisation washes were performed at either 60 or 62°C, dependent on experiment. The chromogenic reaction was performed using NBT/BCIP (Roche), and embryos were then post-fixed in PFA for 60 minutes. Four WMISH experiments were performed for embryonic *Sp5* expression, and three for *Sp5* and *Brachyury2* expression in treated embryos, with a minimum of 3 embryos per stage. Embryos were mounted in 90% glycerol/10% PBS, and photographed under a Leitz DMRB microscope (Leica Microsystems) with Normarski optics. Photographs were taken with the Retiga 2000R camera and the QCapture software suite (QImaging), and processed using Fiji (Schindelin *et al.,* 2012) and Adobe Fireworks CS5.

***Quantitative RT-PCR (qRT-PCR)***

Embryos were treated with DMSO (control) or 10μM CHIR99021 at the 8-cell stage and harvested for RNA at the G4 gastrula stage. cDNA was prepared using a VILO cDNA synthesis kit (Invitrogen). qRT-PCR reactions were run in a LightCycler® 480 Instrument (Roche) using LightCycler® 480 DNA SYBR Green I Master (Roche) according to the standard manufacturer's protocol; typically, 5μl reaction volumes were used. PCR reactions were performed in triplicate for each primer set. Crossing point (Cp) values were calculated with LightCycler® 480 Software (Roche), using the second-derivate maximum algorithm. The average Cp values of all technical replicates were normalized against the Cp values of housekeeping genes (Rpl32 and Polr2l). Statistical significance of the change in mRNA expression was calculated by a two-tailed Student t-test in Microsoft Excel. The following primers were used: Rpl32_Fwd: 5' GGCTTCAAGAAATTCCTCGTT 3'; Rpl32_Rev: 5' GATGAGTTTCCTCTTGCGTGA 3'; Polr2l_Fwd: 5' AAGGTGATCGGGAACAAGTG 3'; Polr2l_Rev: 5' GGTCCACATGTGACAGCAAC 3'; Sp5_Fwd: 5' GAGATCACCGAGCGACGACTA 3'; Sp5_Rev: 5' GGAGACGGCTTCTGAAGCTCT 3'.

*Conflict of interest*

*The authors declare that there is no conflict of interest.*

*Author contributions*

*SCD performed experiments with contributions from IMLS, and analysed genomic data. IK performed and analysed CHIR99021 treatment experiments and qPCR. IMLS conceived the study and contributed reagents. SCD and IMLS wrote the paper. All authors approved the final manuscript.*

*Funding and Acknowledgements*

*SCD was funded by a MASTS (Marine Alliance for Science & Technology Scotland) PhD Studentship. Work in the IMLS lab is in part funded by the European Union Horizon 2020 research and innovation programme under grant agreement numbers 654428 ("CORBEL") and 730984 ("ASSEMBLE+), and the Wellcome Trust ISSF grant number 204821/Z/16/Z. I.K. was supported by grant 15-21285J from the Czech Science Foundation. This work was supported by the* Branchiostoma lanceolatum *genome consortium, which provided access to the* Branchiostoma lanceolatum *genome sequence.*

## References

ALBUIXECH-CRESPO B, LÓPEZ-BLANCH L, BURGUERA D, MAESO I, SÁNCHEZ-ARRONES L, MORENO-BRAVO JA, SOMORJAI I, PASCUAL-ANAYA J, PUELLES E, BOVOLENTA P, GARCIA-FERNÀNDEZ J, PUELLES L, IRIMIA M and FERRAN JL (2017). Molecular regionalization of the developing amphioxus neural tube challenges major partitions of the vertebrate brain. *PLoS Biol* 15: e2001573.

BAUR F, NAU K, SADIC D, ALLWEISS L, ELSÄSSER H-P, GILLEMANS N, DE WIT T, KRÜGER I, VOLLMER M, PHILIPSEN S and SUSKE G (2010). Specificity protein 2 (*Sp2*) is essential for mouse development and autonomous proliferation of mouse embryonic fibroblasts. *PLoS One* 5: e9587.

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

BERTRAND S, LE PETILLON Y, SOMORJAI I and ESCRIVÀ E (2017). Developmental cell-cell communication pathways in the cephalochordate amphioxus: actors and functions. *Int. J. Dev. Biol.* (doi: 10.1387/ijdb.170202sb)

BOUWMAN P and PHILIPSEN S (2002). Regulation of the activity of *Sp1*-related transcription factors. *Mol Cell Endocrinol* 195: 27–38.

COUREY AJ, HOLTZMAN DA and JACKSON SP, TJIAN R, ZENKE M and CHAMBON P (1989). Synergistic activation by the glutamine-rich domains of human transcription factor *Sp1*. *Cell* 59: 827–836.

DAILEY SC, FEBRERO PLANAS R, ROSSELL ESPIER A, GARCIA-FERNÀNDEZ J and SOMORJAI I (2016). Asymmetric distribution of *pl10* and *bruno2*, new members of a conserved core of early germline determinants in cephalochordates. *Front Ecol Evol* 3: 19-30.

FUENTES M, BENITO E, BERTRAND S, PARIS M, MIGNARDOT A, GODOY L, JIMENEZ-DELGADO S, OLIVERI D, CANDIANI S, HIRSINGER E, D'ANIELLO S, PASCUAL-ANAYA J, MAESO I, PESTARINO M, VERNIER P, NICOLAS J, SCHUBERT M, LAUDET V, GENEVIERE AM, ALBALAT R, GARCIA FERNÀNDEZ JG, HOLLAND ND and ESCRIVÀ H (2007). Insights into spawning behavior and development of the European amphioxus (*Branchiostoma lanceolatum*). *J Exp Zool B Mol Dev Evol* 308: 484–493.

FUJIMURA N, VACIK T, MACHON O, VLCEK C, SCALABRIN S, SPETH M, DIEP D, KRAUSS S and KOZMIK Z. (2007) Wnt-mediated down-regulation of *Sp1* target genes by a transcriptional repressor Sp5. *J Biol Chem* 282: 1225-1237.

HAGEN G, MÜLLER S, BEATO M and SUSKE G (1994). *Sp1*-mediated transcriptional activation is repressed by Sp3. *EMBO J* 13: 3843–3851.

HARRISON SM, HOUZELSTEIN D, DUNWOODIE SL and BEDDINGTON RSP (2000). *Sp5*, a new member of the *Sp1* family, is dynamically expressed during development and genetically interacts with *Brachyury*. *Dev Biol* 227: 358–372.

HIRAKOW R and KAJITA N (1994). Electron microscopic study of the development of amphioxus, *Branchiostoma belcheri tsingtauense*: the neurula and larva. *Kaibogaku Zasshi* 69: 1–13.

HIRAKOW R and KAJITA N (1991). Electron microscopic study of the development of amphioxus, *Branchiostoma belcheri tsingtauense*: The gastrula. *J Morphol* 207: 37–52.

HOLLAND LZ (2002). Heads or tails? Amphioxus and the evolution of anterior-posterior patterning in deuterostomes. *Dev Biol* 241: 209–228.

HOLLAND PW, KOSCHORZ B, HOLLAND LZ and HERRMANN BG (1995). Conservation of *Brachyury (T)* genes in amphioxus and vertebrates: developmental and evolutionary implications. *Development* 121: 4283–4291.

HUANG S, CHEN Z, YAN X, YU T, HUANG G, YAN Q, PONTAROTTI PA, ZHAO H, LI J, YANG P, WANG R, LI R, TAO X, DENG T, WANG Y, LI G, ZHANG Q, ZHOU S, YOU L, YUAN S, FU Y, WU F, DONG M, CHEN S and XU A (2014). Decelerated genome evolution in modern vertebrates revealed by analysis of multiple lancelet genomes. *Nat Commun* 5: 5896.

INOUE J, YASUOKA Y, TAKAHASHI H and SATOH N (2017). The chordate ancestor possessed a single copy of the *Brachyury* gene for notochord acquisition. *Zool Lett* 3: 4.

IUCHI S (2001). Three classes of C2H2 zinc finger proteins. *Cell Mol Life Sci* 58: 625–635.

KAWAKAMI Y, ESTEBAN CR, MATSUI T, RODRÍGUEZ-LEÓN J, KATO S and BELMONTE JCI (2004). *Sp8* and *Sp9*, two closely related buttonhead-like transcription factors, regulate *Fgf8* expression and limb outgrowth in vertebrate embryos. *Development* 131: 4763-4774.

KEANE TM, CREEVEY CJ, PENTONY MM, NAUGHTON TJ and MCLNERNEY JO (2006). Assessment of methods for amino acid matrix selection and their use on empirical data shows that *ad hoc* assumptions for choice of matrix are not justified. *BMC Evol Biol* 6: 29.

KENNEDY MW, CHALAMALASETTY RB, THOMAS S, GARRIOCK RJ, JAILWALA P and YAMAGUCHI TP (2016). Sp5 and Sp8 recruit β-catenin and Tcf1-Lef1 to select enhancers to activate Wnt target gene transcription. *Proc Natl Acad Sci USA* 113: 3545–3550.

KUNICK C, LAUENROTH K, LEOST M, MEIJER L and LEMCKE T (2004). 1-Azakenpaullone is a selective inhibitor of glycogen synthase kinase-3β. *Bioorg Med Chem Lett* 14: 413–416.

LACALLI TC, HOLLAND ND and WEST JE (1994). Landmarks in the anterior central nervous system of amphioxus larvae. *Philos Trans R Soc London B Biol Sci* 344: 165-185.

MILLER MA, PFEIFFER W and SCHWARTZ T (2010). Creating the CIPRES Science Gateway for inference of large phylogenetic trees. *2010 Gateway Comp. Environ. Workshop (GCE)* IEEE: 1–8.

NAKAMURA T, UNDA F, DE-VEGA S, VILAXA A, FUKUMOTO S, YAMADA KM and YAMADA Y (2004). The Krüppel-like factor epiprofin is expressed by epithelium of developing teeth, hair follicles, and limb buds and promotes cell proliferation. *J Biol Chem* 279: 626–634.

NAKASHIMA K, ZHOU X, KUNKEL G, ZHANG Z, DENG JM, BEHRINGER RR and DE CROMBRUGGHE B (2002). The novel zinc finger-containing transcription factor *osterix* is required for osteoblast differentiation and bone formation. *Cell* 108: 17–29.

OULION S, BERTRAND S, BELGACEM MR, LE PETILLON Y and ESCRIVÀ H (2012). Sequencing and analysis of the Mediterranean amphioxus (*Branchiostoma lanceolatum*) transcriptome. *PLoS One* 7: e36554.

PEI J and GRISHIN N V (2015). C2H2 zinc finger proteins of the SP/KLF, Wilms tumor, EGR, Huckebein, and Klumpfuss families in metazoans and beyond. *Gene* 573: 91–99.

PETERSEN CP and REDDIEN PW (2009). Wnt signaling and the polarity of the primary body axis. *Cell* 139: 1056–1068.

PRESNELL JS, SCHNITZLER CE and BROWNE WE (2015). KLF/SP transcription factor family evolution: Expansion, diversification, and innovation in eukaryotes. *Genome Biol Evol* 7: 2289–2309.

PUTNAM NH, SRIVASTAVA M, HELLSTEN U, DIRKS B, CHAPMAN J, SALAMOV A, TERRY A, SHAPIRO H, LINDQUIST E, KAPITONOV V V., JURKA J, GENIKHOVICH G, GRIGORIEV I V., LUCAS SM, STEELE RE, FINNERTY JR, TECHNAU U, MARTINDALE MQ and ROKHSAR DS (2007). Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. *Science* 317: 86–94.

SAFFER JD, JACKSON SP and ANNARELLA MB (1991). Developmental expression of *Sp1* in the mouse. *Mol Cell Biol* 11: 2189–2199.

SAHARA S, KAWAKAMI Y, IZPISUA BELMONTE J and O'LEARY DD (2007). *Sp8* exhibits reciprocal induction with *Fgf8* but has an opposing effect on anterior-posterior cortical area patterning. *Neural Dev* 2: 10.

SCHAEPER ND, PRPIC N-M and WIMMER EA (2010). 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 Evol Biol* 10: 88.

SCHAEPER ND, PRPIC N-M and WIMMER EA (2009). A conserved function of the zinc finger transcription factor *Sp8/9* in allometric appendage growth in the milkweed bug *Oncopeltus fasciatus*. *Dev Genes Evol* 219: 427–435.

SCHINDELIN J, ARGANDA-CARRERAS I, FRISE E, KAYNIG V, LONGAIR M, PIETZSCH T, PREIBISCH S, RUEDEN C, SAALFELD S, SCHMID B, TINEVEZ J-Y, WHITE DJ, HARTENSTEIN V, ELICEIRI K, TOMANCAK P and CARDONA A (2012). Fiji: an open-source platform for biological-image analysis. *Nat Methods* 9: 676–682.

SHIMELD SM (2008). C2H2 zinc finger genes of the *Gli, Zic, KLF, SP, Wilms' tumour, Huckebein, Snail, Ovo, Spalt, Odd, Blimp-1, Fez* and related gene families from *Branchiostoma floridae*. *Dev Genes Evol* 218: 639–649.

SOMORJAI I, BERTRAND S, CAMASSES A, HAGUENAUER A and ESCRIVA H (2008). Evidence for stasis and not genetic piracy in developmental expression patterns of *Branchiostoma lanceolatum* and *Branchiostoma floridae*, two amphioxus species that have evolved independently over the course of 200 Myr. *Dev Genes Evol* 218: 703–713.

STAMATAKIS A (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30: 1312–1313.

SU K, ROOS MD, YANG X, HAN I, PATERSON AJ and KUDLOW JE (1999). An N-terminal region of Sp1 targets its proteasome-dependent degradation *in vitro*. *J Biol Chem* 274: 15194–202.

TALLAFUS A, WILM TP, CROZATIER M, PFEFFER P, WASSEF M and BALLY-CUIF L (2001). The zebrafish buttonhead-like factor Bts1 is an early regulator of *pax2.1* expression during mid-hindbrain development. *Development* 128: 4021-4034.

TERAZAWA K and SATOH N (1997). Formation of the chordamesoderm in the amphioxus embryo: Analysis with *Brachyury* and *fork head/HNF-3* genes. *Dev Genes Evol.* 207: 1-11.

THORPE CJ, WEIDINGER G and MOON RT (2005). Wnt/β-catenin regulation of the *Sp1*-related transcription factor *Sp5l* promotes tail development in zebrafish. *Development* 132: 1763–1772.

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

VIZCAÍNO C, MANSILLA S and PORTUGAL J (2015). *Sp1* transcription factor: A long-standing target in cancer chemotherapy. *Pharmacol Ther* 152: 111–124.

WEIDINGER G, THORPE CJ, WUENNENBERG-STAPLETON K, NGAI J and MOON RT (2005). The *Sp1*-related transcription factors *Sp5* and *Sp5*-like act downstream of Wnt/β-Catenin signaling in mesoderm and neuroectoderm patterning. *Curr Biol* 15: 489–500.

WIMMER EA, JÄCKLE H, PFEIFLE C and COHEN SM (1993). A *Drosophila* homologue of human *Sp1* is a head-specific segmentation gene. *Nature* 366: 690–694.

ZHANG Q-J, LUO Y-J, WU H-R, CHEN Y-T and YU J-K (2013). Expression of germline markers in three species of amphioxus supports a preformation mechanism of germ cell development in cephalochordates. *Evodevo* 4: 17.

## Further Related Reading, published previously in the *Int. J. Dev. Biol.*

**The Wnt signaling mediator tcf1 is required for expression of foxd3 during Xenopus gastrulation**
Sylvie Janssens, Olaf Van Den Broek, Ian R. Davenport, Robbert C. Akkers, Fei Liu, Gert Jan C. Veenstra, Stefan Hoppler, Kris Vleminckx and Olivier Destrée
Int. J. Dev. Biol. (2013) 57: 49-54

**Wnt signaling in planarians: new answers to old questions**
Maria Almuedo-Castillo, Miquel Sureda-Gómez and Teresa Adell
Int. J. Dev. Biol. (2012) 56: 53-65

**From the American to the European amphioxus: towards experimental Evo-Devo at the origin of chordates**
Jordi Garcia-Fernàndez, Senda Jiménez-Delgado, Juan Pascual-Anaya, Ignacio Maeso, Manuel Irimia, Carolina Minguillón, Èlia Benito-Gutiérrez, Josep Gardenyes, Stéphanie Bertrand and Salvatore D'Aniello
Int. J. Dev. Biol. (2009) 53: 1359-1366
https://doi.org/10.1387/ijdb.072436jg

**Wnt signaling in hydroid development: ectopic heads and giant buds induced by GSK-3beta inhibitors**
Werner Müller, Uri Frank, Regina Teo, Ofer Mokady, Christina Guette and Günter Plickert
Int. J. Dev. Biol. (2007) 51: 211-220

**Formation and maintenance of the organizer among the vertebrates.**
K Joubin and C D Stern
Int. J. Dev. Biol. (2001) 45: 165-175


**5 yr ISI Impact Factor (2013) = 2.879**