Manuscript Details

Manuscript number	DEVELOPMENTALBIOLOGY_2017_555
Title	De novo neurogenesis in a budding chordate: co-option of larval anteroposterior patterning genes in a transitory neurogenic organ
Article type	Review article

Abstract

During metamorphosis of solitary ascidians, part of the larval tubular nervous system is recruited to form the adult central nervous system (CNS) through neural stem-like cells called ependymal cells. The anteroposterior (AP) gene expression patterning of the larval CNS regionalize the distribution of the ependymal cells, which contains the positional information of the neurons of the adult nervous system. In colonial ascidians, the CNS of asexually developed zooids has the same morphology of the one of the post-metamorphic zooids. However, its development follows a completely different organogenesis that lacks embryogenesis, a larval phase and metamorphosis. In order to describe neurogenesis during asexual development (blastogenesis), we followed the expression of six CNS AP patterning genes conserved in chordates and five neural-related genes to determine neural cell identity in Botryllus schlosseri. We observed that a neurogenesis occurs de novo on each blastogenic cycle starting from a neurogenic transitory structure, the dorsal tube. The dorsal tube partially co-opts the AP patterning of the larval CNS markers, and potentially combine the neurogenesis role and provider of positional clues for neuron patterning. This study shows how a larval developmental module is reused in a direct asexual development in order to generate the same structures.

Keywords	Ascidians, neurogenesis, regeneration, Otx, neural gland, Botryllus schlosseri

Corresponding Author	Stefano Tiozzo
Corresponding Author's Institution	CNRS-UPMC
Order of Authors	Maria Mandela Prünster, Lorenzo Ricci, Federico Brown, Stefano Tiozzo

Suggested reviewers	bernie degnan, Alberto Stolfi, Cristian Canestro
Opposed reviewers	Baruch Rinkevich, Lucia Manni, Ayelet Voskoboynik

Highlights

1. Nervous system develops *de novo* during *Botryllus schlosseri* blastogenetic cycle
2. Asexual neurogenesis origins from a transitory tubular structure, the dorsal tube
3. Dorsal tube formation and neurulation share similar anteroposterior patterning

De novo neurogenesis in a budding chordate: co-option of larval anteroposterior patterning genes in a transitory neurogenic organ

Maria Mandela Prünster[1], Lorenzo Ricci[1, 2], Federico Brown[3,4], Stefano Tiozzo[1,*]

[1]Sorbonne Universités, CNRS, Laboratoire de Biologie du Développement de Villefranche-sur-mer (LBDV), 06230 Paris, France
[2]Harvard University, Department of Organismic & Evolutionary Biology, 52 Oxford Street Cambridge, MA 02138, US
[3] Departamento de Zoologia - Instituto Biociências, Universidade de São Paulo, São Paulo, SP, Brazil, CEP 05508-090
[4] Centro de Biologia Marinha (CEBIMar), Universidade de São Paulo, São Sebastião, SP, Brazil, CEP 11612-109

*Corresponding author: tiozzo@obs-vlfr.fr

ABSTRACT

During metamorphosis of solitary ascidians, part of the larval tubular nervous system is recruited to form the adult central nervous system (CNS) through neural stem-like cells called ependymal cells. The anteroposterior (AP) gene expression patterning of the larval CNS regionalize the distribution of the ependymal cells, which contains the positional information of the neurons of the adult nervous system.

In colonial ascidians, the CNS of asexually developed zooids has the same morphology of the one of the post-metamorphic zooids. However, its development follows a completely different organogenesis that lacks embryogenesis, a larval phase and metamorphosis.

In order to describe neurogenesis during asexual development (blastogenesis), we followed the expression of six CNS AP patterning genes conserved in chordates and five neural-related genes to determine neural cell identity in *Botryllus schlosseri*.

We observed that a neurogenesis occurs *de novo* on each blastogenic cycle starting from a neurogenic transitory structure, the dorsal tube. The dorsal tube partially co-opts the AP patterning of the larval CNS markers, and potentially combine the neurogenesis role and provider of positional clues for neuron patterning. This study shows how a larval

developmental module is reused in a direct asexual development in order to generate the same structures.

KEY WORDS:

Ascidians, neurogenesis, regeneration, AP axis, Otx, neural gland, *Botryllus schlosseri*

INTRODUCTION

During the development of most bilaterian, the central nervous system (CNS) becomes regionalized molecularly along its dorso-ventral and anteroposterior (AP) axes. Such partitioning is defined by a patterned expression of transcription factors (TFs). Dorso-ventral regionalization has been suggested to have evolved independently[1]. The AP order of gene expression is conserved even in animals with no centralized nervous system[2] and such patterning can be important for the axis of entire organisms[3].

In ascidians (Tunicata), the larval CNS is formed like in other chordates by the rolling up of a neural plate into a dorsal hollow tube, which becomes organized along the AP axis into four regions: the sensory vesicle (SV), the neck, the visceral ganglion (VG), and the tail nerve cord [4]. While the homology of these structures to corresponding regions in chordate CNS regionalization is still debated [4–7], the expression of some TF genes characterizing the AP axis are conserved [3]. For instance, along the AP axis, *Otx* is expressed anteriorly [8,9], and *Pax2/5/8* and *Gli* more posteriorly [10], while several *Hox* genes are expressed even more posterior, in the VG [11] and in the nerve cord [12]. This order of gene expression reflects the patterns of expression of orthologous genes not only of vertebrate, but also of the cephalochordate and the hemichordate nervous systems [3].

During their biphasic life history ascidians undergo a severe remodeling of their nervous system. The swimming tadpole-like larva with the typical chordate body plan metamorphoses into a sessile filter feeding zooid[13]. Throughout this process it loses the larval sensory organs and tail used in locomotion. The post-metamorphic zooid substitutes the larval CNS by the adult neural complex. The neural complex [14] consists of a cerebral ganglion, which innervates the body wall [15–17] and the neural gland (NG) [18,19], which opens anteriorly into the branchial chamber through a ciliated funnel-shaped duct (CF) and continues posteriorly forming the dorsal strand, or dorsal organ (DO) [20].

The adult ganglion is derived from cell progenitors of the transitory larval CNS. In *Ciona* only one third of the larval nervous system are neurons and the rest are glia-like cells, classified as ependymal cells [21]. Most larval neurons disappear during metamorphosis, whereas the ependymal cells of the larval brain differentiate into the neurons of the adult ganglion [22]. Interestingly, cell tracing experiments showed that the patterning of the larval CNS provides the positional information along the AP axis of the adult CNS neurons [7,22–24]. This is an example where a transitional structure, the larval CNS, serves as a scaffold to build the adult definitive structure, the adult CNS[25].

In colonial species of ascidians, such as *Botryllus schlosseri*, the post-metamorphic zooid (oozooid), derived from a larva, begins a cyclic asexual budding process named blastogenesis that leads to the development of other adult zooids (blastozooids) [26]. During blastogenesis all organs, including the nervous system, are created anew from pre-existing epithelia, without passing through a larval stage (Suppl. Mat. 1). Regardless their different ontogenetic origin, the overall anatomy of the CNS in oozooids and blastozooids is identical. During blastogenic neurogenesis, the neural complex originates from an epithelial thickening that evaginates from the dorsal inner vesicle of the young bud forming a tubular structure, the dorsal tube, which elongates anteriorly along the AP axes [26,27]. The dorsal tube gives rise to the neural gland and by delamination of migratory cells to the cerebral ganglion [26,27].

To explore whether the larval AP patterning process has been co-opted during blastogenetic CNS development, we studied the expression of six genes with highly conserved roles in chordate CNS AP patterning, and five neural-related genes to determine neural cell identity (Table 1). We followed the expression dynamics of these 11 genes during blastogenetic development and observed that the dorsal tube, partially co-opt the anteroposterior patterning of the larval CNS markers, potentially providing positional clues for the establishment of the entire neural complex.

	Gene / aliases	Expression in Tunicata		Expression in chordates and hemichordates NS
		larval	post-metamorphic	
AP patterning genes	*Otx*	*Halocynthia roretzi,* anterior CNS, [5] *Ciona intestinalis*, SV & ant. ectoderm, [31] *Oikopleura dioica*, OtxB anterior NS, [33]	*Ciona intestinalis* endostyle, pharynx, heart [28], [29] *Oikopleura dioica anterior* endostyle,[32] *Herdmania curvata* pharynx, CF, [34]	[3,30]
	Gli (zinc finger TF)	*C.intestinalis* neck, [10,35]	*C. intestinalis* endostyle, pharynx and heart, [34]	[3]
	IrxB (homeobox TF)	*Ascidia:*Irx behind posterior sensory vesicle, [3]	NA	[3,30,36]
	Pou3 (Pou class III homeobox TF)	*Ciona intestinalis* no orthologue, *Halocynthia roretzi, Oikopleura dioica,* NA supplement figure 3	*Ciona intestinalis* no orthologue, *Halocynthia roretzi, Oikopleura dioica,* NA supplement figure 3	[30,37]
	Hox3 (homeobox TF)	*C. intestinalis* anterior region of the visceral ganglion, [11]	*C. intestinalis* posterior part of the CG, [38]	[3,30]
	Pax3/7 (paired box TF)	*C. intestinalis* 3 patches in brain, post. neural tube, [39]	NA	[30,40,41]
Neural / neural tube marker	*Zic-r.a* (zinc finger TF) Macho-1	*C. intestinalis* in CNS with 2 gaps, [42] *Halocynthia roretzi,* not in CNS, [43]	NA	[42,43]
	Ebf (HLH TF) COE, Unc-3	*C. intestinalis* CNS, [35]	NA	[30,44]
	Etr (ELAV-like family member) Celf/CUGBP	*Halocynthia roretzi* CNS, [45,46]	NA	[47]
	Notch (transmembrane receptor)	*Halocynthia roretzi* preferentially in CNS [45,48]	NA	[49,50]
	Pou4 (Pou class IV homeobox TF)	*C. intestinalis* PNS and posterior sensory vesicle, [51]	*C. intestinalis* endostyle, [39]	[52,53]

MATERIALS AND METHODS

Animal husbandry

Botryllus schlosseri colonies were raised on 50x70x1 mm glass slides as described previously[54]. A *Botryllus* colony consists of three coexisting generations that arise asexually: the adult filter feeding zooids, their buds, called primary buds, and their buds (secondary buds or budlets). Budding (blastogenesis) was staged according to Lauzon et al. (2002)[55]. First, a secondary bud appears as a thickening of the peribranchial epithelia and overlying epidermis of the previous generation (stages A1-A2). Second, the peribranchial evagination closes off and forms a vesicle, whereas the overlying epidermis maintains a connection to the parental zooid giving the impression of a double vesicle (stages B1-B2). Third, organogenesis begins as the inner vesicle folds into three distinct chambers: the central branchial chamber, and two lateral peribranchial chambers (stages C1-C2). While the adult zooid undergoes programmed cell death, and the primary bud takes over his place to become the next generation of zooid (stage D) and the secondary bud becomes the primary (A).

Gene identification and phylogeny

RNA sequences were retrieved by tblastn from the *Botryllus schlosseri* transcriptome database http://octopus.obs-lfr.fr/public/botryllus/blast_botryllus.php, full length sequences of the tunicate proteins are retrieved from Aniseed (https://www.aniseed.cnrs.fr/), and others from NCBI (Suppl. Mat. 2). Alignments were generated using MAFFT, and sequences trimmed by the TrimAl Gappyout method. Maximum likelihood trees were compiled using PhyML [56] (Suppl. Mat. 3-11).

Fluorescent in situ hybridization (FISH)

Antisense mRNA probes were designed within the coding region of each gene (Supp. Mat. 12) FISH was carried out as previously described in Ricci et al. (2016)[57] with the following modifications: 1% Dextran sulfate was added to the Hybridization buffer and the revelation solution. The anti-Digoxigenin Antibody (HRP) (Roche, #11207733910) was pre-adsorbed for 1 hour in hybridization solution with a mix of fixed colonies at different stages. When the tunic was exhibiting a very strong background the animals were manually removed from the tunic after rehydration, post fixed in 4% PFA for 1h and transferred into washing baskets in 24-well plates. DIG-probe detection was performed with bench-made FITC-Tyramide by 3h incubation. For double FISH, the hybridization of DIG labeled and Fluorescein labeled probes was performed at the same time, fluorescein probes were detected with Cy3-Tyramide.

Imaging

Confocal images were acquired using a Leica SP8 (40x/1.1 Water WD 0.6 HCX PL APO CS2) and processed with ImageJ [58]and Inkscape.

RESULTS

Early expression of neural patterning genes during blastogenesis

At the onset of blastogenesis the secondary bud forms a double vesicle, the inner vesicle starts to fold giving rise to various structures including the dorsal tube[26](Materials and methods). The dorsal tube begins as a dorsal thickening of the epithelium of the inner vesicle[27]. In this region *Otx* and *IrxB* were expressed starting from the double vesicle stage (stage B2, Fig. 1 A, B). This expression was maintained when the primordium of the dorsal tube started to evaginate (stage C1, Fig. 1C). At this stage the first AP pattern was detected (Fig. 1 D-D''): *Pou3* was expressed anteriorly in few cells in the forming dorsal tube (Fig. 1D), moving posteriorly *Pou3* was co-expressed with the posterior marker *Hox3* (Fig. 1D') and more posteriorly few cells expressed only *Hox3* (Fig. 1D''). The area of the evaginating dorsal tube co-expressed also the neural makers *Zic-r.a* and *Ebf* (Fig. 1E-G). While the domain of expression of *Zic-r.a* was broader (Fig. 1F), *Ebf* was expressed only in a set of cells at the posterior side of the *Zic-r.a* domain (Fig. 1G).

Dynamic of expression during the elongation of the dorsal tube

In the secondary bud, during stages C2 the forming dorsal tube extends anteriorly as a blind tube (Fig. 2A) to meet the inner vesicle and fuse to the future branchial chamber at stage D (Fig. 2B-C). In the meantime, cells detach from the walls of the dorsal tube to the overlying mesenchyme[26,27]. During the elongation of the dorsal tube *Otx* transcripts were expressed at the anterior pole with a gradual increase in expression to the ventral side, over the dorsal epithelia of the future branchial basket and atrial chambers, as well as in other more posterior regions (Fig. 2D). During dorsal tube elongation, *IrxB* was expressed around the dorsal tube and in scattered cells of the tube (Fig. 2E). The expression of *Gli* was first detected along the entire dorsal tube (Fig. 2F) and then became restricted in the median region of the dorsal tube after the anterior fusion of the tube (Fig. 2G). During dorsal tube elongation *Pou3* expression shifted posteriorly and overlaps with *Hox3* expression (Fig. 2H). During anterior fusion of the dorsal tube *Hox3* was detected in the posterior end of the tube, and in mesenchymal cells around the dorsal tube (Fig. 2I). *Pax3/7* patterns *Ciona* CNS in three regions anterior and posterior part of the SV, the neck and the caudal nerve cord[59]. In the *Botryllus* C2 secondary bud, *Pax3/7* was expressed in the anterior dorsal side of the elongating dorsal tube (Fig. 2J), and in the epidermis that overlies the tube. After the fusion of the tube (stage D), *Pax3/7* transcripts were no longer localized in the epidermis, but in mesenchymal cells overlying the dorsal tube (Fig. 2K). The detection of the pan-neuronal marker ELAV related gene *Etr* was observed from stage C2 in scattered cells of the dorsal tube and in the mesenchymal cells around the tube (Fig. 2J). Some *Etr*+ cells in the tube and in the mesenchyme also co-expressed *Pax3/7* (Fig. 2J). The larval neural tube marker *Zic-r.a* retained expression along the dorsal tube during elongation (Fig. 2L) and after fusion. In stage C2 the neural marker *Ebf* was expressed in single cells along the entire dorsal tube and the overlying mesenchyme (Fig. 2L, M), some of these cells co-expressed the second Pou-domain containing transcription factor *Pou4* (Fig. 2M), a TF necessary for terminal differentiation of specific sensory neurons in different metazoan phyla[51]. *Notch* (Fig. 2N) was expressed in scattered cells all along the tube during its elongation and fusion.

AP patterning of the dorsal tube during the formation of cerebral ganglion

At stage A1 of the colony, the CNS of the primary bud is further developed than the secondary bud at stage C mentioned above (Fig. 3A-B). The posterior side of the dorsal tube detaches [26], and a delamination of cells from the tube migrate ventrally forming the cerebral ganglion [26,27,60]. At this stage *Otx* expression was confined to the anterior pole of the tube, as well as to the dorsal epithelia of the branchial basket and peribranchial chambers (Fig. 3 C, C'). *IrxB* expression was restrained to a central region of the dorsal tube (Fig. 3D, D'). *Gli* maintained its domain of expression in the middle of the dorsal tube posterior to the *Otx* expression domain and showing only a small overlap (Fig. 3E). Only a scattered part of the *Gli* expressing cells co-expressed the neural marker *Ebf* (Fig. 3F, F_a, F'). *Pou3* expression split in two domains: low expression in anterior cells and strong expression on cells located on the posterior third of the dorsal tube (Fig. 3G, G').

The patterning markers *Hox3* (Fig. 3H, H') and *Pax3/7* (Fig. 3I, I') were both expressed in the posterior portion of the dorsal tube. *Pax3/7* was no longer expressed in the anterior tube, but showed an additional expression in mesenchymal cells located in the dorsal and anterior region of the dorsal tube. *Zic-r.a* (Fig. 3J, J') was expressed over the entire length of the dorsal tube. The neural markers *Ebf*, *Etr* (Fig. 3K, K'), and *Pou4* (Fig. 3L) were expressed in scattered cells along the dorsal tube. Only part of the *Ebf*+ cells co-expressed *Etr* (Fig. 3K'). *Notch* (Fig. 3M) was expressed over the entire length of the dorsal tube.

Expression of CNS patterning genes and neural markers in the adult neural complex

The mature neural complex of a *B. schlosseri* blastozooid is composed of four structures: the ciliated funnel (CF), which is in continuity with the neural gland (NG), the dorsal organ (DO), and the cerebral ganglion (CG) [26](Fig. 4A).

The CNS patterning genes that were still expressed in the fully developed neural complex include: *IrxB*, which was expressed in scattered cells in the NG and CG (Fig. 4B); *Pou3*, which was expressed in NG, CG and DO (Fig. 4C); *Hox3*, expressed in the posterior region of the NG, the DO, and in a patch of cells in the postero-ventral side of the CG (Fig. 4D); and *Pax3/7* that was expressed at the anterior pole of the NG and the antero-dorsal region of the CG (Fig. 4E). *Gli* and *Otx* were not detected in the adult neural complex. *Zic-r.a* was expressed throughout the CF, NG and DO but not in the CG (Fig. 4F, F', F"). *Ebf* was expressed in scattered cells of both CG and NG (Fig. 4G, H)*. Etr* was expressed in the entire CG and in scattered cells of the NG (Fig.5H). Expression of the transmembrane protein *Notch* was restricted to the NG region (Fig. 4I), whereas *Pou4* was expressed in the CG and NG (Fig. 4J).

DISCUSSION

Dorsal tube formation and neurulation share similar AP patterning

Our data highlight comparable AP expression patterns between the stereotypical chordate neurulation [3,5,30](Fig. 5), and the formation of the blastogenetic dorsal tube in *B. schlosseri*. As described before the dorsal tube is a developmental structure that generates the adult neural complex during asexual development [26]. Here we show a common tripartite molecular regionalization during chordate CNS tubular formation, independent of its developmental origin, whereas *Otx* is expressed anteriorly, *IrxB* and *Gli* in a central region, and *Hox3* and *Pax3/7* at the posterior end of the tube. The regionalized expression of these genes is transient and begins in the young bud, at stage B2, in a region defined by the presence of TFs associated to ectodermal tissue fates[61]. Within this epithelial domain of the internal vesicle of the young bud, the dorsal tube arises as a thickening, and displays a bipartite expression of the transcription factors *Pou3* and *Hox3.* Pou3 has been shown to regionalize the nervous system during development of many taxa [62,63]. In early stages of larval development of ascidians, the CNS exhibits an early bipartition reflected by an absence of a gap between *Otx* and *Hox* gene expression [6], which is followed by a the later insertion of a *Pax2/5/8a+/Otx-/Hox-* domain that establishes the tripartite partitioning of the larval brain (Wada and Satoh, 2001). Concordantly, in *Botryllus* blastogenesis, we find the formation of a middle region during the progression of the dorsal tube elongation, which corresponds to a region of expression of *IrxB* and *Gli* between the *Otx* and *Hox3* domains at stage D. While present in the *B. schlosseri* genome, in this study we were not able to amplify *Pax2/5/8a*, a marker of the neck region in *Ciona*; hence we used the zinc finger TF Gli, a downstream target of *Pax2/5/8a* [64]. In addition, we analyzed *IrxB,* which in vertebrates is expressed posterior to the *zona limitans intrathalamica*, the border between the thalamus and prethalamus[36,65].Both *Gli* and *IrxB* were expressed in the central gap region between *Otx* and *Hox* in stage A1 bud. Interestingly, before expression became restricted to the mid region prior the dorsal tube differentiation, the expression of both TFs started broader during dorsal tube elongation in the early tubulogenesis. Such extended AP domain might reflect a generally broader expression of both, *Gli and IrxB,* observed in non-ascidian chordates (Fig. 5C)[3].

The tripartite regionalization of the dorsal tube reflects largely the divisions of the larval ascidian brain. In contrast, genetic markers of the nerve cord in the larval tail were not expressed during dorsal tube blastogenetic development. For instance, posterior *Hox5* (Gionti et al, 1998) was detected in the peribranchial chamber only, but not in the dorsal tube (Suppl. Mat. 13).

The AP patterning genes analyzed here are not exclusively expressed during larval CNS development, but are also involved in the regionalization of other tunicate tissues with AP extension, such as the endostyle[32]. They are also known to be expressed in post-metamorphic structures, such as the heart and pharynx [34,66,67]. The three AP tripartite molecular domains observed here may be responsible for the morphogenetic boundaries of the neural gland complex, i.e. CF/NG/DO (Fig. 3.A). Hence, an anterior *Otx*+ domain delimits the territory of the CF, a median *Gli+/IrxB+* delimits the median NG, and a posterior *Pax3/7+/Hox3+* domain defines the posterior DO respectively (Fig. 5). We hypothesize further that the three domains may also be important for generating different neuronal cell types that delaminate and eventually reside in the CG.

Dorsal tube as center of neurogenesis during blastogenetic development

Histological and ultrastructural description of asexual neurogenesis defines the dorsal tube also as a cellular source for cerebral ganglion development. In particular Burighel et al. (1998)[27] described the delamination and migration of cells from the dorsal tubular epithelium which migrate, cluster and differentiate into the neurons of the cerebral ganglion. In addition, the patterning observed in the dorsal tube occurs in tandem with the expression of neural and neurogenic markers such as *Etr*, *Zic-r.a*, *Ebf* and *Pou* genes. Thus expression dynamics suggests a progressive neural fate determination from a population of cells in the dorsal tube epithelia, and in cells delaminating from it, which corresponds to morphological observations present in literature[27,68]. It was previously suggested that during larval development and metamorphosis in *B. schlosseri*, most of the adult neural complex originated from the neurohypophyseal duct in the larval CNS[68], which followed a similar morphogenetic mechanisms described during blastogenesis[69]. On the other hand, in *Ciona* metamorphosis, the post-metamorphic ganglion was formed from ependymal cells of the CNS, and the position of cells along the AP axis reflects their eventual positioning in the adult cerebral ganglion, suggesting that it was the patterning of the larval nervous system that provide the positional information to pattern the adult cerebral ganglion [7,22–24]. While the patterning of the larval CNS is conserved in *B. schlosseri* (Suppl. Mat. 13) we did not detect any patterning genes in the neurohypophyseal duct. It is then intriguing to speculate that, in a non-embryonic context, the patterning of the dorsal tube may provide the spatial cues for neuronal guidance and development in the cerebral ganglion.

Dorsal tube: neural and non-neural nature

During dorsal tube differentiation, not all the *Ebf*+ mesenchymal cells express the neural differentiation markers *Etr* [46]or *Pou4*. In *Ciona, Ebf* is not only responsible for the decision of cholinergic motor neuron over ependymal cells fate in the larval CNS[70] but is also important for muscle formation[71]. Interestingly, we also detected myogenic TFs in the *Botryllus* dorsal tube and delaminating cells (Prünster, in prep.). Therefore, we argue that the dorsal tube and delaminating mesenchymal cells are not exclusively committed to neural fates.

Even though the ascidian specific transcription factor *Zic-r.a* is a maternal factor that determines muscular fate primary[72,73] its zygotic expression in the anterior part of the sensory vesicle, in the neck, the ganglion, and along the nerve cord suggests a secondary role in neural development [74,75]. In *Botryllus* asexual development *Zic-r.a* represents a pan-dorsal tube marker, from the onset of dorsal tube formation until its final differentiation into the CF, NG and DO. Without clear live tracking of the dorsal tube derived epithelial mesenchymal cells it is not possible to follow the dynamics of any delaminating cells. Nevertheless, these patterns of expression suggest that the dorsal tube is not exclusively a source of neural cells but may contribute to other cell types. While the role of the dorsal tube has not been functionally tested, this structure shows complex dynamics of cell commitment, cell differentiation, and cell migration. In this context, the presence of *Notch*+ cells, could play an inhibitory role on their differentiation[76,77]. Unfortunately, chemically blocking Notch pathway resulted into a lethal phenotype (data not shown), so any effect on cell type specification could not be addressed.

Neural complex: neural and non-neural nature

Despite its name, the neural gland and associated structures, including the CF and the DO, have been considered the non-neuronal part of the neural complex[14]. Some authors support the homology of the ascidian neural gland complex with the vertebrate pituitary or adenohypophysis[78,79], whereas others dismiss this homology and infer an osmoregulatory function[19] (reviewed in Burighel and Cloney, 1997) [16]. When analyzed histologically, the NG is a homogeneous structure with no obvious differences along the AP axis[19,80] and composed of mainly phagocytic cells, without evidence of neuronal cell types[81]. Our observations of a differential expression of *Pax3/7* and *Hox3* along the AP axis of the NG in *Botryllus* indicates a molecular maintenance of territory identity after the separation of the CF, NG, and DO. In addition, a continuous expression of the neural marker *Etr* and *Pou4* both in the differentiated ganglion as well as in cells of the NG suggests a perpetual differentiation of neurons in this region.

Conclusion

During non-embryonic developments such as *Botryllus* blastogenesis, different morphogenetic and ontogenetic events lead to similar structures, often co-opting molecular regulatory pathways from embryogenesis[82–84]. The observation of a transitory tubular structure that leads to the development of an adult CNS during asexual development, allows us to draw a raw map of gene expression patterns that show a remarkable similarity to AP patterning of larval CNS development. This is the first study that describes a molecular patterning event during a non-embryonic development of an ascidian. The development of two tubular neurogenic structures, one during embryogenesis and metamorphosis, and another during blastogenesis, strongly suggests co-option of a patterning between two different ontogenesis. In this case, the partial co-option of the dorsal tube patterning could represent a provisional scaffolding in blastogenesis, just like the larval CNS functions as a provisional scaffold for the adult CNS during the sexual development.

ACKNOWLEDGEMENTS

We would like to thank Clare Hudson for her suggestions and the constructive discussions. Laurent Gilletta, Regis Lasbleiz and Sophie Collet for help with the aqua-culture systems. Sonia Lotito, Maxence Burtin, Marta Scelzo and Vincenzo Vellucci for technical and artistic support.

FUNDING SOURCES

This work was supported by ANR (ANR14-CE02-0019-01) and CNRS/PICS (07679) to S.T., FRM (#FDT20140931163) to L.R. and FAPESP Jovem Pesquisador grant (JP 2015/50164-5) to F.B.

REFERENCES

1. Martín-Durán JM, Pang K, Børve A, et al. Convergent evolution of bilaterian nerve cords. *Nature*. 2017;553(7686):45-50. doi:10.1038/nature25030.
2. Pani AM, Mullarkey EE, Aronowicz J, Assimacopoulos S, Grove EA, Lowe CJ. Ancient deuterostome origins of vertebrate brain signalling centres. *Nature*. 2012;483(7389):289-294. doi:10.1038/nature10838.
3. Holland LZ, Carvalho JE, Escriva H, Laudet V, Schubert M, Shimeld SM. Evolution of bilaterian central nervous systems : a single origin ? *Evodevo*. 2013;4(1):1. doi:10.1186/2041-9139-4-27.
4. Hudson C. The central nervous system of ascidian larvae. *Wiley Interdiscip Rev Dev Biol*. 2016;5(5):538-561. doi:10.1002/wdev.239.
5. Wada H, Saiga H, Satoh N, Holland PW. Tripartite organization of the ancestral chordate brain and the antiquity of placodes: insights from ascidian Pax-2/5/8, Hox and Otx genes. *Development*. 1998;125(6):1113-1122. http://www.ncbi.nlm.nih.gov/pubmed/9463358.
6. Ikuta T, Saiga H. Dynamic change in the expression of developmental genes in the ascidian central nervous system : Revisit to the tripartite model and the origin of the midbrain – hindbrain boundary region. 2007;312:631-643. doi:10.1016/j.ydbio.2007.10.005.
7. Dufour HD, Chettouh Z, Deyts C, et al. Precraniate origin of cranial motoneurons. *Proc Natl Acad Sci*. 2006;103(23):8727-8732. doi:10.1073/pnas.0600805103.
8. Hamada M, Shimozono N, Ohta N, et al. Expression of neuropeptide- and hormone-encoding genes in the Ciona intestinalis larval brain. *Dev Biol*. 2011;352(2):202-214. doi:10.1016/j.ydbio.2011.01.006.
9. Moret F, Christiaen L, Deyts C, Blin M, Vernier P, Joly JS. Regulatory gene expressions in the ascidian ventral sensory vesicle: Evolutionary relationships with the vertebrate hypothalamus. *Dev Biol*. 2005;277(2):567-579. doi:10.1016/j.ydbio.2004.11.004.
10. Imai KS, Stolfi A, Levine M, Satou Y. Gene regulatory networks underlying the compartmentalization of the Ciona central nervous system. *Development*. 2009;136(2):285-293. doi:10.1242/dev.026419.

11. Locascio a, Aniello F, Amoroso a, Manzanares M, Krumlauf R, Branno M. Patterning the ascidian nervous system: structure, expression and transgenic analysis of the CiHox3 gene. *Development*. 1999;126(21):4737-4748. http://www.ncbi.nlm.nih.gov/pubmed/10518491.
12. Ikuta T, Yoshida N, Satoh N, Saiga H. Ciona intestinalis Hox gene cluster: Its dispersed structure and residual colinear expression in development. *Proc Natl Acad Sci U S A*. 2004;101(42):15118-15123. doi:10.1073/pnas.0401389101.
13. Cloney RA. Ascidian larvae and the events of metamorphosis. *Integr Comp Biol*. 1982;22(4):817-826. doi:10.1093/icb/22.4.817.
14. Mackie GO, Burighel P. The nervous system in adult tunicates : current research directions 1. *Camera*. 2005;183:151-183. doi:10.1139/Z04-177.
15. Tiozzo S, Murray M, Degnan BM, De Tomaso AW, Croll RP. Development of the neuromuscular system during asexual propagation in an invertebrate chordate. *Dev Dyn*. 2009;238(8):2081-2094. doi:10.1002/dvdy.22023.
16. Burighel P, Cloney RA. Urochordata: Ascidiacea. In: Harrison FW;Ruppert EE, ed. *Microscopic Anatomy of Invertebrates, Volume 15, Hemichordata, Chaetognatha, and the Invertebrate Chordates.* Wiley-Liss. New York: Wiley-Liss; 1997:221-347.
17. Osugi T, Sasakura Y, Satake H. The nervous system of the adult ascidian Ciona intestinalis type A (Ciona robusta): Insights from transgenic animal models. *PLoS One*. 2017;12(6):1-16. doi:10.1371/journal.pone.0180227.
18. Lane NJ. Neurosecretory Cells in the Cerebral Ganglion of Adult Tunicates : Fine Structure and Distribution of Phosphatases. 1972;497:15-17.
19. Deyts C, Casane D, Vernier P, Bourrat F, Joly JS. Morphological and gene expression similarities suggest that the ascidian neural gland may be osmoregulatory and homologous to vertebrate peri-ventricular organs. *Eur J Neurosci*. 2006;24(8):2299-2308. doi:10.1111/j.1460-9568.2006.05073.x.
20. Terakado K. Placode Formation and Generation of Gonadotropin-Releasing Hormone (GnRH) Neurons in Ascidians. *Zoolog Sci*. 2009;26(6):398-405. doi:10.2108/zsj.26.398.
21. Nicol D, Meinertzhagen IA. Cell counts and maps in the larval central nervous system of the ascidian ciona intestinalis (L.). *J Comp Neurol*. 1991;309(4):415-429. doi:10.1002/cne.903090402.
22. Horie T, Shinki R, Ogura Y, Kusakabe TG, Satoh N, Sasakura Y. Ependymal cells of

chordate larvae are stem-like cells that form the adult nervous system. *Nature*. 2011;469(7331):525-528. doi:10.1038/nature09631.

23. Sasakura Y, Hozumi A. Formation of adult organs through metamorphosis in ascidians. *Wiley Interdiscip Rev Dev Biol*. 2017;1:e304. doi:10.1002/wdev.304.
24. Hozumi A, Horie T, Sasakura Y. Neuronal map reveals the highly regionalized pattern of the juvenile central nervous system of the ascidian Ciona intestinalis. *Dev Dyn*. 2015. doi:10.1002/dvdy.24317.
25. Minelli A. *The Development of Animal Form : Ontogeny, Morphology, and Evolution*. Cambridge University Press; 2003.
26. Manni L, Gasparini F, Hotta K, et al. Ontology for the asexual development and anatomy of the colonial chordate Botryllus schlosseri. *PLoS One*. 2014;9(5):e96434. doi:10.1371/journal.pone.0096434.
27. Burighel P, Lane NJ, Zaniolo G, Manni L. Neurogenic role of the neural gland in the development of the ascidian, Botryllus schlosseri (Tunicata, Urochordata). *J Comp Neurol*. 1998;394(2):230-241. http://www.ncbi.nlm.nih.gov/pubmed/9552128.
28. Sasaki A, Miyamoto Y, Satou Y, Satoh N, Ogasawara M. Novel endostyle-specific genes in the ascidian Ciona intestinalis. *Zoolog Sci*. 2003;20(8):1025-1030. http://www.ncbi.nlm.nih.gov/pubmed/12951410.
29. Yoshida K, Nakahata A, Treen N, Sakuma T, Yamamoto T, Sasakura Y. Hox-mediated endodermal identity patterns pharyngeal muscle formation in the chordate pharynx. *Development*. 2017;144(9):1629-1634. doi:10.1242/dev.144436.
30. Albuixech-Crespo B, López-Blanch L, Burguera D, et al. *Molecular Regionalization of the Developing Amphioxus Neural Tube Challenges Major Partitions of the Vertebrate Brain*. Vol 15.; 2017. doi:10.1371/journal.pbio.2001573.
31. Hudson C, Lemaire P. Induction of anterior neural fates in the ascidian Ciona intestinalis. *Mech Dev*. 2001;100(2):189-203. http://www.ncbi.nlm.nih.gov/pubmed/11165477.
32. Cañestro C, Bassham S, Postlethwait JH. Evolution of the thyroid: Anterior-posterior regionalization of the Oikopleura endostyle revealed by Otx, Pax2/5/8, and Hox1 expression. *Dev Dyn*. 2008;237(5):1490-1499. doi:10.1002/dvdy.21525.
33. Cañestro C, Bassham S, Postlethwait J. Development of the central nervous system in the larvacean Oikopleura dioica and the evolution of the chordate brain. *Dev Biol*. 2005;285(2):298-315. doi:10.1016/j.ydbio.2005.06.039.

34. Hinman VF, Degnan BM. Retinoic acid perturbs Otx gene expression in the ascidian pharynx. *Dev Genes Evol*. 2000;210(3):129-139. doi:10.1007/s004270050019.
35. Islam AFMT, Moly PK, Miyamoto Y, et al. Distinctive Expression Patterns of Hedgehog Pathway Genes in the Ciona intestinalis Larva : Implications for a Role of Hedgehog Signaling in Postembryonic Development and Chordate Evolution Distinctive Expression Patterns of Hedgehog Pathway Genes in the . 27(2):84-90. doi:10.2108/zsj.27.84.
36. Rodríguez-Seguel E, Alarcón P, Gómez-Skarmeta JL. The Xenopus Irx genes are essential for neural patterning and define the border between prethalamus and thalamus through mutual antagonism with the anterior repressors Fezf and Arx. *Dev Biol*. 2009;329(2):258-268. doi:10.1016/j.ydbio.2009.02.028.
37. Spaniol P, Bornmann C, Hauptmann G, Gerster T. Class III POU genes of zebrafish are predominantly expressed in the central nervous system. *Nucleic Acids Res*. 1996;24(24):4874-4881. doi:10.1093/nar/24.24.4874.
38. Nakayama S, Satou K, Orito W, Ogasawara M. Ordered expression pattern of Hox and ParaHox genes along the alimentary canal in the ascidian juvenile. *Cell Tissue Res*. 2016:65-75. doi:10.1007/s00441-016-2360-7.
39. Mazet F, Hutt JA, Milloz J, Millard J, Graham A, Shimeld SM. Molecular evidence from Ciona intestinalis for the evolutionary origin of vertebrate sensory placodes. *Dev Biol*. 2005;282(2):494-508. doi:10.1016/j.ydbio.2005.02.021.
40. Matsunaga E, Araki I, Nakamura H. Role of Pax3/7 in tectum development. 2001;4077:4069-4077.
41. Miyamoto N, Wada H. Hemichordate neurulation and the origin of the neural tube. *Nat Commun*. 2013;4:1-8. doi:10.1038/ncomms3713.
42. Gostling NJ, Shimeld SM. Protochordate Zic genes define primitive somite compartments and highlight molecular changes underlying neural crest evolution. *Evol Dev*. 2003;5(2):136-144. http://www.ncbi.nlm.nih.gov/pubmed/12622730.
43. Aruga J. The role of Zic genes in neural development. *Mol Cell Neurosci*. 2004;26(2):205-221. doi:10.1016/j.mcn.2004.01.004.
44. Dubois L, Vincent A. The COE - Collier/Olf1/EBF - Transcription factors: Structural conservation and diversity of developmental functions. *Mech Dev*. 2001;108(1-2):3-12. doi:10.1016/S0925-4773(01)00486-5.
45. Akanuma T, Hori S, Darras S, Nishida H. Notch signaling is involved in nervous

system formation in ascidian embryos. *Dev Genes Evol*. 2002;212(10):459-472. doi:10.1007/s00427-002-0264-x.

46. Yagi K, Makabe KW. Isolation of an early neural maker gene abundantly expressed in the nervous system of the ascidian, Halocynthia roretzi. *Dev Genes Evol*. 2001;211(1):49-53. doi:10.1007/s004270000118.

47. Knecht AK, Good PJ, Dawid IB, Harland RM. Dorsal-ventral patterning and differentiation of noggin-induced neural tissue in the absence of mesoderm. *Development*. 1995;121(6):1927-1935.

48. Hori S, Saitoh T, Matsumoto M, Makabe KW, Nishida H. Notch homologue from Halocynthia roretzi is preferentially expressed in the central nervous system during ascidian embryogenesls. *Dev Genes Evol*. 1997;207(6):371-380. doi:10.1007/s004270050126.

49. Holland LZ, Rached L a, Tamme R, et al. Characterization and developmental expression of the amphioxus homolog of Notch (AmphiNotch): evolutionary conservation of multiple expression domains in amphioxus and vertebrates. *Dev Biol*. 2001;232(2):493-507. doi:10.1006/dbio.2001.0160.

50. Louvi A, Artavanis-Tsakonas S. Notch signalling in vertebrate neural development. *Nat Rev Neurosci*. 2006;7(2):93-102. doi:10.1038/nrn1847.

51. Candiani S, Pennati R, Oliveri D, et al. Ci-POU-IV expression identifies PNS neurons in embryos and larvae of the ascidian Ciona intestinalis. *Dev Genes Evol*. 2005;215(1):41-45. doi:10.1007/s00427-004-0444-y.

52. Candiani S, Oliveri D, Parodi M, Bertini E, Pestarino M. Expression of AmphiPOU-IV in the developing neural tube and epidermal sensory neural precursors in amphioxus supports a conserved role of class IV POU genes in the sensory cells development. *Dev Genes Evol*. 2006;216(10):623-633. doi:10.1007/s00427-006-0083-6.

53. He X, Treacy MN, Simmons DM, Ingraham HA, Swanson LW, Rosenfeld MG. Expression of a large family of POU-domain regulatory genes in mammalian brain development. *Nature*. 1989;340(6228):35-41. doi:10.1038/340301a0.

54. Langenbacher AD, Rodriguez D, Di Maio A, De Tomaso AW. Whole-mount fluorescent in situ hybridization staining of the colonial tunicate Botryllus schlosseri. *Genesis*. 2015;53(1):194-201. doi:10.1002/dvg.22820.

55. Lauzon RJ, Ishizuka KJ, Weissman IL. Cyclical Generation and Degeneration of Organs in a Colonial Urochordate Involves Crosstalk between Old and New : A

Model for Development and Regeneration. 2002;348:333-348. doi:10.1006/dbio.2002.0772.

56. Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. http://www.lirmm.fr/~gascuel. Accessed January 25, 2018.

57. Ricci L, Cabrera F, Lotito S, Tiozzo S. Redeployment of germ layers related TFs shows regionalized expression during two non-embryonic developments. *Dev Biol*. 2016;416(1). doi:10.1016/j.ydbio.2016.05.016.

58. Rueden CT, Schindelin J, Hiner MC, et al. ImageJ2: ImageJ for the next generation of scientific image data. *BMC Bioinformatics*. 2017;18(1):1-26. doi:10.1186/s12859-017-1934-z.

59. Mazet F, Hutt J, Millard J, Shimeld S. Pax gene expression in the developing central nervous system of Ciona intestinalis. *Gene Expr Patterns*. 2003;3(6):743-745. doi:10.1016/S1567-133X(03)00137-6.

60. Manni L, Lane NJ, Burighel P, Zaniolo G. Are neural crest and placodes exclusive to vertebrates? *Evol Dev*. 2001;3(5):297-298. http://www.ncbi.nlm.nih.gov/pubmed/11710760.

61. Ricci L, Cabrera F, Lotito S, Tiozzo S. Re-deployment of germ layers related TFs shows regionalized expression during two non-embryonic developments. *Dev Biol*. 2016;(in press).

62. Wollesen T, McDougall C, Degnan BM, Wanninger A. POU genes are expressed during the formation of individual ganglia of the cephalopod central nervous system. *Evodevo*. 2014;5(1):1-15. doi:10.1186/2041-9139-5-41.

63. Alvarez-Bolado G, Rosenfeld MG, Swanson LW. Model of forebrain regionalization based on spatiotemporal patterns of POU-III homeobox gene expression, birthdates, and morphological features. *J Comp Neurol*. 1995;355(2):237-295. doi:10.1002/cne.903550207.

64. Imai KS, Stolfi A, Levine M, Satou Y. Gene regulatory networks underlying the compartmentalization of the Ciona central nervous system. 2009;293:285-293. doi:10.1242/dev.026419.

65. Sena E, Feistel K, Durand B. An Evolutionarily Conserved Network Mediates Development of the zona limitans intrathalamica, a Sonic Hedgehog-Secreting

Caudal Forebrain Signaling Center. *J Dev Biol*. 2016;4(4):31. doi:10.3390/jdb4040031.

66. Yoshida K, Nakahata A, Treen N, Sakuma T, Yamamoto T, Sasakura Y. *Hox* -mediated endodermal identity patterns the pharyngeal muscle formation in the chordate pharynx. *Development*. 2017;(March):dev.144436. doi:10.1242/dev.144436.
67. Ogasawara M, Tanaka KJ, Makabe KW, Satoh N. Expression of endostyle-specific genes in the ascidian Halocynthia roretzi. *Dev Genes Evol*. 1996;206(3):227-235. doi:10.1007/s004270050048.
68. Manni L, Lane NJ, Sorrentino M, Zaniolo G, Burighel P. Mechanism of neurogenesis during the embryonic development of a tunicate. *J Comp Neurol*. 1999;412(3):527-541. http://www.ncbi.nlm.nih.gov/pubmed/10441238.
69. Manni L, Burighel P. Common and divergent pathways in alternative developmental processes of ascidians. *Bioessays*. 2006;28(9):902-912. doi:10.1002/bies.20462.
70. Kratsios P, Stolfi A, Levine M, Hobert O. Coordinated regulation of cholinergic motor neuron traits through a conserved terminal selector gene. *Nat Neurosci*. 2012;15(2):205-214. doi:10.1038/nn.2989.
71. Razy-Krajka F, Lam K, Wang W, et al. Collier/OLF/EBF-Dependent Transcriptional Dynamics Control Pharyngeal Muscle Specification from Primed Cardiopharyngeal Progenitors. *Dev Cell*. 2014;29(3):263-276. doi:10.1016/j.devcel.2014.04.001.
72. Nishida H, Sawada K. macho-1 encodes a localized mRNA in ascidian eggs that specifies muscle fate during embryogenesis. *Nature*. 2001;409(February):724-729. doi:10.1038/35055568.
73. Sawada K, Fukushima Y, Nishida H. Macho-1 functions as transcriptional activator for muscle formation in embryos of the ascidian Halocynthia roretzi. *Gene Expr Patterns*. 2005;5(3):429-437. doi:10.1016/j.modgep.2004.09.003.
74. Satou Y, Yagi K, Imai KS, Yamada L, Nishida H, Satoh N. Macho-1-related genes in Ciona embryos. *Dev Genes Evol*. 2002;212(2):87-92. doi:10.1007/s00427-002-0218-3.
75. Satou Y, Imai KS. Ascidian Zic Genes. In: Springer, Singapore; 2018:87-106. doi:10.1007/978-981-10-7311-3_6.
76. Nye JS, Kopan R, Axel R. An activated Notch suppresses neurogenesis and myogenesis but not gliogenesis in mammalian cells. *Development*.

1994;120(9):2421-2430. doi:10.1016/0168-9525(91)90220-k.

77. Tanigaki K, Honjo T. Two Opposing roles of RBP-J in Notch signaling. *Curr Top Dev Biol*. 2010;92(C):231-252. doi:10.1016/S0070-2153(10)92007-3.

78. Pestarino M. A pituitary-like role of the neural gland of an ascidian. *Gen Comp Endocrinol*. 1985;60(2):293-297. doi:10.1016/0016-6480(85)90326-0.

79. Burighel P, Lane NJ, Zaniolo G, Manni L. Neurogenic role of the neural gland in the development of the ascidian, Botryllus schlosseri (Tunicata, Urochordata). *J Comp Neurol*. 1998;394(2):230-241. doi:10.1002/(SICI)1096-9861(19980504)394:2<230::AID-CNE7>3.0.CO;2-3.

80. Lane NJ. The Neural Gland in Tunicates : Fine Structure and Intracellular Distribution of Phosphatases. 1971;93:80-93.

81. Ruppert EE. Structure, Ultrastructure and Function of the Neural Gland Complex of Ascidia interrupta (Chordata, Ascidiacea): Clarification of Hypotheses Regarding the Evolution of the Vertebrate Anterior Pituitary. *Acta Zool*. 1990;71(3):135-149. doi:10.1111/j.1463-6395.1990.tb01189.x.

82. Ricci L, Chaurasia A, Lapébie P, et al. Identification of differentially expressed genes from multipotent epithelia at the onset of an asexual development. *Sci Rep*. 2016;6. doi:10.1038/srep27357.

83. Tiozzo S, Christiaen L, Deyts C, Manni L, Joly J-S, Burighel P. Embryonic versus blastogenetic development in the compound ascidian Botryllus schlosseri: Insights from Pitx expression patterns. *Dev Dyn*. 2005;232(2). doi:10.1002/dvdy.20250.

84. Tiozzo S, De Tomaso AW. Functional analysis of Pitx during asexual regeneration in a basal chordate. *Evol Dev*. 2009;11(2). doi:10.1111/j.1525-142X.2009.00316.x.

FIGURE LEGENDS

Fig. 1. AP patterning and neurogenic markers are expressed early in the *B. schlosseri* secondary bud stage B2.

Confocal sections of budlets (A, B, D-G), dorsal is upwards. (C) Schematic drawing of *Botryllus schlosseri* budlet in stage C1 shows the first event of dorsal evagination of the inner vesicle (green) and the formation of peribranchial chambers (ochre), with an expansion of mRNA expression projected on top. Genes (yellow or magenta) are indicated in the lower left corner. Dashed lines indicate the border with the primary bud (Pb) and dashed circles outline the gonads. Scale bar is 50 μm. Nuclei are stained with Hoechst (cyan).

Fig. 2. Dynamic anteroposterior marker genes are expressed coupled with neurogenic genes in *B. schlosseri* during dorsal tube elongation and fusion.

(A) Schematic drawing of *Botryllus schlosseri* budlet in stage C2 shows dorsal evagination, elongation, and tubularization of the inner vesicle. Overlying mesenchymal cells (green) move ventrally (green arrow), while the peribranchial chambers form (ochre). States of expression (close-up in grey) are represented in the upper right corner of corresponding figures.

(B-C) Schematic drawing of the budlet in stage D shows the dorsal tube (green) attaching anteriorly to the branchial chamber (light ochre); presumptive gonads (blue) are located laterally (B) and transversally (C). (E, F, H, J, L, M, N) Confocal sections of budlets in stage C2 and stage D (D, G, I, K).

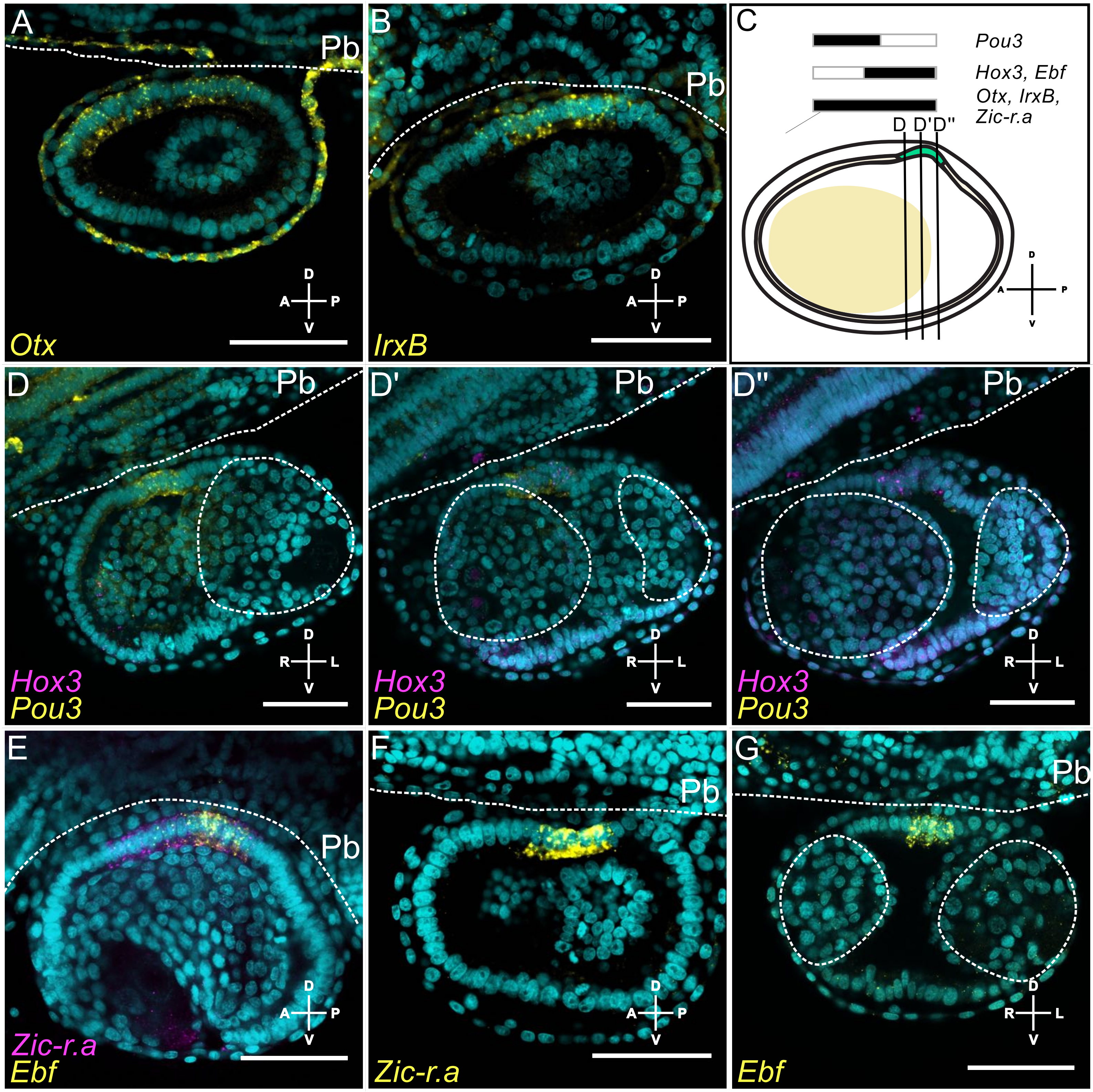
Genes (yellow or magenta) are indicated in the lower left corner in. The color border indicates colony stages (A-C). Dashed lines indicate the border with the primary bud (Pb). Asterisk indicates unspecific staining of the probe in the tunic. The scale bar is 50 μm. Nuclei are stained with Hoechst (cyan).

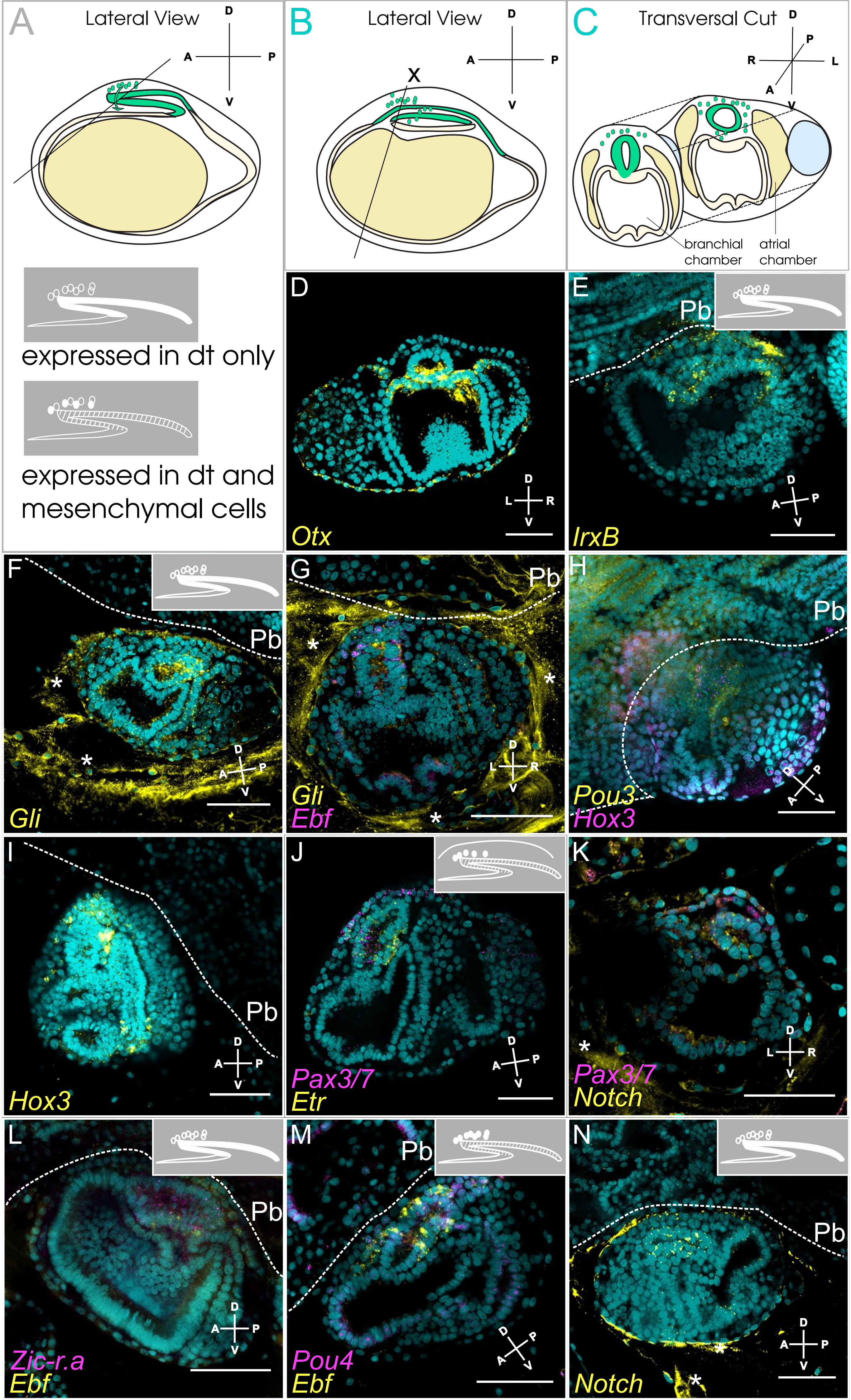
Fig. 3. AP neurogenic genes during ganglion development in the *B. schlosseri* primary bud.

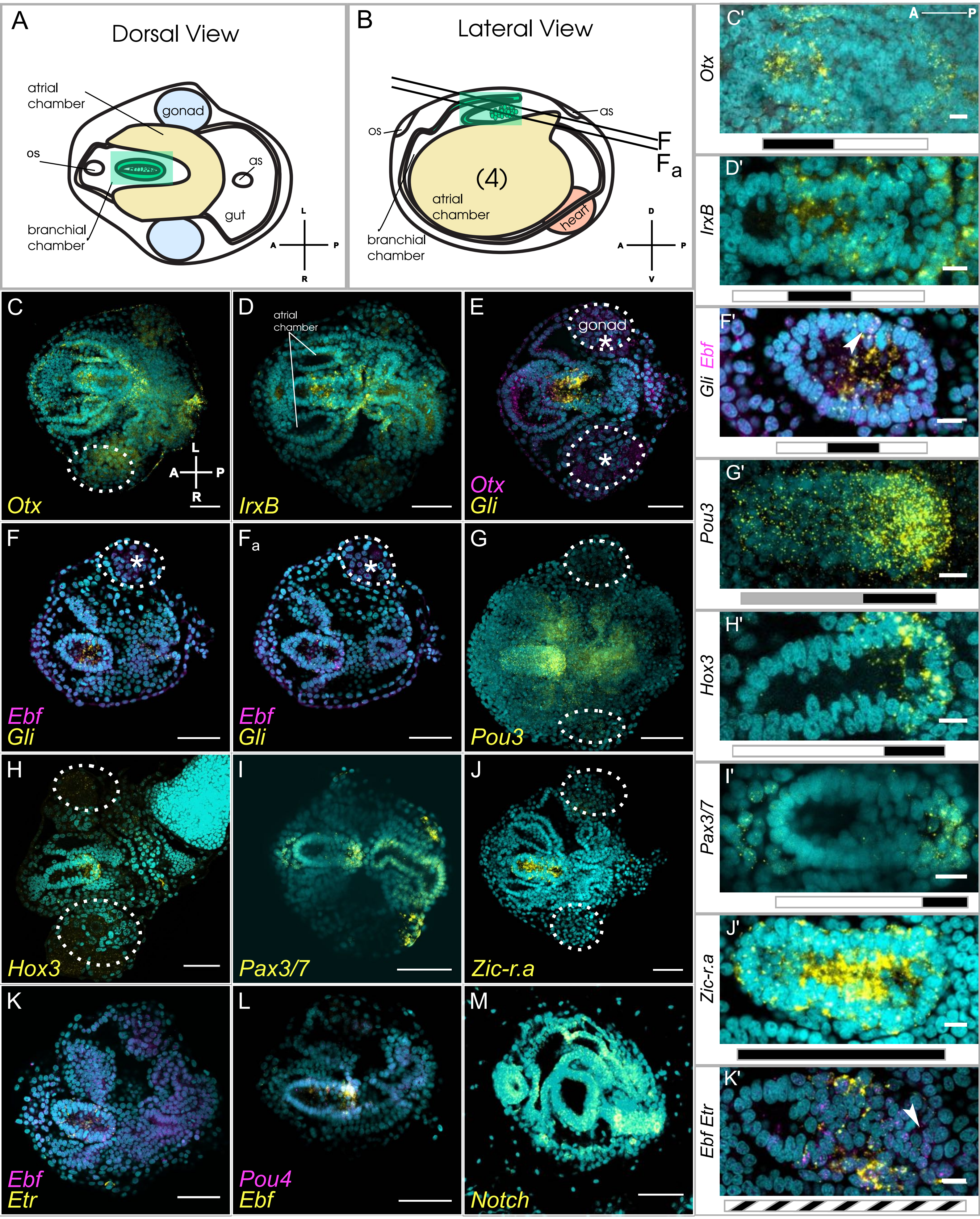
(A-B) Schematic drawing of *Botryllus schlosseri* bud in stage A1, the dorsal tube is detached posteriorly and the cerebral ganglion is shown as a cluster of cells ventral to the tube (green), and the forming peribranchial chambers are fused dorsally (ochre). The green box in (A) represents a close-up view of the dorsal tube shown in the pictures on the right column. Oral siphon (os) and atrial siphon (as) remain closed. (C-F',H,J-M) Confocal images of the dorsal side in stage A1 buds (G,I). Genes are indicated in the lower left corner of the pictures in yellow or magenta. (C',D',F",G'-K') Close-up view of the DT, black bar indicates the expansion of expression, dashed bar indicates scattered expression, grey relative low expression. (F') Arrow indicates co-expression of *Ebf* and *Gli* (K') *Ebf* expression. Asterisks indicate nonspecific staining of the probe trapping in the tunic and gonad. Scale bar is 50 μm and 10 μm in right column. Nuclei are stained with Hoechst (cyan).

Fig. 4. AP patterning and neurogenic markers in the neural complex of *B. schlosseri.*
(A) Schematic drawing of *Botryllus schlosseri* neural complex. (B,C,E,G,I,J) Confocal projection images of stacks of the neural complex and (D,F, F',F", H) confocal sections. (F) Bright field images of a zooid prior to siphon opening, close-up in (F'). Genes (yellow or magenta) are indicated in the lower left corner. Scale bar is 50 μm. Nuclei are stained with Hoechst (cyan).

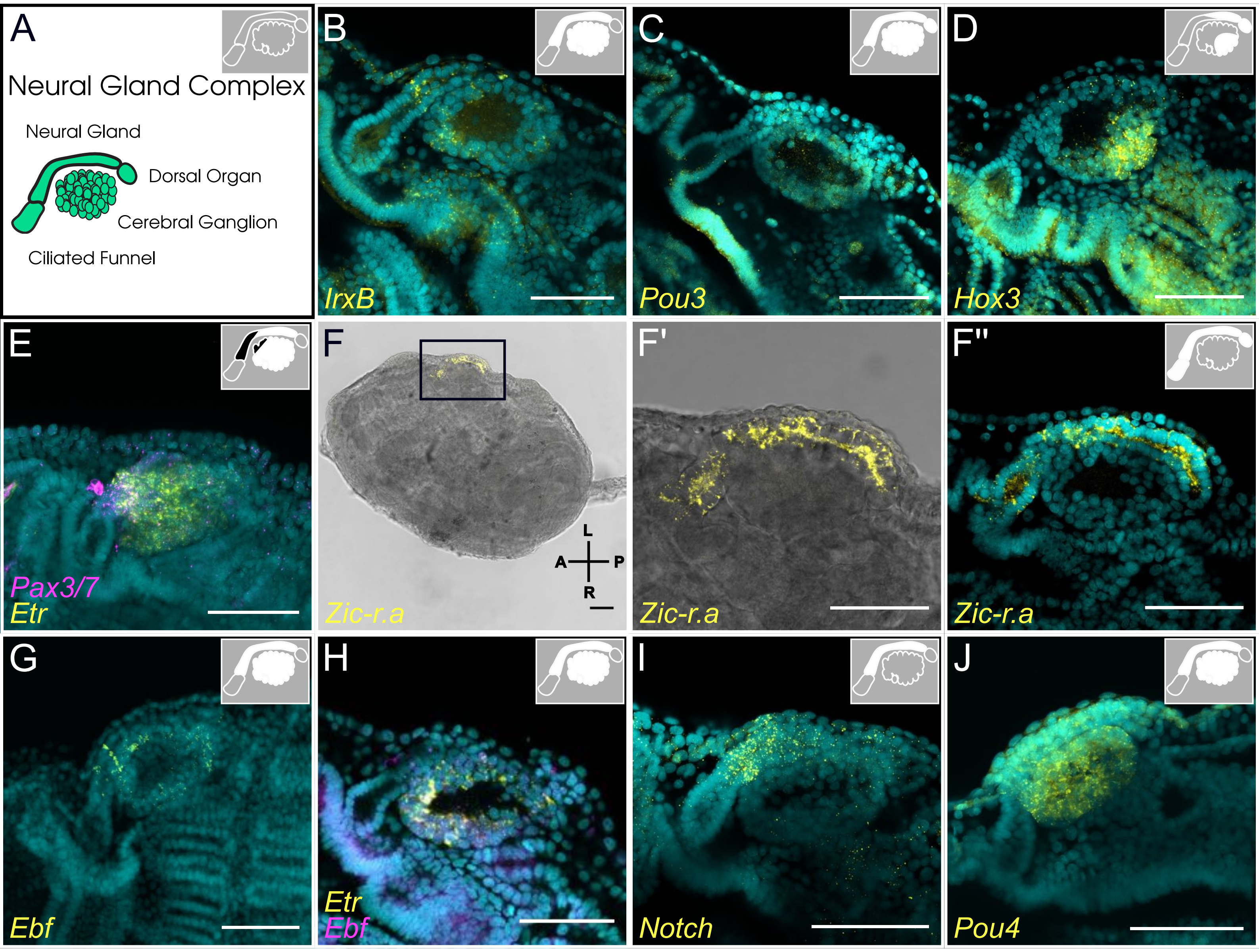
Fig. 5. AP patterning genes during larval CNS development and asexual dorsal tube development.
(A) AP patterning domains of expression in the larval CNS of *Ciona* and the dorsal tube in a stage A1 bud of *Botryllus schlosseri*
(B) Stage B2 (C) adult neural complex prior to siphon opening in *B. schlosseri*. (D) Chordate AP patterning genes in Amphioxus or a generic vertebrate (rat or chicken). Gene expression domains based on Alvarez-Blado et al.1995, Hudson & Lemaire, 2001, Matsunaga et al, 2001, Mazet et al. 2003, Imai et al. 2009, Holland et al. 2013, Albuxech-Crespo et al. 2017. ND, neurohypophyseal duct; SV, sensory vesicle; PSV, posterior sensory vesicle; VG, visceral ganglion; DT, dorsal tube; CG, cerebral ganglion; CF, ciliated funnel, NG, neural gland, DO, dorsal organ, CG cerebral ganglion, Tel, Di, Mes, Rhomb, Tel-,Di-,Mes-,Rhombenchephalon;

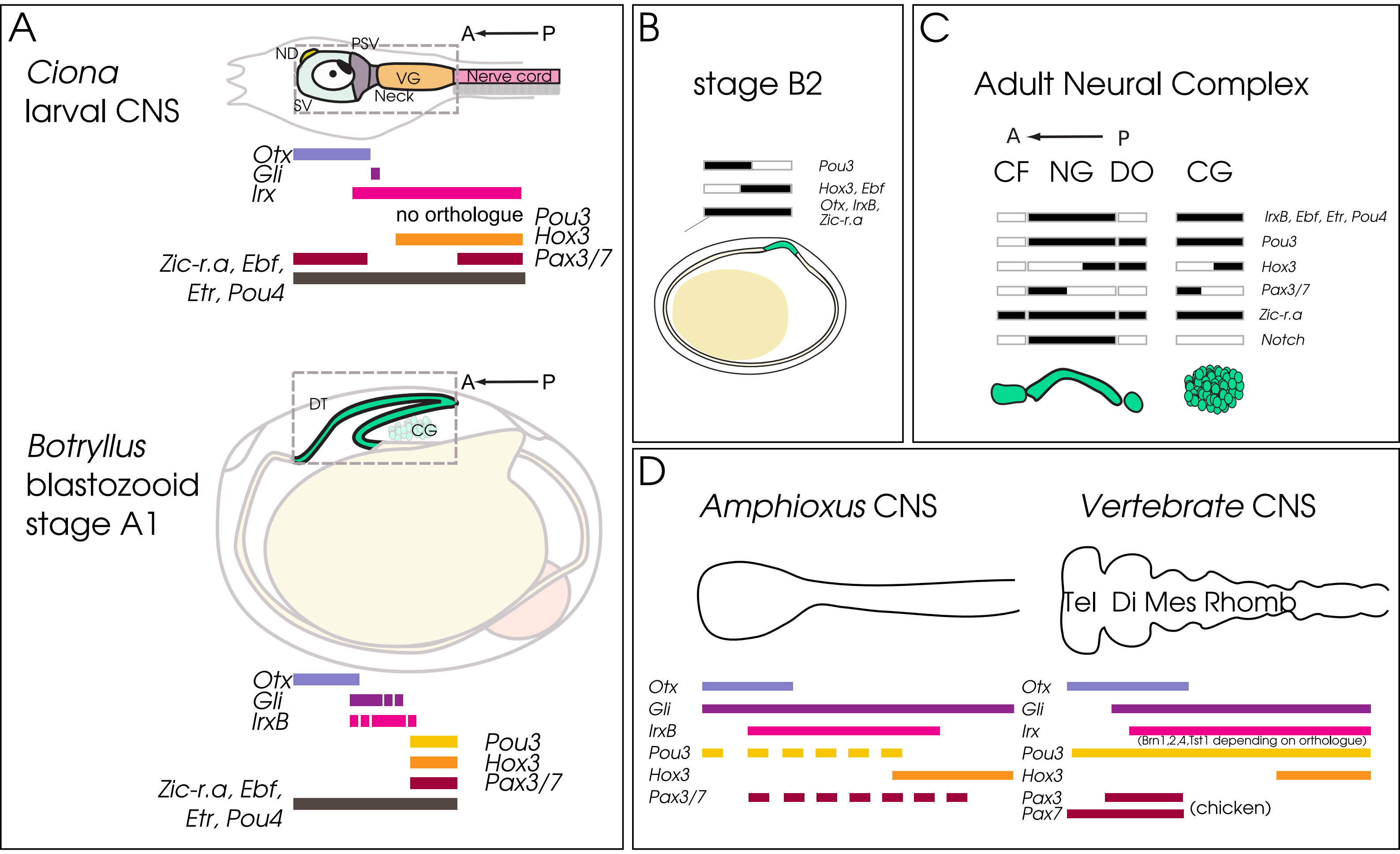

A
Pb
D
A
P
V
Otx
B
Pb
IrxB
C
Pou3
Hox3, Ebf
Otx, IrxB,
Zic-r.a
D D'D''
D
A
P
V
D
Pb
R
L
Hox3
Pou3
D'
Pb
Hox3
Pou3
D''
Pb
Hox3
Pou3
E
Pb
Zic-r.a
Ebf
F
Pb
Zic-r.a
G
Pb
Ebf

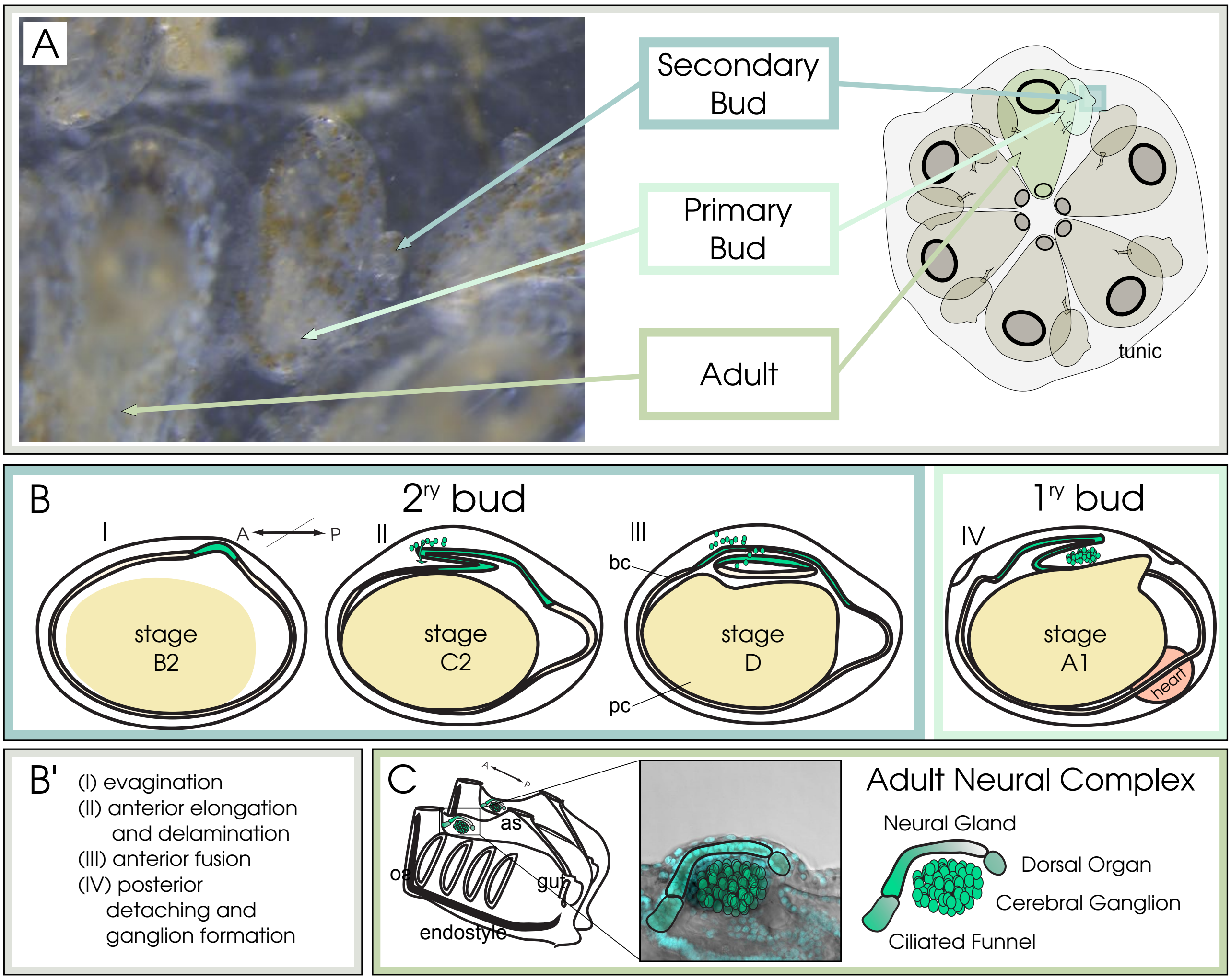

A
Lateral View
D
A
P
V
expressed in dt only
expressed in dt and
mesenchymal cells
B
Lateral View
X
C
Transversal Cut
R
L
branchial
chamber
atrial
chamber
D
Otx
E
Pb
IrxB
F
Gli
G
Gli
Ebf
H
Pou3
Hox3
I
Hox3
J
Pax3/7
Etr
K
Pax3/7
Notch
L
Zic-r.a
Ebf
M
Pou4
Ebf
N
Notch



A
Dorsal View
atrial chamber
gonad
os
as
gut
branchial chamber
B
Lateral View
os
as
F
Fa
(4)
atrial chamber
heart
branchial chamber
C
Otx
D
atrial chamber
IrxB
E
gonad
Otx
Gli
F
Ebf
Gli
Fa
Ebf
Gli
G
Pou3
H
Hox3
I
Pax3/7
J
Zic-r.a
K
Ebf
Etr
L
Pou4
Ebf
M
Notch
C'
Otx
D'
IrxB
F'
Gli Ebf
G'
Pou3
H'
Hox3
I'
Pax3/7
J'
Zic-r.a
K'
Ebf Etr


A
Neural Gland Complex
Neural Gland
Dorsal Organ
Cerebral Ganglion
Ciliated Funnel
B
IrxB
C
Pou3
D
Hox3
E
Pax3/7
Etr
F
L
A
P
R
Zic-r.a
F'
Zic-r.a
F''
Zic-r.a
G
Ebf
H
Etr
Ebf
I
Notch
J
Pou4


A
Ciona
larval CNS
A P
ND
PSV
SV
VG
Neck
Nerve cord
Otx
Gli
Irx
no orthologue
Pou3
Hox3
Pax3/7
Zic-r.a, Ebf,
Etr, Pou4
A P
DT
CG
Botryllus
blastozooid
stage A1
Otx
Gli
IrxB
Pou3
Hox3
Pax3/7
Zic-r.a, Ebf,
Etr, Pou4
B
stage B2
Pou3
Hox3, Ebf
Otx, IrxB,
Zic-r.a
C
Adult Neural Complex
A P
CF NG DO CG
IrxB, Ebf, Etr, Pou4
Pou3
Hox3
Pax3/7
Zic-r.a
Notch
D
Amphioxus CNS
Vertebrate CNS
Tel Di Mes Rhomb
Otx
Gli
IrxB
Pou3
Hox3
Pax3/7
Otx
Gli
Irx
(Brn1,2,4,Tst1 depending on orthologue)
Pou3
Hox3
Pax3
Pax7
(chicken)



Suppl. Fig. 1 *Botryllus schlosseri* colony and development of the dorsal tube during blastogenesis (asexual propagation).
(A) Overview of colony of *Botryllus schlosseri* showing three generations of individuals -adult, primary and secondary bud- connected with each other by vasculature and enbedded in a common tunic.
(B) Schematic draving of bud development, sagittal view, with focus on the forming neural complex (green). (B') Major steps of dorsal tube development, I, dorsal evagination of the inner vesicle, II, the structure becomes tubular by anterior elongation, and cells delaminate; III, the tube fuses anteriorly with the proximal branchial basket (bc), flanked by latteral peribranhcial chambers (pc); the delaminated cells migrate ventrally IV, to form the cerebral ganglion, the tube detaches posteriorly.
(C) Detail of adult neural complex, sagittal view. Its location in a schematic drawing of adult blastozooid with oral siphon (os), atrial siphon (as) containg the Neural gland complex of *Botryllus schlosseri*. Close view of the neural complex in brightfield and Hoechst (cyan) containig 4 structures.

Sequence Names>

<table>
<tr><td colspan="2">Irx</td></tr>
<tr><td colspan="2">#Bs_IrxB
>Bot_trin2_104970_c0_seq1
MMDGRLVAPRIPAI
PMYGQPFNTSPFPAGYLPYGPGDASAFYPQLTIGENADAWRAGLTQAAAAACYDPALAAA
HYGYGYGSMIGDGGRRKNATRETTSTLKAWLNEHRKNPYPTKGEKIMLAIITKMTLTQVS
TWFANARRRLKKENKMTWSPRNRCGEDDDDDIIEGEDENNNLTDAHRNMKSDRASSRSV
S
PTAQVNPKQENGQRIPSFDLPQQPAMHPFLPRQAITGSFTIQRQSTRPDSAFQRQEDMPA
HRSAPEDQDVDLESVDEDAKQSAMSKRSRSTDLQTHDHYRRRSQSPARAVPKSHQAFQP
T
SISSFSRGHHQKSHNAQSIGPNRTSRRDTISPRSHPYSRNVVVSPASTAASRSSSNSPTK
PKIWSLAECAASNDNDHAKSVVDTPTNIAISKSSPRPGISAVFGNGPRPMQGLPTPHLHP
AMNPEAHATLRNWMDGMMHQKMAVAMAAMSGGQAPPTSFFHGGLLPQHLGMNPAAIAM
AA
SFPRVPESGRFPQNFPGFADFAAAQRAHMAATHAADLHAHPENVPHRKSPLSDAGKQSF
D
SSATRHRIISNEDKRKSPLSTASSPRLEIAEVSAFTALNTRPRSGDSGRASVETNNECLT
QSKESSSPK
#Bs_Nr002
>Bot_trin2_109147_c0_seq1
MSYSTLPAPSDIFGVNSGLRSRVFPMAT
DRRSDDSHSPTSEPRSPHYRSALQPAELPPHLVSLYNSPYAAAAVAAAMAGQGSGGATTS
GSHFGHFGRVPNSFYPVMVGGEAWRRPPYLSQLDMYWPGLRSLEAVVGPRTKAPREATS
I
LKSWLNDHKDNPYPTKSEKMTLSIVTHMTLTQVSTWFANARRRIKKENKMTWTPKTRSGE
RDDNSSNDEDHDRDFYGRQRSPFSVSRDSTDGSLPSFDGCRRPMIHPVSNTIPKCLSSPM
PPPNTAHPLFSGFNGFTAERRMANENIATSAAEKKFPYGRSAANNGNATDQTKVATDAKA
CRNKSHESDDEKKVARAPFIHLSERTSHTEDDCAFQTNANPDIRLSPSQSRENSPRGSEC
TPSPRVKIEHEEDSTRSARAAHHPVTRDHKADAITSWMKDFQHLVATGQEDRARQALAAF
ALAAKAGTYPDLGTNAFEAYRGTHKGDRADRQSAFRNLSLFNNPYNTRIDTGDLQKRRLS
DGNCRRNSRAESDSSDAESKSSIPDFHTRKRSIEISAARPEASGDISHQKTVLRTQFASG
DNSNYKD
#Bs_Nr003
>Bot_trin2_118029_c0_seq1
MLIPPTLNLGNTDAEVNSRERFY
STTGIYSTNLAPTATVGPELSPDTSENQSLYQSAAAEGRMFLNAPKMQSDLPGFHMHAAA
AVVTGHPGSFPLNSPASYAVGAATTPSHGLVPHHYGSPTLDPTQYYPSLAPGYDMKTGDM
WNGGYSQPPHTYLQQYEQGYNPYSYDGRYFPGDINDGGRRKNATRESTNTLKAWLQEHK
K
NPYPTKGEKIMLAIITKMTLTQVSTWFA</td></tr>
<tr><td>Cr_Nr001</td><td>Cr_KH2012:KH.C11.485.v1.A.SL1-1</td></tr>
<tr><td>Bf_Nr001</td><td>ACF10237.1 iroquois A isoform 1 [Branchiostoma floridae]</td></tr>
</table>

Bf_Nr002	ACF10239.1 iroquois B [Branchiostoma floridae]
Bf_Nr003	ACF10241.1 iroquois C [Branchiostoma floridae]
Pm_Nr001	phmamm.CG.MTP2014.S526.g09766.01.p
Pm_Nr002	phmamm.CG.MTP2014.S480.g09228.01.p
Mo_Nr001	Moocul.CG.ELv1_2.S24038.g00958.01.p
Mo_Nr002	Moocul.CG.ELv1_2.S130713.g15314.01.p
Hr_Nr001	Harore.CG.MTP2014.S91.g04367.01.p
Hr_Nr002	Harore.CG.MTP2014.S184.g03498.01.p
Mm_Irx1	NP_034703.2 iroquois-class homeodomain protein IRX-1 [Mus musculus]
Mm_Irx2	AAG10083.1 iroquois-class homeobox protein IRX2 [Mus musculus]
Mm_Irx3	CAA75233.1 Iroquois homeobox protein 3 [Mus musculus]
Mm_Irx4	NP_061373.1 iroquois-class homeodomain protein IRX-4 [Mus musculus]
Mm_Irx5	AAF42871.1 iroquois homeobox protein 5 IRX-5 [Mus musculus]
Mm_Irx6	NP_001288066.1 iroquois-class homeodomain protein IRX-6 isoform 2 [Mus musculus]
Nv_Irx	AGD98928.1 Irx [Nematostella vectensis

Gli
#Bs_GLI >Bot_trin2_109349_c0_seq1 MERYIGTQEHQLTITDSKI KTVNREDLYLEKQKYFAAANCAHPTPLYPQLRRAGSISSGSESAFSPVCNNKCDEFQVSS TTDAACTRSRSPSPADVKGSKNQEGDLYIPEDERTNLEVSLRHHRVSAGTGKTETATATS TTSSEEDEERLRMQKQGNRGRRGAIGATDHRPPSCGTPLGKIAEETNAKKLGLNEYPRAH DDEYKQQKHNNVAFLAPSNPHDVNFTNDHHQSTVEARSGNFYVQPHIPVDSKIRDGRYPF DPRQLSYHDEPGMGNTLPDISLLPISPHSPMQPSLNGAPGLFMGSANNSSLDFYMRSLTS PTMSILSQARGLSPRHSMQGRDGEKLQFYYGNNATGSSNAATSVGSNGAIQSANTVQNG Q TQNALNTMAYYQLAQQQSLLAAVANAGVDRSLGMSLDRNGDMSRLSSPFYIRSRQGQKR A LSISPSHSDISIDLMIRTSPNSLLPFTGSVGNSRNSSIGSAGSYGHLSARGISPQMGFNT RLASPMHTQLLTSRLAAASAGFSPGSFTHMPAPAAPMHPQMTRPQNHFHPQHQITHQDIG HRPNQVPLARMPDGVNKTESQHHYDHGLIKTEPHSPAVNSAMRMTGPSQPRNADETETV Y ETNCGWESCHREFDTQEQLVHHINNDHIHGEKKEFVCRWKDCARDQKPFKAQYMLVVHM R RHTGEKPHKCTFEGCSKAYSRLENLKTHLRSHTGEKPYVCEYPGCSKAFSNASDRAKHQN

RTHSNEKPYVCRIPNCTKRYTDPSSLRKHVKTVHGPDAHVTKRMKAEREREEAKKMEDDR
TVKQEADSVRKDEKDQNNNNQPPGQGCNPISPGSQNGAGDTNGPGNLNPRSNQSTCTS
GS
DMSTTPQNSPHAHHPYNNNDSGVELNAGYGHDHDSDGDIVVDENPLPDSTSGGVGLQSR
R
RTSLRQNIIPRMVNQKMQSLSIGNAGMNPPGIGTDALGEYILPSFPNPKTERGSNHSPYP
HFPQQGSGDGTNEVSSTTLINAQNHARPPSKPSHHSKTLPGGLAPMNRHMTLLAPDRRDS
GTSSGHPSRKPSAVSNASRRSSQNTNQVLVNGSYDPISLGSSRKSSGISVAPGNSGAASS
PMLPSLNPYTLHRLTSKFNEVTGRPPPTPLDRDTYTRSQIGKWLQEDAVARSNAQHTGNT
SYHTSMAPPMAPPPHHHNSSATPNRRRSDAPFNRATRTPLPQEISGNANRRASDPVRKTS
QQQFDQNRPSNVQRYYSWNNMNPLPGIRNYHNADKQIPTGPPGMQGYNNFLSAGNLNVA
R
QRSDSNGSFFSSESGYQSNGSNMALNNAGDYQASPMANVESQPYFNSRHASQQGYHNM
PG
VNENILQENLNADIMSGVNGMYPQNNSHACPPNSVMPGGPQRPFSNHPHIQHNPAITDHY
QRGASSTSYHQILQHPEQTAFPNADPPISHPNYRPNDQYRQMQPRPPASAQGQHHFVHP
A
QTFYPANQNGPYSNSNVVPRPPSSHNSGNSGVNRMRASRMRNPNFRAQPSAGGNPQAG
NV
PVNFVPDAEIPDTYLQQEFIDSIPPLHVSGSFDDSNKNRSSMKMPADSTTGPDGPDNGEP
ELRKPVTEIEDNFDAMSGTTNTDVTLPSMSGLSTRLATPTLQKPYNASLQQQEINEQMAL
QDCWDTNPSSNMALNSMSSMLRTLEEEDKFLQMMPH

#Bs_Gli002
>Bot_trin2_114258_c0_seq3
MTASLLSAIADADMNHDNSAHYVSDKSEVAFLDYN
TDTGDPSRSSADSGINSSYDFQQTYQPSAEHVMQSGYSHIPEHTMASHDNDMCDIDGQGF
VGYHQRNECSSTCTANDSPGQTLFVMSPMMQQQGQCTFVAPSIRVLPETPCPSSQGPTP
V
PPCSAKSPFASVASPRNQLQERFSKLRMEQNQTPNANTKTQYKASRSEQQNEASHIGNIQ
EDSKDAGIDYNTVDPHTATDRNGGNTVSVKSEHADPYPFHQSGMRQHHSTDFTQQDYIPA
PTDKTPSYSRTNSCNSLFSHQSTYVLGVPPAYSAQSNSLNPYRVLPPVTPQQSSDDHFQP
PGSAMSTFSNLSGLLSGYSPGGSDFAGSPRNSGKSRGLTSQARKRTLSVSPFSIDGIEIT
TLIRNSPTSLFLTSRGPSPANGDSLPSYGGTYGHLSARKSISPSAASTFSQRLLLATPVS
PNITQNTKQNDSSGRHYLSVHELHSRTHNPFVASTDDYLSNAYVNVPQHHGHRNQYSRTN
RYGPPETQQTHIVSQSPSMLVVASHEDNIAHENSFYEEQYPTNYNPNMVKTEVNTSGYNE
FPVHSHTVPPSFLPRHQHSQRDYNYQANYPPTSNPPPFPYSPGKMEHTAVQLPTSQPHQ
M
THITIPHNSDASGSYHDEKDAQHTSALLSETGIPICRWVNCNQAFKTLEDLVKHIEKNHI
DQRKGEEFICYWNGCPRKCKPFNARYKLVIHMRVHSGERPNKCSFEGCTKAFSRLENLKI
HMRSHTGERPYVCGKDGCEKAFSNSSDRAKHQRTHQDRKPYACEIPGCNKRYTDPSSLR
K
HIKSHAHMKKRKNEPDLVSCMTIHPLTVRQDAYSTDRHVKMEKSTLPPAPQSNHLTQHLG
AMPQAIHRSAHGNHNSELSRHYDAEQHSNRHMMETNRQEQTLQQLMPPPTGPPLVRRQV
N
ARSGMHPTSAGRSVSSQSPMMVDSRMDLGSRPEYYSLSRASNNRDMIGGFGQAPMEGF
QP

QPPPRPPTSGSYRYERNASITPNRLLHPNYQIGHSHLGGHYQQSYTPRALSARSDMSVDA
VSDHMLPGTPYSMTLDVDRKAYNAVDRSPSQMSLIYADGPR-

#Bs_Zic-r.a
>Bot_trin2_108071_c0_seq2
MKESTEN
DGNPLVIHENQLQQSVYPVCHQAYPTFYECPYPTTANCNSNNNNNSTANQQCFYPYAQFQ
QKVPALDLTQPTQSYAQQQYYDGRYRDNASTMLPYAGLSNYRRFRSEDFVPVTSDEVIYD
ATYSSSPLAGPYQESQASFAIGFHSASHNCNYKFSHNCRVENSHSLADIYPYVLPQTMTP
QVDTGYEEFVCKWRTRVSDCKSSGPERQQTCNAMLWTQLDLVNHINTNHVGGPEQADHS
C
YWENCSRRRKTFKAKYKLVNHMRVHTGEKPFACPFPNCGKMFARSENLKIHKRIHTGERP
FGCPHPGCDRRFANSSDRKKHSHVHTSDKPYTCKVAGCDKTYTHPSSLRKHMRLHEEKG
D
MIPFAGTSSPDSELSDANSSKSGPDTFCQSRDGRRTVTCSQASTRIVSVSPFIGEQTISP
KQEFIHSPYPLCYPADHPIELHKEGVIRQWPQTSASPSDDWYYYQSQKLGGMPTPPSDDR
DAPKDFR

Cr_Gli001	KH2012:KH.C7.334.v1.A.SL1-1
Cr_Gli002	KH2012:KH.C5.53.v1.A.ND1-1
Hr_Gli001	Harore.CG.MTP2014.S26.g04635.01.p
Hr_Gli002	Harore.CG.MTP2014.S76.g07624.01.p
Mo_Gli001	Moocci.CG.ELv1_2.S315764.g08156.01.p
Pm_Gli001	phmamm.CG.MTP2014.S516.g09647.01.p
Mm_Gli001	AAC09169.1 zinc finger transcription factor GLI [Mus musculus]
Bf_Gli001	CAB96572.1 AmphiGli protein, partial [Branchiostoma floridae]
Dm_cuint	NP_001245402.1 cubitus interruptus, isoform B [Drosophila melanogaster]

POU

#Bs_Pou3
>Bot_trin2_137286_c0_seq1
MLSQVPNGELAYGSPLQDGCG
KYSSHSEVGKCRNKTRIHENPHMPNHSPVMSYSTPGLSTYACLDSQPPIRSDTSQEQESF
TKISDEHYRPYPNGYQFSNHCYQFNQSGYHRALPIPSLQEQLYSVHDSPRQRIPSRSPHT
SQVTEDFVIKESPAYTNDANVQSWHSFVHSAREISESSRLNTTSNAGCYPIANSVCASDK
AGCVYSYQNQPGQYPYCYSRNYYPASTNLQNRTWNPRPIDLSLKQDHCDGYGEMDYQPT
H
FTSYSPLRIDERNLLSDERTLNKLPCDDMSLNGWTEEDMRQFSKVFKHRRTKLGYTQSDV
GTSLGELYGSVFSQTTICRFEAQQLSLKNMCKLRPLLSRWLQHKDNKHETLTPDIDQIDS
ENGPGRKRKKRTSIEAEVKAVLEKHFKLKPKPMTQEIVSIAEQLSLEKEVVRIWFCNRRQ
KEKKVNEQVMRSQNPT

<table>
<tr><td colspan="2">#Bs_Pou4
>Bot_trin2_112211_c0_seq4
MPFHSTASMAVSSSSSPLENNSHHHNAIGCSSTSITHADIPPGDL
TAEPRELEAFAERFKQRRIKLGVTQCDVGQALAKLKINGVTSLSQSTICRFESLTLSHNN
MVALKPILTTWLELAEEEYRRKMEQSGLAEKKRKRTSIAAPEKRSLEAYFLVQPRPSSEK
IAAIAEKLDLKKNVVRVWFCNQRQKQKRMKFSAFNGENGGM
#Bs_Pou4b
>Bot_trin2_112211_c0_seq1
MYSSMLAQQHHHHQPGSPAIGSIANTSNGSPAPEQKY
APLHNAADAMRRHCMPSPVAYSGNLFAGFDESFLARAEALAAVDYTAKSQFKPEPPVTSY
YSMGPPPPHHMPPQTHHHHHHHGMSMAAGAVPNPIYGAMPPPPPPPGIHHPSPLHSVPQ
C
VTQGSLNITDSGVDLMEQLSNPSHCSSFSSYAPTSFSTPPGIIHISEAPTYNSQPSMPFH
STASMAVSSSSSPLENNSHHHNAIGCSSTSITHADIPPGDLTAEPRELEAFAERFKQRRI
KLGVTQCDVGQALAKLKINGVTSLSQSTICRFESLTLSHNNMVALKPILTTWLELAEEEY
RRKMEQSGLAEKKRKRTSIAAPEKRSLEAYFLVQPRPSSEKIAAIAEKLDLKKNVVRVWF
CNQRQKQKRMKFSAFNGENGGM</td></tr>
<tr><td>Mo_Pou4</td><td>Moocul.CG.ELv1_2.S117415.g13959.01.p</td></tr>
<tr><td>Pm_Pou4</td><td>phmamm.CG.MTP2014.S453.g08914.01.p</td></tr>
<tr><td>Hr_Pou3</td><td>Harore.CG.MTP2014.S51.g14448.01.p</td></tr>
<tr><td>Od_Pou3</td><td>AAW23073.1 POU3 [Oikopleura dioica]</td></tr>
<tr><td>Ci_Pou41</td><td>NP_001027972.1 POU domain, class 4, transcription factor [Ciona intestinalis]</td></tr>
<tr><td>Od_Pou31</td><td>AAT47873.1 brain-specific homeobox/POU domain protein 3 [Oikopleura dioica]</td></tr>
<tr><td>Sk_Pou</td><td>NP_001161509.1 BRN3 transcription factor [Saccoglossus kowalevskii]</td></tr>
<tr><td>Hs_Brn3</td><td>CAA50589.1 Brn-3b [Homo sapiens]</td></tr>
<tr><td>Ci_Pou6</td><td>BAE06650.1 transcription factor protein [Ciona intestinalis]</td></tr>
<tr><td>Ci_Pou2</td><td>KH2012:KH.C4.85.v1.A.ND1-1</td></tr>
<tr><td>Ci_Pou4</td><td>KH2012:KH.C2.42.v1.A.SL1-1</td></tr>
<tr><td>Cs_Pou2</td><td>Cisavi.CG.ENS81.R41.1314919-1326878.12152.p</td></tr>
<tr><td>Cs_Pou4</td><td>Cisavi.CG.ENS81.R19.4764818-4767613.19272.p</td></tr>
<tr><td>Hs_Pou3</td><td>NP_006227.1 POU domain, class 3, transcription factor 3 [Homo sapiens]</td></tr>
<tr><td>Bf_Pou3</td><td>AAL85498.1 transcription factor AmphiBrn1/2/4 [Branchiostoma floridae]</td></tr>
<tr><td>Bj_pou4</td><td>BAQ21920.1 pou4, partial [Branchiostoma japonicum]</td></tr>
<tr><td>Hc_Pou4</td><td>AAB62538.1 class IV POU protein [Herdmania curvata]</td></tr>
</table>

Hox
#Bs_Nr006 >Bot_trin2_108737_c0_seq8 MTESTDTPRAYLGTTFAPNLTSDQSYRTYEAAAETVVYQPSVMP PAVQAPYYYQYSYQSPSVPGTLGSPVTSYFPHHPEYGTYHHGMEWNHVPVHDQNGDRN MD TVATEEHNAASVPQDSIVLPVVDKPDYLHDHEITRSKDSTYDVRYRDSAVACKLSSLKQE KTSDIDSHEDPDKFSSPERVQVPSQHNPAPVTETTWSAQPGFHDHGLPVRNQHELSPTVT HPTNVALEGSQSEDLMHRELSATNSHARFGGSLELSPRATPQIPAANFGFPHPLPVNSYS YPNYQSNDYIRYGYAPSLQQTYGSYSEGYYETTRADADAKLRMSVQHKTKPVSAQPDSTA QTKPHKKIRTSFTKDQVQQLEDDFLRNNYLTRLRRYELAMKLNLSERQIKVWFQNRRMKW KRQSVVAHHSHGSQDKVSPMSNSRSVESLSADESNPVIYQQSLLTKPDYANGYTTRHDCR PVTYKTTNQSASTMADSCFLRKHEDSTFLPNLQEIGTGVTVISSATHSDTAVTYPDSETT NATLSVAGSRHDMASMVGMSRADMRQDYLSIVKANHARSNAHSTKSHEDATQLLYASEES ERVRNGATSSAIPANVDLNDIKPPNLQAVTENSSQMNIIPGINEIDNQR #Bs_Nr005 >Bot_trin2_99455_c0_seq1 MLSARNK HLGVGISSRLQNELGRNARIDTDVVGMNKNEMDSTLEVPRSNGNMMISPLLMKRFMTRQA ELDSSIRSSSRCSTASPSMRDHGRMSSTPQSSSDNDESLVEVCSTEQYTRSKGSAFRIDA ILSENTESLKQSPPSKQFQETERMHSTHAYAASPLAGTSTIRSRRARRQHGTNVKIEARE TDSVQSSSSCSSTENTADPTNLGSRDCGHDNLGSPERIGSEKESPRHTRDFKAENPECCD GKSVQNRNYVSCGPYPMNPFAFMAMMSMGRENTSKISRDFDRMSSAASESRESIHQSLN G QYSLPGAFSCGPNALFPRHHVAVNPHPSQPFVPIQAVNAFATNPRFFDGIGPHLPALDLM RGGSFLHGFGDYPGHIHPGFFNNKARRPRTAFTSQQLLELESQFKQNKYLTRPRRYEVAR DLCLTETQVKIWFQNRRMKWKRSKAEREERHKESLTSTS-   #Bs_Nr004 >Bot_trin2_54769_c0_seq1 MTYRKRVLCLMVKFYCGRVYREEYRQHICKRDVHHRRKLLLLKMSVNNKDSDYS RYAYRDCEIGDYFSSSQTSASSSLINTPGGNFSDYARQGNEKSTYGSNSWLSNSLLTTEH FGHRPGHGNLVSSNLDLQDRIGSNPCCLPQVETNQEISNKKHTEFAPTPNYNTMRNSSQE HSLQDSEIRNGSNPGSTINSFPWMNMTAGKDTAIYPWMRRIHTKPNQTTDQLKRPRTAYT RFQTLELEKEFHFNQYLTRRRRIEIAQDLSLTERQIKIWFQNRRMKWKKSNTLRPSQRYD ATFDPHL #Bs_Nr003 >Bot_trin2_84483_c0_seq1 MCDHVSLKYSSMPAIESHPYYMGHVKSWPAEKQPNDSEY ASFRPDAGIWNNHPYAQHDKTPNLEYTTRVVPEQEAPFDFSTPYQEFPGGVGRAVTGQRF SKFPQNAHHVPKGEPIAVDSYSFADSFWDSNRERYISPSGYSNGYIPTNSYGSIPVYGQT QPPFKSQHSPPFEYTTGPFRNDANRNRSTETDDTESSCDADAKDKALKFPWMKTTKSHHY EWKAQWQQAMGMTSPFYPEVEENKRTRTAYARWQLLELEKEFHFSRYISRPRRIELAAML

SLTERHIKIWFQNRRMKWKKD

#Bs_Nr002
Bot_trin2_106698_c0_seq2
MAALHEDVAVTSLPQPFGVEMNVATAVSADSC
PNVAETLTPLRAQNQNSSNSNYLQGMQNMLRSHQEDLSGLENLNTSAMPGPTGDFFQTG
D
SEFNGVNDRLVKSELRNGGTNIYGKNGMDGMVSIGDHEGKLGKCMKSKGESPPTVYPWM
K
RIHASHVFNSGAEQGKRPRTAYTRHQVLELEKEFHFNRYLTRRRRIEIAHSLCLTERQVK
IWFQNRRMKWKKDNKLPNTKSRVLQGSGMSGSFDQNISPTLSSPSNDGIPMHSRSPMNS
M
MHGNSHSPEHRTINGYESRGPASRPLPENYFGGYDVTAASQHQPQRRTFVNSAVERMLN
G
SPGFFYQRQSDFGSGKPDGGAVGPFQGKFDPLDSQSPTSMLHRSLMSFTSPPTATVGWP
D
QLTSPGFQLRQPSLPPAYMGQGVPINTLPYHLSTYQ

#Bs_Nr001
Bot_trin2_114311_c0_seq3
MTSVRETGFINSQPSLAECFNLGPKENSLSRSREFPISTSTSS
ASQRAEKFRSVDRSDCSLAKRRSEIEDFTHHPERCVMQKLPSPPIVSCATMSGDTVQEYQ
RWDSFTSHCKINSPDAIPTGVGLMSRMRSQSQYNASNGRSSYVKQDHKPEMQGYPYPHP
H
HRFDVAGEMPIPMDLQPVPRDIMNGLNQIPHDNGYNDRVHGYPHNLDRQMHHIPSDVAPT
LLPQGHCVPDLQRWRDGSSNAMTSMYDRAMVNNFGNYDPHQSPAMIPNYMTTPHPDYP
WMREKKAGKKHQETRRVRPPGTSRRLRTAYTNTQLLELEKEFHFNKYLCRPRRIEIATLLDL
TERQVKVWFQNRRMKHKRQQQNKHDDQKAGDESANSEEEVNSDSDEDDGKSKGGTTS
SDNDSQSTGSSATKPVKSGKKKSKDSKPKSRQSEGNQTLEGVDHDQDLANCSDKENGSD
SGI
HSDCKLDKSVDTKTAKELPAAKQWRVNEDEALLDETDSDRIHTLTRSVSEEKPTDVTSSP
EHAQLQNKMAGQPVDYGSTAHENDTIAESLKDPINFHPLTSSPSLRSALVQPKGPGSGNV
KPTSDNANDKYPFASYHGPSKSHSDASVAIPPQRRYSERNDMARNRGQPQKPPITVDRSF
SVPSVSPYETHRNGMHSPNRYQANDSRNAPPARYPAHDTQQYMPETQVYPNTQGQSPD
LA
FHAGGQSHGFLAFGQRPGVNPAQLRTAGQNPATPAIDGQRSVYSSDSPSSQGNFKPPSG
D
FYPNVRHADRDYLNHHEHRYGHDRWPHAKQDATLEYEARFSAPYSHMNDHYRNYGSSHV
K
AGHDFALGGAMEPFRDADLVNQPYKLQSKPFAKVGIAQTAGLLNSTQKVLDM

#Bs-Hox10
Bot_trin2_111932_c0_seq2
MPGGVYPGESQYQLQVSSGYNQSTDKDSGHGVSNHMSQSAIATMSTNNLP
PAEHSSKISSTPEDLNRHSLPLPDYQNSFGHQLMKDRASMQLPDYQQFGNGQPIDVGMNG
THMFNRVDIEHNPSPVSDKKFSYSTLTTKPKKQNGQYVQQLAGDDPPCYGQRVKAEFLSA

<table>
<tr><td colspan="2">DVFSPRSSASDASSCDRRMMTSSSVDSLSTDKSPAKRDALEYDVDAPYQDKGEKKSEDE
D
PTKHWLTANGRKKRVPYTKYQLLELEKEFHYNQYLSRERRQEVAKAVTLSDRQVKIWFQN
RRMKWKKEKKEEKVRDGISLPPPPHILHPHNHGNPMHQAHFSPLTHYSGTAAAFAAAASM
GHHHHHPSAGGIGTHHHPHFNGSISGHNGPPSHFQHPSYSTPHSAMPGSMAAVAAADFF
S
SFQHHHGYQVPRDPSGGLQLGCMYN

#Bs-Hox3
Bot_trin2_111400_c0_seq1
MSFPAIQPLESGSFELNHDLDSGANIQSAFSGRDFYCRRH
DALRNEISDDEVAFTRRCNEIKFESTGNGNFPNGLSPLYSDVTDTSGELIENNIGPGNGT
KSDKVIDFSCYGMSQSNSFFADDMPSQSTALNSVEKSLHGNLESSIPISLRSAEPDRHFI
EKYLGDLRNVNGKENVVTSRQNKPLSAIANDLKCRDDNDHYVPQHETQTRFERPYLGEQL
SRNPDQTTHSSESRDRVENDEVKGKKNGGETRPRTSEASGLVEQAEFRHPQANFQQRFL
N
HKLRRSEHAAKSNVEQRGSMLIVADAQNGDSAEDSFLACDDLQASCGGKVISAKSSPTAQ
RASEYMSPDSHTDISQRSSERSHYRGEMQSHREAKDNTVEGYSPTRGRLINGRTGVAVND
KSKVSQAANLFPTGTGSVEKHGHERRDSRDSGSSKSTEKDTMVECHQNGDTPEMGEGLD
E
SSPVSKHIYPWMKETRQNAKRKQSTSSNSSNGGDKSSKRERTAYTNAQLVELEKEFHFNR
YLCRPRRIEMAQMLNLTERQIKIWFQNRRMKYKKDQKLKVVPMARDDGFFHQHHHNPNMA
TDPLMSRNNDHRHYVGGGHFHPGDGSYRPLPTEGPGSTPGCSSGSRCYHSHRQHGSHV
SQ
DAGLPNTHMMPHGAAHSFESMSKSYPGPTHKGRLTSPSSVSSNGTLSPGNGQSSNFTPDI
GARSASSSGSEAQNMGQSTRGGYMPSSISPHQTPVEDDAGMLREKTARAQTQALARVAS
R
PSGYDSLFHDNGFPAFQNQIDDLAARKPQGNAAQITAPRAPAYGNFSSVYDRAKELAAAS
NFNTGRYGQFHNPLESLSDCVQRSPGEIQDRARHFAGHGHPPADVFAGSFASGHCGPAR
L
ERRVERTPLDFSRSKSSSPRMQMNDVYMNNRLTHL*

#Bs_Hox1
Bot_trin2_97118_c0_seq1
MSSFMEYPVCSGDPYAVRPPANMMYNPVDGISSFNIKRTISSEDSIDTTS
PMYSCVMHDSSALRKATDENMTQLETNGEQTYTSLSHQLSIYELQQQHYQARDAPGPQTA
TPSPNSSSPPVSESYSHPSPQQTPLSPPTYPQITNTTTYSTNLMGTYANGYAYNPPLKTD
PASSSLYSLNPTSPYPTSPNQPGPPQNSYSSCNPIVYSGNLHATSYPMHAQYPPCPMPIN
GHHSGLDTSIVGYFHSGNTQRRSPHDYGERDKLDNAASTNSDLANTYDWMKIKRNPPKTT
FYNHGKMDPYGYVNQVGNGRTNFSTKQLTELEKEFHYNKYLTRARRHEIADSLSLNETQV
KIWFQNRRMKQKKRDKEAEKILKANATQLQPVKIAPDEFCANDNVNCDKKTIASEKQDSG
TASSRKPDKSDAYASFDTPPKKCQRQSPSGFSPSSSLSNSSSPDRATQRHPTAGNETATR
IYEDSKSNLSELTVPVRSYSNNSLLATTL*</td></tr>
<tr><td>Cr_Nr001</td><td>KH2012:KH.S765.3.v1.A.SL1-1</td></tr>
<tr><td>Cr_Nr002</td><td>KH2012:KH.L171.16.v1.A.SL1-1</td></tr>
</table>

Cr_Nr003	KH2012:KH.C1.1098.v1.B.SL2-1
Cr_Nr004	KH2012:KH.C1.959.v1.A.SL2-1
Cr_Nr005	KH2012:KH.S765.3.v1.A.SL2-1
Cr_Nr006	KH2012:KH.L132.17.v1.A.SL1-1
Cr_Nr007	KH2012:KH.C1.38.v1.A.ND1-1
Cr_Nr008	KH2012:KH.C2.917.v1.A.ND1-1
Cr_Nr009	KH2012:KH.C14.554.v1.A.ND1-1
Cr_Nr010	KH2012:KH.C8.511.v1.A.ND1-1
Cr_Nr011	KH2012:KH.C2.957.v1.A.nonSL1-1
Cr_Nr012	KH2012:KH.C1.922.v1.A.ND1-1
Hr_Nr001	BAD52255.1 homeo domain transcription factor Not [Halocynthia roretzi]
Hr_Nr002	Harore.CG.MTP2014.S26.g09261.01.p
Hr_Nr003	Harore.CG.MTP2014.S11.g04683.01.p
Hr_Nr006	Harore.CG.MTP2014.S26.g15504.01.p
Hr_Nr005	Harore.CG.MTP2014.S327.g04987.01.p
Hr_Nr007	Harore.CG.MTP2014.S1.g00959.01.p
Hr_Nr009	Harore.CG.MTP2014.S151.g05226.01.p
Bf_NOTO_like	XP_002601133.1 notochord homeobox-like protein [Branchiostoma floridae]
Od_Nr001	AAT47829.1 Hox4.1 [Oikopleura dioica]
Pm_Nr001	phmamm.CG.MTP2014.S288.g06741.01.p
Pm_Nr002	phmamm.CG.MTP2014.S288.g06745.01.p
Pm_Nr003	phmamm.CG.MTP2014.S1196.g14047.01.p
Pm_Nr004	phmamm.CG.MTP2014.S4.g00163.01.p
Pm_Nr005	phmamm.CG.MTP2014.S245.g06061.01.p
Pm_Nr006	phmamm.CG.MTP2014.S245.g06060.01.p
Pm_Nr007	phmamm.CG.MTP2014.S288.g06741.01.p
Pm_Nr008	phmamm.CG.MTP2014.S245.g06059.01.p
Pm_Nr009	phmamm.CG.MTP2014.S486.g09313.01.p
Mo_Nr001	Moocci.CG.ELv1_2.S647652.g29801.01.p
Mo_Nr002	Moocci.CG.ELv1_2.S269453.g06393.01.p
Mo_Nr003	Moocci.CG.ELv1_2.S456844.g14957.01.p

Mo_Nr004	Moocci.CG.ELv1_2.S84862.g01155.01.p
Mo_Nr005	Moocci.CG.ELv1_2.S581361.g23190.01.p
Mo_Nr006	Moocci.CG.ELv1_2.S269453.g06390.01.p
Bf_Nr001	ABX39493.1 AmphiHox9 [Branchiostoma floridae]
Bf_Nr002	ABX39492.1 AmphiHox8 [Branchiostoma floridae]
Bf_Nr003	ABX39491.1 AmphiHox7 [Branchiostoma floridae]
Bf_Nr004	ABX39490.1 AmphiHox6 [Branchiostoma floridae]
Bl_Nr005	ACJ74385.1 Hox5 [Branchiostoma lanceolatum]
Bf_Nr006	AAF81909.1 homeodomain-containing protein Hox11, partial [Branchiostoma floridae]
Bf_Nr007	AAF81905.1 homeodomain-containing protein Hox14, partial [Branchiostoma floridae
Bf_Nr008	Bf_H008-AAF81904.1 homeodomain-containing protein Hox13, partial [Branchiostoma floridae]
Bf_Nr009	AAF81903.1 homeodomain-containing protein Hox12, partial [Branchiostoma floridae]
Bf_Nr010	BAA78620.2 AmphiHox1 [Branchiostoma floridae]
Bf_Nr011	BAA78622.1 AmphiHox4 [Branchiostoma floridae]
Bf_Nr012	BAA78621.1 AmphiHox2 [Branchiostoma floridae]
Bf_Nr013	CAA48180.1 Amphihox3 [Branchiostoma floridae]
Bl_Nr014	ACJ74389.1 Hox10 [Branchiostoma lanceolatum]
Bb_MOX	XP_019632267.1 PREDICTED: homeobox protein MOX-1-like [Branchiostoma belcheri]

Pax
#BsPax37 >Bot_trin2_113268 MFSWEIRDQLIKDMLCDRASAPTVSAISRILRAKGCDTENETTEGDNGDAAASVSDTGS NGSVEKAEPGDGAESEYGSEPDLPLKRKQRRSRTTFNATQLDDLERCFERTHYPDIYTRE ELAQRTQLTEARVQVWFSNRRARWRKQMAAQQLPPLHAHHAHPHISHHFGPYHPSIAAAG MSAHNYMLQAAAAAQQHQHSPTSSHHLPHHAHSVHYPPSSGHAHMGPATGAVTGSSFES T FPSPSSHDLSTVHRHALSSPLHHAHPYAQSAMRYDGAHDATFNAAAVGYSLVGASAAAAA AARYQFSSSAFTSENYSPQMPGGHEQSAFNSASQPDSQLPSSEMASTEYGRHSSDSLSG Q RGLNLTSARNDDRSPAASSPNCHQRMNPGNIAHRDSRSASSVGASEKISRTVKNPAGSPS SSNDARSAGINDPSYNQTTSGQIPHHQQTTNLISHRSEHVGGNANQPAVQGSDAQMAVGY

RARAGSDQTYTASAALAHGFPVGHDPNVPAAAHQYASHNPFASCQYSGYGQG
#BsPax258A
>Bot_trin2_112027_c0_seq1
MSCTWGPAMTAGHGTMNHPFARGMTPT
MSSLTSSRSGHGGINQLGGVYVNGRPLPDPVRQTIVDMAHQGVRPCDIARQLRVSHGCVS
KILARYYETGSIKPGIIGGSKPKVATPKVVDKICEYKRQNPTMFAWEIRDRLLLECICDS
ENVPSVSSINRIVRNKAAEKSKHSGLPPSLNIPTTAMSMHQNGCIMDGGDLRTPYSINGI
LGLHGHSQAGPMPLSVRMALPGTVPIPSANTPHHQSQHPQFAMKMAEVACTGNYHPPSP
T
HTSAIARSGTSLVKYPQHVKECDMTAVPTLGMIMPHQAQKRHILYHQDGTYDHGQLSVLS
RTAEEITSSGDDLHPQSQALPLSSRHNAIYGNHSRLVNTPQTISAEYSVAPSIRQSRPSS
LSPQHRYDLISNPSTIRGAPHGIEIKQEITNSTPISSIPNPVPISDSTGLEAGTQMNHYH
QSPQNCNDRSSPPKGRKTKSPPETTLTELKPSPAKHPNHYHDTSPLRPTHHAILPTIALR
TAASTYELSENARLTTTASRMMHDSTSNPQISSSPYTSLNVVRNDHGAYLQENSYGQFYS
VSSNYCENLFQSSHNHHHQQKPDHFAIPVTSSLPQKSHSEGAALSESTRAGCSPVHAVNS
SNDYRIRPQAETRGSYLLDATSSPGSMSMSSQHSPNGSSYESGNLHSYQNATATTHGLPA
GQYGDHGLQCPTYTSYGEHWRNAQIGIVKGVSPHSPEPSLMKPRLVDLHAKSNERSVSVG
MISAQ
#BsPax258A
>Bot_trin2_102851_c0_seq1
MSETSSEVAVSEAANSSETMKRPLENPENTPAAKLLKVSSDGEVIIEMVNDG
ELMELSKKSLQQESEDATSQDVVILQGEVAEEETVMTTDNASDDFSLDMIGDKHTQVQVV
VEQEVSDDNLILQQRTETRPKKQKRSLSNTATVAAPQGVNPAMMSNLLHTAAAVAASQNA
AAATVQVVPGQATGTIQLVPASEVQNTQTLMASGIAPQSLSPQTIGVGEDGKPQMIPQPT
NQPVFTYLVTSQGTLISAHSSDGTPLISPAESEKNTNKAKAEEGDDAGHGSGVNQLGGMY
VNGRPLPEPIRKRIVDLSHQGVRPCDISRQLRVSHGCVSKILARFYETGSIRPGVIGGSK
PKVATPPVVVKITEYKRENPTMFAWEIRDRLLSEGVCTPDSVPSVSSINRIVRSKTSDFM
KDNQHLTNQANMTSPLLSLKDKPQFQAMVQGMQASDNQREPDSMQQYISIEMVPSSAFG
Q
GADNTVQAAVSQTSASVASQVVSAASLSLPPPPVVPSAQGAKNVQALPQQIFANLSPQQQ
FDLLKKVSISETIVTSQKDDANAGANQSVSSEQIVTVQGQNLTPLTTQQQEQVVMQLKDL
NKLLTNVSNTVDQEQAAAAKNADLSSGAQLQITRADGQQSFITVSQQQLQQLTEQGIVLQ
RASPEKTKPDDTNAEESASHDDSEKNARAAGHAQSIAEQEQDILKAIQNSVLVKPEPMDP
AEGQESSAVSQLLLAAQQQQLQQQNKQVGGNGAEDDTRMLTAATLADVQNSASSIALTNS
NNSNSAQQQILSASPNVLASLQQQLAGLQQQQQMMQQHPARAGTAQAQQDQAGIVHILS
A
AAQIQQQAGPNANSNNNNSLQRQLTVAQMLSAAGQSAKTGTELQPVGTLLSQLRPVSIQN
SSTIGAADSVSSSVATTLPPLALPSATTATQQVPQIGASSTNTSQPFNTIYNQAWKLATN
GQRVGGEVTVVSDSASKVACAISERD
#BsPax19
>Bot_trin2_110969_c0_seq1
MIQSPTQSQPVTFSSPHPGPPGMVSIMQTSGTICRTEQTFGEVNQLGGVFVNGRPLPNAIR
LRIVELAQLGIRPCDISRQLRVSHGCVSKILARYNETGSILPGAIGGSKPRVTTPSVVNA
IKDYKVRDPGIFAWEIRDRLLSDCICDKYNVPSVSSISRILRNKIGNVLHPNNPLNPNYP
ASCQENMIKDVSPPPGVHHQGVYSNAASAMYTPYMLQPGSKPHPLPPSVPGPASFRHPA
V
PYPMYCHSVNEILGFPRFDQPHISVPSSTGTLQSIDGHYGNLNDVTGPHDISNNMSTKIE

<table>
<tr><td colspan="2">EWAKVPAYGVTPPAQATSVNINGTSTNGDKNHPSMSPNNHDHQQQPQNYSENTDTTYAS
P
PPSSIPQSTAYPNGHSYMSHQYSASMNGWRHDSVKSGRSFSSSPNTVNHLTMPLAPSHQ
H
QSQMYPPIQTISMPSQSLAPQQMSIYPKYQGSLSHGNESKCVSMTKSPGHPHPTVSAMTA
VNAS
#BsPax6
>Bot_trin2_186131_c0_seq1
MATHQGYNQESVMPPHISSYQVPYSSAVALYGISNKE
SCEATGHSGVNQLGGAFVNGRPLPGDTRQRIVKLAHEGHRPCDISRLLQVSNGCVSKILA
RYVLAFTAGPYDFSTIVCRISVTAATAK</td></tr>
<tr><td>Cr_Nr001</td><td>KH2012:KH.C10.150.v1.A.SL1-1</td></tr>
<tr><td>Cr_Nr002</td><td>KH2012:KH.C9.68.v1.A.ND1-1</td></tr>
<tr><td>Cr_Nr003</td><td>KH2012:KH.C12.116.v1.A.ND4-1</td></tr>
<tr><td>Cr_Nr004</td><td>KH2012:KH.C11.488.v1.A.nonSL5-1</td></tr>
<tr><td>Cr_Nr005</td><td>KH2012:KH.C10.150.v1.A.SL1-1</td></tr>
<tr><td>Hr_Nr001</td><td>Harore.CG.MTP2014.S558.g07589.01.p</td></tr>
<tr><td>Hr_Nr002</td><td>Harore.CG.MTP2014.S142.g11549.01.p</td></tr>
<tr><td>Hr_Nr003</td><td>Harore.CG.MTP2014.S169.g05511.01.p</td></tr>
<tr><td>Hr_Nr004</td><td>Harore.CG.MTP2014.S133.g04597.01.p</td></tr>
<tr><td>Mo_Nr001</td><td>Moocul.CG.ELv1_2.S110138.g11515.01.p</td></tr>
<tr><td>Mo_Nr002</td><td>Moocul.CG.ELv1_2.S118792.g14053.01.p</td></tr>
<tr><td>Mo_Nr003</td><td>Moocul.CG.ELv1_2.S44711.g02365.01.p</td></tr>
<tr><td>Mo_Nr004</td><td>Moocul.CG.ELv1_2.S115052.g13842.01.p</td></tr>
<tr><td>Mo_Nr005</td><td>Moocul.CG.ELv1_2.S95411.g07942.01.p</td></tr>
<tr><td>Pm_Nr001</td><td>phmamm.CG.MTP2014.S25.g01019.01.p</td></tr>
<tr><td>Pm_Nr002</td><td>phmamm.CG.MTP2014.S649.g10911.01.p</td></tr>
<tr><td>Pm_Nr003</td><td>phmamm.CG.MTP2014.S208.g05384.01.p</td></tr>
<tr><td>Pm_Nr004</td><td>phmamm.CG.MTP2014.S491.g09372.01.p</td></tr>
<tr><td>Pm_Nr005</td><td>phmamm.CG.MTP2014.S21.g00870.01.p</td></tr>
<tr><td>MmPax1</td><td>CAB38370.1 Pax1 protein [Mus musculus]</td></tr>
<tr><td>MmPax2</td><td>AAI50485.1 Pax2 protein [Mus musculus]</td></tr>
<tr><td>MmPax3</td><td>AAH48699.1 Pax3 protein [Mus musculus]</td></tr>
<tr><td>MmPax4</td><td>BAA24517.1 Pax4 [Mus musculus]</td></tr>
<tr><td>MmPax5</td><td>EDL02420.1 paired box gene 5 [Mus musculus]</td></tr>
</table>

MmPax6	AAH36957.1 Pax6 protein [Mus musculus]
MmPax7	AAG16663.3 paired box transcription factor PAX7 [Mus musculus]
MmPax8	CAA40725.1 Pax8 [Mus musculus]
MmPax9	NP_035171.1 paired box protein Pax-9 [Mus musculus]
Bf_Pax1	AAA81364.1 AmphiPax-1 [Branchiostoma floridae]
Bf_Pax6	CAA11368.1 Pax6 [Branchiostoma floridae]
Bf_Pax37	EEN66816.1 paired box protein [Branchiostoma floridae]
Bf_Pax258	EEN69749.1 paired box protein [Branchiostoma floridae]
Pd_Pax258	AGC12568.1 Pax258 [Platynereis dumerilii]
Pd_Pax6	CAJ40659.1 Pax6 protein [Platynereis dumerilii]
Hau_258	AFD97529.1 Hau-Pax2/5/8 [Helobdella sp. MS-2000]
Hau_Pax6	ABN09915.2 paired box 6A transcription factor [Helobdella sp. MS-2000]
Hau_37A	ABI17942.1 Pax 3/7A [Helobdella sp. MS-2000]
Hau_37B	ABL74454.1 Pax 3/7B, partial [Helobdella sp. MS-2000]
Cw_PaxA	BAF56224.1 Pax-A [Coeloplana willeyi]

Zic
#Bs_Zic-r.a >Bot_trin2_108071_c0_seq2 MKESTENDGNPLVIHENQLQQSVYPVCHQAYPTFYECPYPTTANCNSNNNNNSTANQQCF YPYAQFQQKV PALDLTQPTQSYAQQQYYDGRYRDNASTMLPYAGLSNYRRFRSEDFVPVTSDEVIYDATY SSSPLAGPYQ ESQASFAIGFHSASHNCNYKFSHNCRVENSHSLADIYPYVLPQTMTPQVDTGYEEFVCKW RTRVSDCKSS GPERQQTCNAMLWTQLDLVNHINTNHVGGPEQADHSCYWENCSRRRKTFKAKYKLVNHM RVHTGEKPFAC PFPNCGKMFARSENLKIHKRIHTGERPFGCPHPGCDRRFANSSDRKKHSHVHTSDKPYTC KVAGCDKTYT HPSSLRKHMRLHEEKGDMIPFAGTSSPDSELSDANSSKSGPDTFCQSRDGRRTVTCSQAS TRIVSVSPFI GEQTISPKQEFIHSPYPLCYPADHPIELHKEGVIRQWPQTSASPSDDWYYYQSQKLGGMPT PPSDDRDAP KDFR* #Bs_ZicL >Bot_trin2_143376_c0_seq1 MSYALGYEPKDDEKEEGENEEAGKQRNPGSVPHQIRHPMHHQQQQVG

FPNGYPQGYHQEPAQVQYAQNWHGYQGPPAVYHQHQNMMPYQMPQRFGHLMQHCQQ
APVRIQPLVCLWVPSKDRENGAPCKHTFGCIRDLVEHISNEHVRGHDHTDHSCYWKTCTR
EKGF
HAKYKLVNHIRIHTGERPFICTNKTCGKRFARSENLKIHLRIHSGKKPFGCPAKGCERFF
ANSSDRKKHSYTHLNTKPYSCKRCGKKYSHPSSLRKHEKSHNSKQQSRNQTTDGCDSSG
Q
SPVSSIGADESLPMTSTPVTRPAETQPISQQPIPADVGYAIPQSNIELSSMSAVSQHNVD
CGIVTNPPMDMPGYIGQQSENLPQFHPQQIFRQTPPPIQSAVQNAHYPHILQQEVQDTDF
GNYPVSNPGQASAHNVQHHQHANMMANSPQGCMEHPYQPIFQNGGNECFGGENPFFDA
EN
QVVEFHQNIPGDAADHYXXXXXXX

Mm_Zic1	NM_009573.3
Mm_Zic2	NM_009574.3
Mm_Zic3	AAH80734.1
Mm_Zic4	NP_033602.2
Mm_Zic5	NP_075363.1
Bf_Zic	CAB96573.1
Cs_Macho	BAB68349.1
Pm_Macho	ACH56519.1
Hr_Macho	BAB19958.1
Mt_Macho	BAE54349.1
Hr_Zic	BAC23063.1
Mt_Zic	BAE54350.1
Pm_ZicL	phmamm.CG.MTP2014.S216.g05538.01.p
Cs_ZicLb	BAB68357.1
Cs_ZicL	BAB68356.1
Mm_Gli	AAC09169.1

Ebf/COE
#Ebf >Bot_trin2_108641_c0_seq2 MAMIVPDFGSDMTSALRDQPMLPITGWMQTNLVEPIHMPQPGILRIHRAEFSIQP PKNLRKSNFFNFSLLFYDNRGLPARATKTEFKNFVEEASSEHCSGTEKANNGVHYEITFI LSNGVSVLQDIYVRLVDAANRERIICYEGQDKNPEMRRVLLTHEIMCSRCCDHKSCGNRN ETPSDPTFDKGQGPYDIKFFLKCNQNCLKNAGNPRDMRRFAVAISTDVDMKFPLAYSETM FVHNNSKHGRRAKTDLENGKRSSGPCITSVCPSEGWVTGGTTVIIVGNNFTRDMQVAFDT

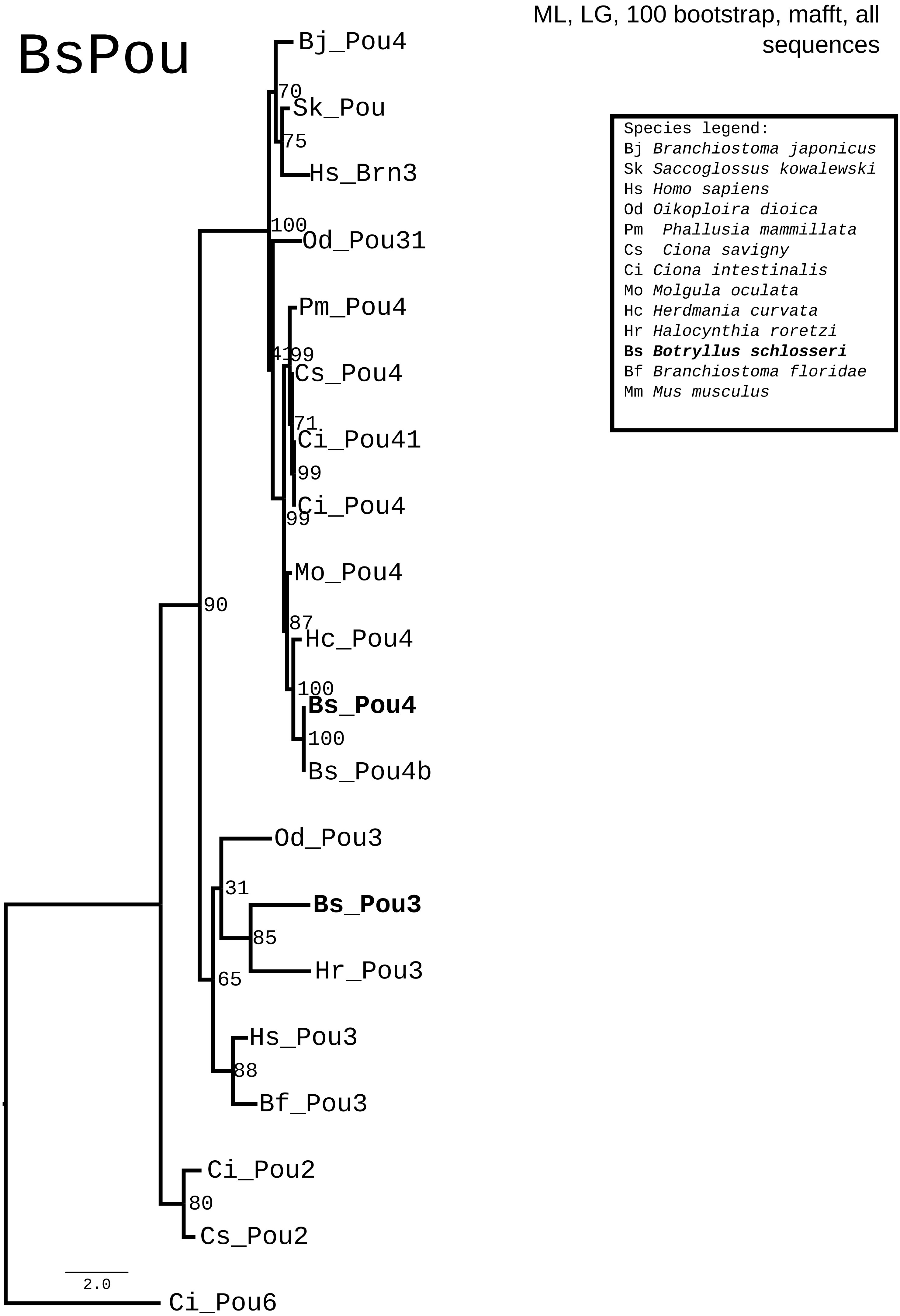
TIVFTEFVSEHALKIQSPGYHLTGPVEVTITIKNKQYCKNQPGRFTYTALNEPSIEYGFQ RLMKMIPRHQADPLQISKEMILTRAAELLEAFSVRPYSQLQSAPPPPLQNSFTSTSSTMM ASAMGGYNVPNQYAPFAAQDRMDSNGYSPHAGPQHSSNISLSSIGAVTNAPPYGSPVNG Y VGPPTFTNMTNSSANMFCNAGLLPQSPNPAMNCLSGSGTTPGIFSFSPANMISAVKQKSA FAPVIRAHNTPSPNHSGVSETTIHEINGYT*	
Ci_COE	NP_001072030.1 transcription factor protein [Ciona intestinalis]
Cs_COE	Cisavi.CG.ENS81.R5.521849-529571.00806.p
Mo_COE	Moocul.CG.ELv1_2.S78092.g05653.01.p
Pm_COE	phmamm.CG.MTP2014.S906.g12706.02.p
Mm_COE1	NP_001277638.1 transcription factor COE1 isoform 1 [Mus musculus]
Mm_COE2	NP_001263316.1 transcription factor COE2 [Mus musculus]
Mm_COE3	NP_001106886.1 transcription factor COE3 isoform 1 [Mus musculus]
Mm_COE4	NP_001103983.1 transcription factor COE4 [Mus musculus]
Bf_COE	CAE45569.1 putative transcription factor Coe [Branchiostoma floridae]
Hr_COE	Harore.CG.MTP2014.S9.g07806.02.p
Af_COE	XP_012344675.1
Nv_COE	AAS48917.1

Etr	
#BsETR >Bot_trin2_111868_c0_seq1 MVMAVHGTTQLLQQGLPGPAGM RGCANPPTSFGLPFLNNVIPGPSSVAHLSQVESPVSMSVNGFFPTYTSNMAGVHTYPQIC HTPSSVSMKEVLPKDEDAIKLFIGQVPKQWTEMDLRPIFEPFGEIYELSVLHDKFTRMHK GCAFLTYCKKSSAIQAQDYLHEKKTLPGMNHPMQVKPADTVNKGEDRKLFVGMLGKKQDE EDVRRIFEKFGTIEECTILRTTDGQSKGCSFVKLSSAAAAKAAIEAIHGKLTMPGASSSI VVKLADTDRERAVRKMNQIANSYGPVALQLSPYGSTIPQMAHHQLIAAAPNVIPTAGWSP MTALAATTPTLAHLTPGSAVMQTGTGPQQITSIPSTPQSPATPITTLSLVPSQVLDQGSP FSSAQEAMYPLQACPAPTPPTIEMVQAPAYPQQPYTVVYMPSGQYAQIPTQHLTQSQLTS LSPAQAAPAAVGNPTAPQKEGPEGCNLFIYHLPAEFQDQDLANAFQRFGTVISAKVFIDR ATSQSKCFGFVSFDNSGSAQAAIQGMNGYQIGPKRLKVQLKRPKEQSRPY	
HrETR1	BAB40781.1 HrETR-1 [Halocynthia roretzi]
MoETR	Moocul.CG.ELv1_2.S81791.g06058.01.p
Cs-ETR1	KH2012:KH.C6.128.v1.A.SL2-1
Pm-ETRa	phmamm.CG.MTP2014.S359.g07762.01.p

Pm-ETRb	phmamm.CG.MTP2014.S359.g07762.02.p
BfHu	Q966S6_BRABE Hu/elav class neuron-specific RNA binding protein OS=Branchiostoma belcheri
DrETR1	BAA95118.1 Etr-1 [Danio rerio]
HsHuB	AAH30692.1 ELAVL2 protein [Homo sapiens]hub
HsHuC	AAA58677.1 huc [Homo sapiens]
HsHuD	AAH36071.1 ELAVL4 protein [Homo sapiens]hud
HsHuR	AAH03376.1 ELAV (embryonic lethal, abnormal vision, Drosophila)-like 1 (Hu antigen R) [Homo sapiens]
DmELAV	AAA28506.1 elav protein [Drosophila melanogaster]
Ce_ETR	A8DYS0_CAEEL ELAV-Type RNA binding-protein family OS=Caenorhabditis elegans GN=etr-1 PE=1 SV=1
Mm_CUGBP 1	NP_001231832.1 CUGBP Elav-like family member 1 isoform 2 [Mus musculus]
Mm_CUGBP 6	NP_780444.2 CUGBP Elav-like family member 6 isoform 1 [Mus musculus]
Mm_CUGBP 4	NP_001139764.1 CUGBP Elav-like family member 4 isoform A [Mus musculus]
Mm_CUGBP 3	NP_001276542.1 CUGBP Elav-like family member 3 isoform 1 [Mus musculus]
Mm_CUGBP 2	NP_001153765.1 CUGBP Elav-like family member 2 isoform 1 [Mus musculus]
Mm_CUGBP 5	NP_795928.2 CUGBP Elav-like family member 5 isoform 1 [Mus musculus]

Notch
#Bs_Notch >Bot_trin2_113352_c1_seq1 MKKFSEDMFSPDDDMFKPPPRVISKYEEDHLHGSVPAEAR KPTPEFEVQKDKRKWTPGHYEAARTPSREATTPLMRDNHQDIDACGPDGITPLMVATTMG GGIEVLEEEEIGEGNEAEGSENMIASLLMQGASLQSQTERTGETPLHLAARYARADAAKR LLDAGADANMKDNTGRTPLHAAVAADALGVFQILIRSRATDLDAKTMDGTTPLMLAARHA VEGMVDDLINAHAEVDTVDNHGKTALHWAAAVNNVEALSSLLRAGAKKDAQTEREETALF LAAREGSYEAVKLLLEFQANRDVTDHMDRLPRDIASEKLHSDIIHLLDDFNIVRSPNSMD NSPLMYNMGLSKAQPAKNKRRPKGDAKAARRTKKKASTDDLEPSNNAPPNMNPIYQSRPI LSDGQSPGSMSSPSTYSDPTHSPPITNPANLPPYYPRGMPQGPSMEKHTHPMHHQSQMS M HKMM

Hr_Notch	BAA25571.1 HrNotch protein [Halocynthia roretzi]
Ci_Notch	NP_001037825.1 Notch precursor [Ciona intestinalis]
Mo_Notch	Moocul.CG.ELv1_2.S43508.g02253.01.p
Nv_Notch	AEW42991.1 notch, partial [Nematostella vectensis]
Bf_Notch	CAC19873.1 putative notch receptor protein [Branchiostoma floridae]
Mm_Notch2	AAC52924.1 Notch-2, partial [Mus musculus]
Hs_Notch2	AAA36377.2 NOTCH 2 [Homo sapiens]


BsPou
ML, LG, 100 bootstrap, mafft, all sequences
Bj_Pou4
70
Sk_Pou
75
Hs_Brn3
100
Od_Pou31
Pm_Pou4
41
99
Cs_Pou4
71
Ci_Pou41
99
Ci_Pou4
99
Mo_Pou4
90
87
Hc_Pou4
100
Bs_Pou4
100
Bs_Pou4b
Od_Pou3
31
Bs_Pou3
85
Hr_Pou3
65
Hs_Pou3
88
Bf_Pou3
Ci_Pou2
80
Cs_Pou2
2.0
Ci_Pou6
Species legend:
Bj *Branchiostoma japonicus*
Sk *Saccoglossus kowalewski*
Hs *Homo sapiens*
Od *Oikoploira dioica*
Pm *Phallusia mammillata*
Cs *Ciona savigny*
Ci *Ciona intestinalis*
Mo *Molgula oculata*
Hc *Herdmania curvata*
Hr *Halocynthia roretzi*
Bs *Botryllus schlosseri*
Bf *Branchiostoma floridae*
Mm *Mus musculus*

Pou

```
Bs_POU3     ML-------- -SQVPNG--- -----ELAY- GSP------- ---------- ---------- ----------
Bs_Pou4     ---------- ---------- ---------- ---------- ---------- ---------- ----------
Bs_Pou4b    -------MYS SMLAQQ---- -------HHH HQPGSPAIGS IANTSNGSPA ---------- ----------
Hc_pou4     M------MFT NMLTHH---- -------HHH HRPGAPMGGI SGSSITGSPS STG------- ----------
MoPou4      QVKSSLIMFT NMLTPH---- -------HSM HRPGSPLCPV SMTSLSGSSQ NSL------- ----------
Pm_Pou4     ---------- ---------- ---------- ---------- ---------- ---------- ----------
Ci_pou41    -------MFT NMLAPH---- -------HH- HRPPSPPTTM TPLNT----- ---------- ----------
Ci_pou4     -------MFT NMLAPH---- -------HH- HRPPSPPTTM TPLNT----- ---------- ----------
Cs_POU4     -------MFT NMLATH---- -------HH- HRPPSPVPTV TTLNT----- ---------- ----------
Sk_pou_NP   MM-------- -MQKDH---- -------VSM HPP------- ---------- ---------- ----------
Hs_Brn3     ---------- ---------- ---------- ---------- ---------- ---------- ----------
Bj_pou4     ---------- ---------- ---------- --P------- ---------- ---------- ----------
Od_pou31    METEQ--EFA TPTTPSS--- -----EIEHK YVPTQPLTPI TSHNINSSII SVA------- ----------
Hs_Pou3     MA-------T AASNPY---- ---------- -LPGNSL--L AAGSIVHSDA AGAGGGGGGG GGGGGGGAGG
Bf_Pou3     MA-------T TTSLPS---- ---------- -SPYIVS--S TAAPIVTTQD DN-------- ----------
Od_Pou3     MGLSL--TGG SAQVPAH--- -----HLQS- ASPPPP---- ---------- ---------- ----------
Ci_pou2     MSKAT----- ---------- ---------- ---------- ---------- ---------- ----------
Cs_pou2     ---------- ---------- ---------- ---------- ---------- ---------- ----------
Hr_Pou3     MLSQL--YGY DQTNPVAFSY FNMKCNTDDQ FDPMQNTTKY GAENDARSPK ---------- ----------
Ci_Pou6     M--------- ---------- ---------- ---------- ---------- ---------- ----------

            71
Bs_POU3     ----LQDGCG KYSSHSEVGK CRNK------ ---TRIHENP ---------- ---------- ----------
Bs_Pou4     ---------- ---------- ---------- ---------- ---------- ---------- ----------
Bs_Pou4b    ---------- --PEQKY--- ---------- --------AP LH-------- ---------- ----------
Hc_pou4     ---------- SPPEQKY--- ---------- --------AP LH-------- ---------- ----------
MoPou4      ---------- SPPEQKY--- ---------- --------AP LH-------- ---------- ----------
Pm_Pou4     ---------- ---------- ---------- ---------- ---------- ---------- ----------
Ci_pou41    ---------- SYGEQKY--- ---------- --------AP LH-------- ---------- ----------
Ci_pou4     ---------- SYGEQKY--- ---------- --------AP LH-------- ---------- ----------
Cs_POU4     ---------- SYGEQKY--- ---------- --------AP LH-------- ---------- ----------
Sk_pou_NP   ---------- IPAAPKY--- ---------- --------SP LH-------- ---------- ------RHD-
Hs_Brn3     ---------- ---------- ---------- ---------- ---------- ---------- ----------
Bj_pou4     ---------- ITAGPKY--- ---------- --------AP LH-------- ---------- ----------
Od_pou31    ---------- ---------- ---------- ---------- ---------- ---------- ----------
Hs_Pou3     GGGGMQPGSA AVTSGAYRG- ---------- --------DP ---------- ---------- ----------
Bf_Pou3     ----MQ--VP SVSQASY--- ---------- ---------- ---------- ---------- ----------
Od_Pou3     ----MENNSQ SNTAPPNCS- ---------- --------TP TS-------- ---------- ----------
Ci_pou2     ---------- ---------- ---------- ---------- ---------- ---------- ----------
Cs_pou2     ---------- ---------- ---------- ---------- ---------- ---------- ----------
Hr_Pou3     ----GQEQEQ QSTSHEQIAS YYNKGLDAGQ TTSAMVETNP VYPNYPYFYP STHVPQAFAT PYFPNTSYNE
Ci_Pou6     ---------- ---------- ---------- --------KP INLD------ ---------- ----------

            141
Bs_POU3     ---------- ---------- ----HMPNHS PV-------- ---------M SYSTPGLSTY ACLDSQPP--
Bs_Pou4     ---------- ---------- ---------- ---------- ---------- ---------- ----------
Bs_Pou4b    ---NAADAMR RHCMPSPV-- ----AYSGNL FA-------- ---------- ---------- -GFDESFL--
Hc_pou4     ---NAADAMR RHCMPSPV-- ----TYSGNL FA-------- ---------- ---------- -GFDESFL--
MoPou4      ---NAADAMR RHCMPSPV-- ----AYSGNL FA-------- ---------- ---------- -GFDESFL--
Pm_Pou4     ---------- ---------- ---------- ---------- ---------- ---------- ----------
Ci_pou41    ---NAADAVR RHCMASPV-- ----AYSGNL FA-------- ---------- ---------- -GFDESFL--
Ci_pou4     ---NAADAVR RHCMASPV-- ----AYSGNL FA-------- ---------- ---------- -GFDESFL--
Cs_POU4     ---NAADAVR RHCMASPV-- ----AYSGNL FA-------- ---------- ---------- -GFDESFL--
Sk_pou_NP   ---HSSEMMR RHCMATT--- ------TSNI FG-------- ---------- ---------- -GLDDSLL--
Hs_Brn3     ---------- ---MNT---- ---------- ---------- ---------- ---------- ----------
Bj_pou4     ---RTSEAMR RHCMTTP--- ------PSNI FA-------- ---------- ---------- -GLDESLL--
Od_pou31    ---ASASAID RKPPITTVAT GHVSAFQSNL FA-------- ---------- --------GV TGFDETFF--
Hs_Pou3     ---SSVKMVQ SDFMQGAMAA ----SNGGHM LS-------- ---------H AHQW-----V TALPHAAAAA
Bf_Pou3     ---RDIQKLP TDYLGQQQA- ----VSNGHP LS-------- ---------Q AHQW------ -VLPHT----
Od_Pou3     -ALQSSSIQP QQRIVSPEAT P-AEVLPGHT PR-------- ---------A AHTDARGITS ENVDESLP--
Ci_pou2     ---------- ---------- ---------- ---------- ---------- ---------E SATKERCAEG
Cs_pou2     ---------- ---------- ---------- ---------- ---------- ---------- ----------
Hr_Pou3     FTHYSDKKIN NEEMQAYRAS ----QGQSNA FSLFPYRSIT SEKELNYGKH NHQQ-----F DNIQAQNSPG
Ci_Pou6     ---------- ---------- ---------- ---------- ---------- ---------- ----------
```

```
           211
Bs_POU3    -------IRS DTSQEQESFT KISDEHYRPY PNGYQFSNHC YQFNQSGYHR ALPIPSLQEQ LYSVHDSPRQ
Bs_Pou4    ---------- ---------- ---------- ---------- ---------- ---------- ----------
Bs_Pou4b   ---------A R-AEALAAVD YTAKSQ---- ---------- ---------- -FKPEPPVTS YYSMGPPPPH
Hc_pou4    ---------A R-AEALAAVD YTGKNQ---- ---------- ---------- -FKTDSTVSS YYPMAAPPPH
MoPou4     ---------A R-AEALAAVD YTNKGQ---- ---------- ---------- -FKPESSVST YYPMSAPAPH
Pm_Pou4    ---------- ---------- ---------- ---------- ---------- ---------- ----------
Ci_pou41   ---------A RQAEALAAVE YPGKSQ---- ---------- ---------- -FKPEPTNHS YYP---PAPH
Ci_pou4    ---------A RQAEALAAVE YPGKSQ---- ---------- ---------- -FKPEPTNHS YYP---PAPH
Cs_POU4    ---------A RQAEALAAVD YPGKNQ---- ---------- ---------- -FKPEPSNHS YYP---PAP-
Sk_pou_NP  ---------A R-AEALAAVD AVSVAKG--A P--------- --------GL NFKNDTGYHH HMN---GSNH
Hs_Brn3    ---------- --IPCTSAAS SSSVPIS--H P--------- ---------- SALAGTHHHH HHH---HHHH
Bj_pou4    ---------A R-AEALAAID VVSGSGVKHS P--------- --------QL -LRPDTG--- -YPGMNGTAH
Od_pou31   -----SATGA K-QEYTSALS TPMEVK---- P--------- --------SL SFKPETS--T YYPMNSSMVH
Hs_Pou3    AAAAAAAVEA SSPWSGSAVG MAGSPQQPPQ PPPPPP---- -------QGP DVKGGAGRDD LHA--GTALH
Bf_Pou3    ---------E GSPWPSSLAA AAGIGVF--- ---------- -------ITA DIK--QGREE LHG---AILH
Od_Pou3    -------ASH NNNEAGSSAL GASLNNHPLF SNWAQ--NHG GQENHCG--- -----SPRTP ILSMSSALPQ
Ci_pou2    AINLSKNSRS SSVDSNNDV- ---------- ---------- ---------- ---------- ----------
Cs_pou2    ---------- ---------- ---------- ---------- ---------- ---------- ----------
Hr_Pou3    SQTSENSNYS SSGNTSSPVP VEKSPCSTTY L--------- ---------- -----ASQYD LYS-------
Ci_Pou6    ---------- --TQSSSDVR YLNLSNI--- ---------- ---------- ---------- ----------

           281
Bs_POU3    RI-PSRS--- --PHTSQVTE DFVIKESPAY TNDA-NVQSW HSFVHSAREI ----SESSRL NTTSNAG---
Bs_Pou4    ---------- ---------- ---------- ---------- ---------- ---------- ----------
Bs_Pou4b   HMPPQTH--H HHHHHGMSMA -AGAVPNPIY GAMP----PP PPPPGIHHPS ----PLHSVP QCVT------
Hc_pou4    HMAPQSH--- ---HASVPMA PSPVPPPPVY GTM------- PPPPGLHHLQ ----PYMGYH SAYT------
MoPou4     HMAPHQS--- --PNSSSSTM GSSMVHPPMY STM------- -PPPAIHHPS ----AIHPMT QSLG------
Pm_Pou4    ---------- ---------- ----MPPPVY SSM------- HPPSAFHHTA ----AL---- ----------
Ci_pou41   YVTQQHT--- ---------- -TSAVPPPVY SSM------- HPPPAFHHSA ----AL---- ----------
Ci_pou4    YVTQQHT--- ---------- -TSAVPPPVY SSM------- HPPPAFHHSA ----AL---- ----------
Cs_POU4    YATQQHT--- ---------- -SSAVPPPVY SSM------- HPPSAFHHSA ----AL---- ----------
Sk_pou_NP  H--PHLS--- ---------- ------APTS RSI------- HP-------- ---------- ----------
Hs_Brn3    Q--PHQA--- ---------- ----LEGELL EHL------- SP-------- ---------- ----------
Bj_pou4    S--P-AT--- ---------- -SVAGPGPHV S--------- HPLP--HH-- ---------- ----------
Od_pou31   S--HHLS--- ---------- -SPSVVVPQL SSY------- ---------- ---------- ----------
Hs_Pou3    HRGPPHL--- ---------- ---------- -------GPP PPPPHQGHPG ----GWGAAA AAAA------
Bf_Pou3    HR-PQQQ--- ---------- ---------- -------QQQ QAQQQQQQQQ ----QQQQQQ HPQI------
Od_Pou3    HS-PTSS--- --HSTATGDQ GSPTGYSAAF LSAKTEPTVA HSLSSLYPQV ----ATSSGH SSLAPFGLDA
Ci_pou2    ---PEDLTGN RL-------- ---------- ---------- ---------- ----KKTNGL HEKSNSN---
Cs_pou2    ---------- ---------- ---------- ---------- ---------- ---------- ----------
Hr_Pou3    ---AQQS--- ---------- -VVTQSQPFY ERPA-AMHSY HAYHDVKQSQ YPTESEYHTH NVVSSSGYP-
Ci_Pou6    ---PKSY--- ---------- ---------- ---------- ---------- ---------- ----------

           351
Bs_POU3    --CYPIANSV CASDKAGCVY SYQNQPGQYP YCY--SRNYY PASTNLQNRT WNPRPIDL-- ----------
Bs_Pou4    ---------- ---------- ---------- ---------- ---------- ---------- ----------
Bs_Pou4b   ----QGSLNI TDS------G VDLMEQLSNP SHC--SSFSS YAPTSFS--- -TPPGIIH-- ----------
Hc_pou4    ----QGPLSI TES------G AEFNGSNLVT HPTV-LLFHR INPRPFT--- -APSGMMP-- ----------
MoPou4     ----QTSLSL SEN------N NDIMEQLGNT -----SSFSS YQPPTLN--- -SSSGMGS-- ----------
Pm_Pou4    ----QSTLE- ---------T NDVIDQLGSS H----GAFPT YQHPTFN--- -TSSVMMS-- ----------
Ci_pou41   ----QTSLD- ---------S SDVIDQLGSS H----CGLPT YQPPTFN--- --SSVMMT-- ----------
Ci_pou4    ----QTSLD- ---------S SDVIDQLGSS H----CGLPT YQPPTFN--- --SSVMMT-- ----------
Cs_POU4    ----PGTLD- ---------T NDVIEQLGSS H----CGLPT YQPPSFN--- --SSVMMT-- ----------
Sk_pou_NP  ------GLD- ---------H GELLGD---- --PMSMAISM SNPMPFTSIG DSITTMAQ-- ----------
Hs_Brn3    ------GLA- ---------L GAMAGPDGAV -------VST PAHAP----- -HMATMNP-- ----------
Bj_pou4    ------PHSF ADS------E ASPFDQIPTT ---------- SLPLSMSMTD PNLTVSS--- ----------
Od_pou31   ------AVSL GAE------G TESLIADGQA YPTIGYGMGL NEAMYISAPS HSQNGLMN-- ----------
Hs_Pou3    ----AAAAAA AAA------H LPSMAGGQQP PP---QSLLY SQPGG----- FTVNGMLSAP PGPGGGGGGA
Bf_Pou3    ----VWNTPV TSA------H IQMTSNGSSP LL---SSHAH SLSG------ YSVNGMVS-- PG--------
Od_Pou3    TRSHPTASSL NAA------Q SQVTWPSSTP GWP------S SAASGIGGDS WYASQMSA-- ----------
Ci_pou2    --LIFVPQSI VEQHN----- ---------- ---------- ------SGQN LSAHTTAT-- ----------
Cs_pou2    ---------- ---------- ---------- ---------- ---------- ---------- ----------
Hr_Pou3    --CPRVATTI SDPYHG---E QRFSNEEKQP HH---RSLEY NQIENIENTE FRACTQVS-- ----------
Ci_Pou6    ---------- ---------- ---------- ---------- ---------E DVVNTVLE-- ----------
```

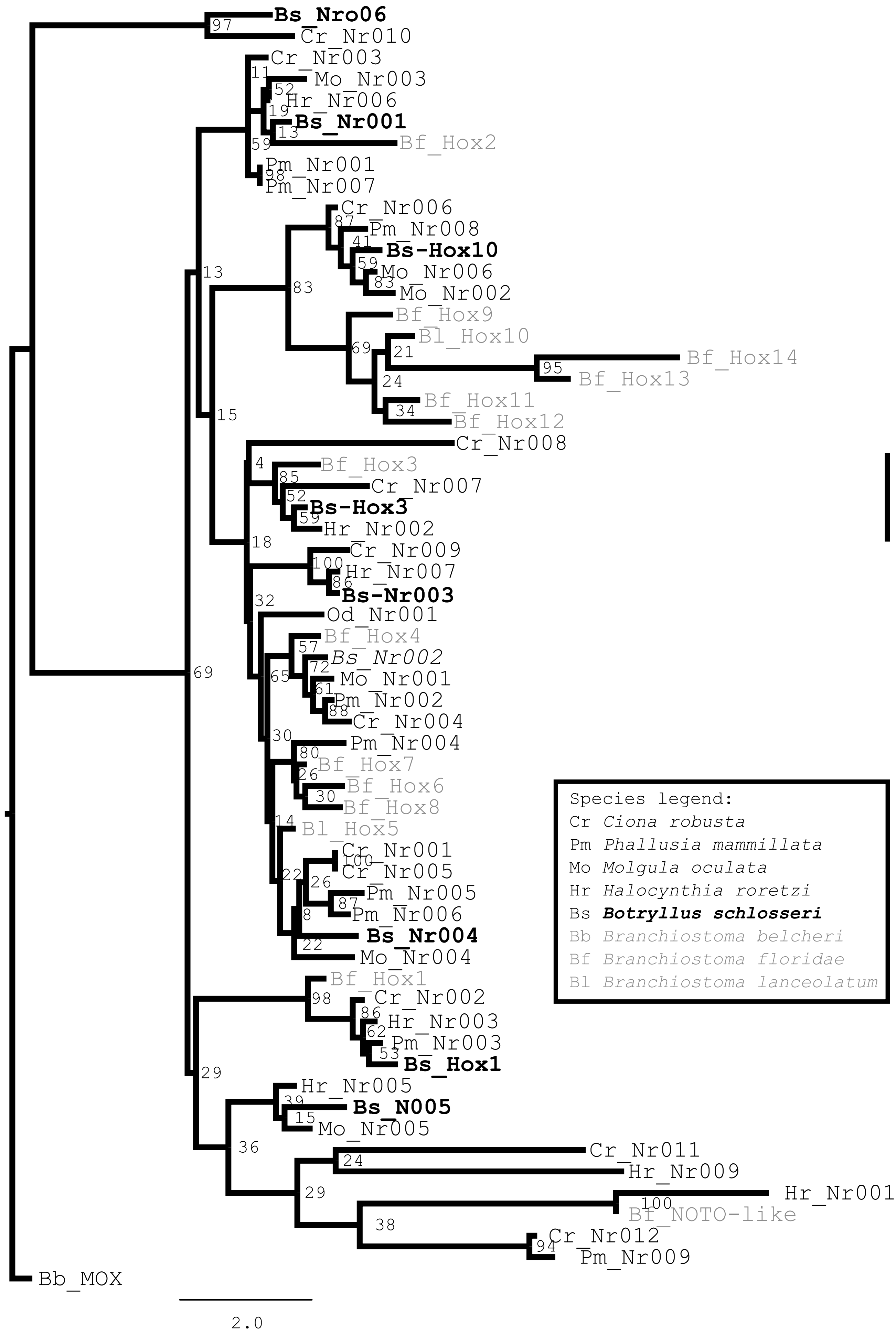
```
           421
Bs_POU3    -----SLKQD HC-------- --------DGY GEMDYQPTHF TSYSPLRIDE RNL------- ----------
Bs_Pou4    ---------- ---------- ---------- ---------- --MPFHST-- ASMAVSSSSS PLE--NNSHH
Bs_Pou4b   -ISE----AP TY-------- -------NSQ P--------- -SMPFHST-- ASMAVSSSSS PLE--NNSHH
Hc_pou4    -TSEVSTSVP SF-------- -------NTQ PSS------- -SMSFHPG-- AAMATPASTA SSD--TSMQH
MoPou4     DISSFAASQP HY-------- ------MSTL PAPHTGFNNS APGSFPAP-- SSMTTTSTA- ------SSES
Pm_Pou4    --SADPVSQP HY-------- -------ISH TAP----PYN PTSHFHPV-S SSIHPSIPPP SSTTATTTQ-
Ci_pou41   ---SDVTSQP HY-------- -------ITH PAP----TYN PPNHFHPT-S SAMHSTVSVT PISAPIITQA
Ci_pou4    ---SDVTSQP HY-------- -------ITH PAP----TYN PPNHFHPT-S SAMHSTVSVT PISAPIITQA
Cs_POU4    ---SDMSSQP HY-------- -------ISH PAP----TYN PPNHFHPT-- SVMHSTSAAT PISAPISTQA
Sk_pou_NP  HHQSHVSMSG HH-------- ------HHVM AAP------- ---------H SMM------- ----------
Hs_Brn3    MHQAALSMA- ---------- -------HAH GLP------- ---------- SHM------- ----------
Bj_pou4    ---------- ---------- -------VSS MAP----THI PGMHPTPT-- GQM------- ---------T
Od_pou31   SGLAQNSMMA HYSGLGT--- ------NQIH GGN------- ---------S THM------- ----------
Hs_Pou3    GGGAQSLVHP GLVRGDTPEL AEHHHHHHHH AHP------- -----HPPHP HHAQ------ -----GPPHH
Bf_Pou3    -----QQMHP VL-------- -------QDH VPQ------- -----DSPIP QH-------- ----------
Od_Pou3    -----AGVAA HP-------- ------YSHY TNP-YTGSAL GAGLHDSFAA QNL------- ----------
Ci_pou2    -----QPIPP YL-------- ------HHQR VNS------- ---------- ---------- --------ET
Cs_pou2    ---------- ---------- ---------- ---------- ---------- ---------- ----------
Hr_Pou3    ---------- ---------- -------DDH PIP------- -----DKVWP HQN------- ----------
Ci_Pou6    ERQSLSVVEV DK-------- ------VKGY VVN------- ---------- ---------- ----------

           491
Bs_POU3    ---------- ---------- ----LSDERT LNKLPCDDMS LN-------- ---------- -GWTEEDMRQ
Bs_Pou4    HNAI------ ---------- --GCSSTSIT HADIPPGDLT A--------- ---------- ---EPRELEA
Bs_Pou4b   HNAI------ ---------- --GCSSTSIT HADIPPGDLT A--------- ---------- ---EPRELEA
Hc_pou4    HT-------- ---------- ---ACSPPIT HPSLPVGDYE T--------- ---------- ---DPRELEA
MoPou4     GQTP------ ---------- --GNCSPPII PHTLPTSEYD T--------- ---------- ---DPRELEA
Pm_Pou4    ---------- ---------- ---PASPNTI VANFPASEYD A--------- ---------- ---DPRELEA
Ci_pou41   GKYGSGCTTP RDQWDDDVTN VFVTGSPPPS VPSFTSNEYD A--------- ---------- ---DPRELEA
Ci_pou4    ---------- ---------- ----GSPPPS VPSFTSNEYD A--------- ---------- ---DPRELEA
Cs_POU4    GKTCREPTTG R--YKPHIIT SYVPGSPQPS VPSFSSNDYD A--------- ---------- ---DPRELEA
Sk_pou_NP  ---------- ---------- ---------- --THGMTDLD T--------- ---------- ---DPRELEA
Hs_Brn3    ---------- ---------- --G------- ----CMSDVD A--------- ---------- ---DPRDLEA
Bj_pou4    GMSP------ ---------- --HHLVPNPH HPMSMHDPND T--------- ---------- ---DPRELEA
Od_pou31   ---------- ---------- ---------- --PSASAEYD T--------- ---------- ---DPRELEA
Hs_Pou3    GGGG------ ---------- --GGAGPGLN SHDPHSDEDT P--------- ---------- ---TSDDLEQ
Bf_Pou3    ---------- ---------- ---------- --HDNSDDDT P--------- ---------- ---TSDDLEQ
Od_Pou3    ---------- ---------- ---------- --FSLGDDQQ P--------- ---------- ---STDDLER
Ci_pou2    GKRS------ ---------- --YPEDDVLS IDNESEDQYQ PSPDDINDLD MHRGHGYDLE EAKQLEQLEL
Cs_pou2    ---------- ---------- ---------- ---------- ---------- ---------- ----------
Hr_Pou3    ---------- ---------- ---------- --NSQQIKHE E--------- ---------- ---TMDEMKI
Ci_Pou6    ---------- ---------- --------IE NENQDVCDLT L--------- ---------- ---TAHEVGK

           561
Bs_POU3    FSKVFKHRRT KLGYTQSDVG TSLGELY--- GSVFSQTTIC RFEAQQLSLK NMCKLRPLLS RWLQHKDNKH
Bs_Pou4    FAERFKQRRI KLGVTQCDVG QALAKLKING VTSLSQSTIC RFESLTLSHN NMVALKPILT TWLELAEEEY
Bs_Pou4b   FAERFKQRRI KLGVTQCDVG QALAKLKING VTSLSQSTIC RFESLTLSHN NMVALKPILT TWLELAEEEY
Hc_pou4    FAERFKQRRI KLGVTQCNVG QALAKLKLPG VGSLSQSTIC RFESLTLSHN NMVALKPILM TWLEQAEAEY
MoPou4     FAERFKQRRI KLGVTQCDVG QALANLKLPG VGSLSQSTIC RFESLTLSHN NMVALKPILM TWLEKAEEEY
Pm_Pou4    FAERFKQRRI KLGVTQNDVG RALANLKLPG VGSLSQSTIC RFESLTLSHN NMVALKPILI AWLEKAEEEH
Ci_pou41   FAERFKQRRI KLGVTQNDVG QALANLKLPG VGSLSQSTIC RFESLTLSHN NMVALKPILI TWLEKAEEEH
Ci_pou4    FAERFKQRRI KLGVTQNDVG QALANLKLPG VGSLSQSTIC RFESLTLSHN NMVALKPILI TWLEKAEEEH
Cs_POU4    FAERFKQRRI KLGVTQNDVG QALANLKLPG VGSLSQSTIC RFESLTLSHN NMVALKPILI TWLEKAEEEH
Sk_pou_NP  FAEHFKQRRI KLGVTQCDVG SALGNLKLPG VGSLSQSTIC RFESLTLSHN NMIALKPILS AWLEEAEKAA
Hs_Brn3    FAERFKQRRI KLGVTQADVG SALANLKIPG VGSLSQSTIC RFESLTLSHN NMIALKPILQ AWLEEAEKSH
Bj_pou4    FAEHFKQRRI KLGVTQADVG QALGKLKLPG VGSLSQSTIC RFESLTLSHN NMVALKPVLE AWLGEAEQQA
Od_pou31   FAERFKQRRI KLGVTQADVG QALGKLKIPG VGSLSQSTIC RFESLTLSHN NMVALKPILQ AWLERAENAA
Hs_Pou3    FAKQFKQRRI KLGFTQADVG LALGTLY--- GNVFSQTTIC RFEALQLSFK NMCKLKPLLN KWLEEADSST
Bf_Pou3    FAKQFKQRRI KLGFTQADVG LALGTLY--- GNVFSQTTIC RFEALQLSFK NMCKLKPLLQ KWLEEADSSS
Od_Pou3    FARDFKQKRI KLGYTQADVG LALGTLY--- GNVFSQTTIC RFEALQLSFK NMCKLKPLLS RWLDEADSAT
Ci_pou2    FAKDFKQKRI KMGFTQGDVG QAMGCLY--- GNDFSQTTIS RFEALNLSFK NMIKLKPMLE RWLLDANNVL
Cs_pou2    ---------- ---------- ---------- ---------- ---------- --------LE RWLMDANNVL
Hr_Pou3    FSRLFKQRRV KLGFTQSDVG IALGSLY--- GYVFSQTTIC RFEAMQLSFK NMCKLKPLLT KWLQDTDIDN
Ci_Pou6    FATEFRTKRI WLGISIETFH SSFDSWV--- GEEV-KFNIR DFETLDLTME ESIYLLPDLK EWLEQVEKAK
```

```
           631
Bs_POU3    ETLTPDI--- ----DQIDSE NGPGRKRKKR TSIEAEVKAV LEKHFKLKPK PMTQEIVSIA EQLSLEKEVV
Bs_Pou4    RRKMEQ---- -----SGL-- ---AEKKRKR TSIAAPEKRS LEAYFLVQPR PSSEKIAAIA EKLDLKKNVV
Bs_Pou4b   RRKMEQ---- -----SGL-- ---AEKKRKR TSIAAPEKRS LEAYFLVQPR PSSEKIAAIA EKLDLKKNVV
Hc_pou4    RRKMED---- -----SGL-- ---AEKKRKR TSIAAPEKRS LEAYFSLQPR PSSEKIAAIA DKLDLKKNVV
MoPou4     RRKMEE---- -----QGI-- ---NEKKRKR TSIAAPEKRS LEAYFTVQPR PSSEKIAAIA EKLDLKKNVV
Pm_Pou4    RRKMES---- -----AGL-- ---GDKKRKR TSIAAPEKRS LEAYFAVQPR PSSEKIAAIA EKLDLKKNVV
Ci_pou41   RRKMEN---- -----AAM-- ---GEKKRKR TSIAAPEKRS LEAYFAVQPR PSSEKIAAIA EKLDLKKNVV
Ci_pou4    RRKMEN---- -----AAM-- ---GEKKRKR TSIAAPEKRS LEAYFAVQPR PSSEKIAAIA EKLDLKKNVV
Cs_POU4    RRKMEN---- -----AAM-- ---GEKKRKR TSIAAPEKRS LEAYFAVQPR PSSEKIAAIA EKLDLKKNVV
Sk_pou_NP  KSKDKN---- -----PDIFS --NGEKKRKR TSIAAPEKRS LEAYFAVQPR PSSEKIAAIA EKLDLKKNVV
Hs_Brn3    REKLTK---- -----PELFN --GAEKKRKR TSIAAPEKRS LEAYFAIQPR PSSEKIAAIA EKLDLKKNVV
Bj_pou4    RAKQRS---- -----PDIFN --NAEKKRKR TSIAAPEKRS LEAYFAVQPR PSSEKIAAIA EKLDLKKNVV
Od_pou31   KAKKEA---- -----DGY-- -TGNEKKRKR TSIAAPEKRS LEAYFTVQPR PSSEKIAAIA EKLDLKKNVV
Hs_Pou3    GSPTSI---- -----DKIAA --QGRKRKKR TSIEVSVKGA LESHFLKCPK PSAQEITNLA DSLQLEKEVV
Bf_Pou3    GSPSSI---- -----DKIAA --QGRKRKKR TSIEVTVKGA LESHFLKQPK PSAQEIAQLA DSLQLEKEVV
Od_Pou3    GGMPTIG--- ---GDKLNQQ -FQGRKRKKR TSIELSTKNA LEMEFIKQPK PSAAEITNTA ESLQLEKEVV
Ci_pou2    RESKDV---V IPTPIQLDPI -SISRKRKKR TSIDAKKRTT LDKQFERNPK PSSEDLQAIA EECKMEKEVV
Cs_pou2    RESKDGHSVV IPPQTNLDPH -TMSRKRKKR TSIDAKKRTT LDRRFIMNPK PSSEELQSIA EECDMEKEVV
Hr_Pou3    GTLNCAI--- ----EDTDH- -VATRKRKKR TSIEQVVKSA LEKQFQIKQK PTTQEICKIS TNLGLDKEVI
Ci_Pou6    AEGDVAL--- ----AKFLG- --KDLKAKQR KCLNEKTKAK LMKLYEDTPN PTKKQMQELS KEIMIAEFSI

           701
Bs_POU3    RIWFCNRRQK EKKVNEQVMR SQNPT----- ---------- ---------- ---------- ----------
Bs_Pou4    RVWFCNQRQK QKRMKFSAFN GENGG----- ---------- ---------- ---------- ----------
Bs_Pou4b   RVWFCNQRQK QKRMKFSAFN GENGG----- ---------- ---------- ---------- ----------
Hc_pou4    RVWFCNQRQK QKRMKFSAFN GENII----- ---------- ---------- ---------- ----------
MoPou4     RVWFCNQRQK QKRMKFSAFN GENVV----- ---------- ---------- ---------- ----------
Pm_Pou4    RVWFCNQRQK QKRMKFSAFN AEAML----- ---------- ---------- ---------- ----------
Ci_pou41   RVWFCNQRQK QKRMKFSAFN GDVIM----- ---------- ---------- ---------- ----------
Ci_pou4    RVWFCNQRQK QKRMKFSAFN GDVIM----- ---------- ---------- ---------- ----------
Cs_POU4    RVWFCNQRQK QKRMKFSAFN GDVIM----- ---------- ---------- ---------- ----------
Sk_pou_NP  RVWFCNQRQK QKRMKFSAQG ---------- ---------- ---------- ---------- ----------
Hs_Brn3    RVWFCNQRQK QKRMKYSAGI ---------- ---------- ---------- ---------- ----------
Bj_pou4    RVWFCNQRQK QKRMKFSAQQ GGR------- ---------- ---------- ---------- ----------
Od_pou31   RVWFCNQRQK QKRMKFSAYN QTAQS----- ---------- ---------- ---------- ----------
Hs_Pou3    RVWFCNRRQK EKRMTPPGIQ QQ-------- ---------- ---------- ---------- ----------
Bf_Pou3    RVWFCNRRQK EKRMTPPALN GPPGPPGS-- ---------- ---------- ---------- ----------
Od_Pou3    RVWFCNRRQK AKRLLGPERK KPDSTGENL- ---------- ---------- ---------- ----------
Ci_pou2    RVWFCNRRQK QKRISQHEQQ EMSPP-DSP- ---------- ---------- ---------- ----------
Cs_pou2    RVWFCNRRQK EKRISMHEQQ QELSSPDSP- ---------- ---------- ---------- ----------
Hr_Pou3    RIWFCNRRQK EKRHAIPDRK AIAGP----- ---------- ---------- ---------- ----------
Ci_Pou6    EQWFHKKRRF DKRENRIKAR LEAKKKNEER LNKEKNAYWD EKLKNLMSDY YKEKEFNKKL FERLKSEGKR

           771
Bs_POU3    ---------- ---------- ---------- ---------- ---------- ---------- ----------
Bs_Pou4    ---------- ---------- ---------- ---------- ---------- ---------- ----------
Bs_Pou4b   ---------- ---------- ---------- ---------- ---------- ---------- ----------
Hc_pou4    ---------- ---------- ---------- ---------- ---------- ---------- ----------
MoPou4     ---------- ---------- ---------- ---------- ---------- ---------- ----------
Pm_Pou4    ---------- ---------- ---------- ---------- ---------- ---------- ----------
Ci_pou41   ---------- ---------- ---------- ---------- ---------- ---------- ----------
Ci_pou4    ---------- ---------- ---------- ---------- ---------- ---------- ----------
Cs_POU4    ---------- ---------- ---------- ---------- ---------- ---------- ----------
Sk_pou_NP  ---------- ---------- ---------- ---------- ---------- ---------- ----------
Hs_Brn3    ---------- ---------- ---------- ---------- ---------- ---------- ----------
Bj_pou4    ---------- ---------- ---------- ---------- ---------- ---------- ----------
Od_pou31   ---------- ---------- ---------- ---------- ---------- ---------- ----------
Hs_Pou3    ---------- ---------- ------IPDD VYS------- ---------- --------QV GIVSADT---
Bf_Pou3    ---------- ---------- ------VPTT VSS------- ---------- --------EV DRASALR---
Od_Pou3    ---------- ---------- --------SD SYSI------ ---------- ---------- ----------
Ci_pou2    ---------- ---------- --RNMVVPTT SYSVPWQFPS NE-------- -------DEP STSNNIRR-Q
Cs_pou2    ---------- ---------- --EHISAPTT SYSVPWKFTE AE-------- -------CEP STSTVLRRPT
Hr_Pou3    ---------- ---------- ---------- ---------- ---------- ---------- ----------
Ci_Pou6    TSKGRKRKLK QTTKETEKKN LMEDKVASAT SEDVPPQRNL RERKNRKKYI EDESFNNDEV SVSEMQENIQ
```

```
          841
Bs_POU3   ---------- ---------- ---------- ---------- ---------- ---------- ----------
Bs_Pou4   ---------- ---------- ---------- ---------- ---------- ---------- ----------
Bs_Pou4b  ---------- ---------- ---------- ---------- ---------- ---------- ----------
Hc_pou4   ---------- ---------- ---------- ---------- ---------- ---------- ----------
MoPou4    ---------- ---------- ---------- ---------- ---------- ---------- ----------
Pm_Pou4   ---------- ---------- ---------- ---------- ---------- ---------- ----------
Ci_pou41  ---------- ---------- ---------- ---------- ---------- ---------- ----------
Ci_pou4   ---------- ---------- ---------- ---------- ---------- ---------- ----------
Cs_POU4   ---------- ---------- ---------- ---------- ---------- ---------- ----------
Sk_pou_NP ---------- ---------- ---------- ---------- ---------- ---------- ----------
Hs_Brn3   ---------- ---------- ---------- ---------- ---------- ---------- ----------
Bj_pou4   ---------- ---------- ---------- ---------- ---------- ---------- ----------
Od_pou31  ---------- ---------- ---------- ---------- ---------- ---------- ----------
Hs_Pou3   ---------- ---------- ---------- --------P- ---------- ---------- ----------
Bf_Pou3   ---------- ---------- ---------- --------P- ---------- ---------- ----------
Od_Pou3   ---------- ---------- ---------- ---------- ---------- ---------- ----------
Ci_pou2   PAKVN---EI SPIATVKSFT QPQIKQENHK SSMSSTPLPS LFPHQMLPSN ---PNQVSST LLISSAAVTK
Cs_pou2   PQNSNI--EL NPSATVKSFP HNRIKQESQK MAT----LPA VFGHGVL-SN ---PNQ---T PILLSAAAAK
Hr_Pou3   ---------- ---------- ---------- ---------- ---------- ---------- ----------
Ci_Pou6   QQQNDLFSHM SNSQTIE--- ----KEENDN KDN---PTPH NYSDPLPPVN SIFSRQV--T STIPTSTVTK

          911
Bs_POU3   ---------- ---------- ---------- ---------- ---------- ---------- ----------
Bs_Pou4   ---------- ---------- ---------- ---------- ---------- ---------- ----------
Bs_Pou4b  ---------- ---------- ---------- ---------- ---------- ---------- ----------
Hc_pou4   ---------- ---------- ---------- ---------- ---------- ---------- ----------
MoPou4    ---------- ---------- ---------- ---------- ---------- ---------- ----------
Pm_Pou4   ---------- ---------- ---------- ---------- ---------- ---------- ----------
Ci_pou41  ---------- ---------- ---------- ---------- ---------- ---------- ----------
Ci_pou4   ---------- ---------- ---------- ---------- ---------- ---------- ----------
Cs_POU4   ---------- ---------- ---------- ---------- ---------- ---------- ----------
Sk_pou_NP ---------- ---------- ---------- ---------- ---------- ---------- ----------
Hs_Brn3   ---------- ---------- ---------- ---------- ---------- ---------- ----------
Bj_pou4   ---------- ---------- ---------- ---------- ---------- ---------- ----------
Od_pou31  ---------- ---------- ---------- ---------- ---------- ---------- ----------
Hs_Pou3   ---------- ---------- ---------- ---------- ---------- -------PPH HGLQTSV---
Bf_Pou3   ---------- ---------- ---------- ---------- ---------- -------PSA EGIMQSAG--
Od_Pou3   ---------- ---------- ---------- ---------- ---------- ---------- --------N-
Ci_pou2   ----SKEIFQ AVSARDPKHN ADTGLGAALP SHLLPQHTLD NMAYNPRLVQ -------PNH SRLQSKQSI-
Cs_pou2   ----SKGIQQ AVSLHDHK-- --LGPSPSGP NHLFQLNKMD SMAFNNR-VQ -------PNH GQSQSKTNN-
Hr_Pou3   ---------- ---------- ---------- ---------- ---------- ---------- ----------
Ci_Pou6   HISDQNNTTQ VCSSQTPV-- ------ETSP NQIMYPRNIS TLCSNMKTTP VRCVIKIPSN TNLKAQQQNL

          981
Bs_POU3   ---------- ---------- ---------- ---------- ---------- ---------- ----------
Bs_Pou4   ---------- ---------- ---------- ---------- ---------- ---------- ----------
Bs_Pou4b  ---------- ---------- ---------- ---------- ---------- ---------- ----------
Hc_pou4   ---------- ---------- ---------- ---------- ---------- ---------- ----------
MoPou4    ---------- ---------- ---------- ---------- ---------- ---------- ----------
Pm_Pou4   ---------- ---------- ---------- ---------- ---------- ---------- ----------
Ci_pou41  ---------- ---------- ---------- ---------- ---------- ---------- ----------
Ci_pou4   ---------- ---------- ---------- ---------- ---------- ---------- ----------
Cs_POU4   ---------- ---------- ---------- ---------- ---------- ---------- ----------
Sk_pou_NP ---------- ---------- ---------- ---------- ---------- ---------- ----------
Hs_Brn3   ---------- ---------- ---------- ---------- ---------- ---------- ----------
Bj_pou4   ---------- ---------- ---------- ---------- ---------- ---------- ----------
Od_pou31  ---------- ---------- ---------- ---------- ---------- ---------- ----------
Hs_Pou3   ---------- ---------- ---------- ---------- ---------- ---------- ----------
Bf_Pou3   ---------- ---------- ---LAAQNIA ASLAQ----- ---------- ---------- ----------
Od_Pou3   ---------- ---------G ---------- ---------- ---------- ---------- ----------
Ci_pou2   ---------- ---------T SVPIPATTVL NSLF------ ---------- ---------- ----------
Cs_pou2   ---------- ---------T SLPLPTTTVL NSLLN----- ---------- ---------- ---IGS----
Hr_Pou3   ---------- ---------- ---------- ---------- ---------- ---------- ----------
Ci_Pou6   INGQVSGKAK DKEFHKQTLS DMPITLTDTE SKFFNSSIPI CELVPQISDS TKSNTANTPH ISDIGNEIPQ
```

```
           1051
Bs_POU3     ---------- ---------- ---------- ---------- ---------- ---------- ----------
Bs_Pou4     ---------- ---------- ---------- ---------- ---------- ---------- ----------
Bs_Pou4b    ---------- ---------- ---------- ---------- ---------- ---------- ----------
Hc_pou4     ---------- ---------- ---------- ---------- ---------- ---------- ----------
MoPou4      ---------- ---------- ---------- ---------- ---------- ---------- ----------
Pm_Pou4     ---------- ---------- ---------- ---------- ---------- ---------- ----------
Ci_pou41    ---------- ---------- ---------- ---------- ---------- ---------- ----------
Ci_pou4     ---------- ---------- ---------- ---------- ---------- ---------- ----------
Cs_POU4     ---------- ---------- ---------- ---------- ---------- ---------- ----------
Sk_pou_NP   ---------- ---------- ---------- ---------- ---------- ---------- ----------
Hs_Brn3     ---------- ---------- ---------- ---------- ---------- ---------- ----------
Bj_pou4     ---------- ---------- ---------- ---------- ---------- ---------- ----------
Od_pou31    ---------- ---------- ---------- ---------- ---------- ---------- ----------
Hs_Pou3     ---------- ---------- ---------- ---------- ---------- ---------- ----------
Bf_Pou3     ---------- ---------- ---------- ---------- ---------- ---------- ----------
Od_Pou3     ---------- ---------- ---------- ---------- ---------- ---------- ----------
Ci_pou2     ---------- ---------- ---------- ---------- ---------- ---------- ----------
Cs_pou2     -------TSF AMNNSHHTMP SMVNSCTNSN -----FTLDS KPGAA----- ---------- ----------
Hr_Pou3     ---------- ---------- ---------- ---------- ---------- ---------- ----------
Ci_Pou6     QLLKPTPSKV AQDNFKVAGP PKLN-CSKTN VLLMVFSQTH HPDISMIQFL SAQLKLSTSS VEKWFEMMRL

           1121
Bs_POU3     -----
Bs_Pou4     ----M
Bs_Pou4b    ----M
Hc_pou4     -----
MoPou4      -----
Pm_Pou4     -----
Ci_pou41    -----
Ci_pou4     -----
Cs_POU4     -----
Sk_pou_NP   ----R
Hs_Brn3     -----
Bj_pou4     -----
Od_pou31    ----E
Hs_Pou3     -----
Bf_Pou3     ----Q
Od_Pou3     -----
Ci_pou2     -----
Cs_pou2     ----S
Hr_Pou3     ----Q
Ci_Pou6     QNNIS
```

BsHox

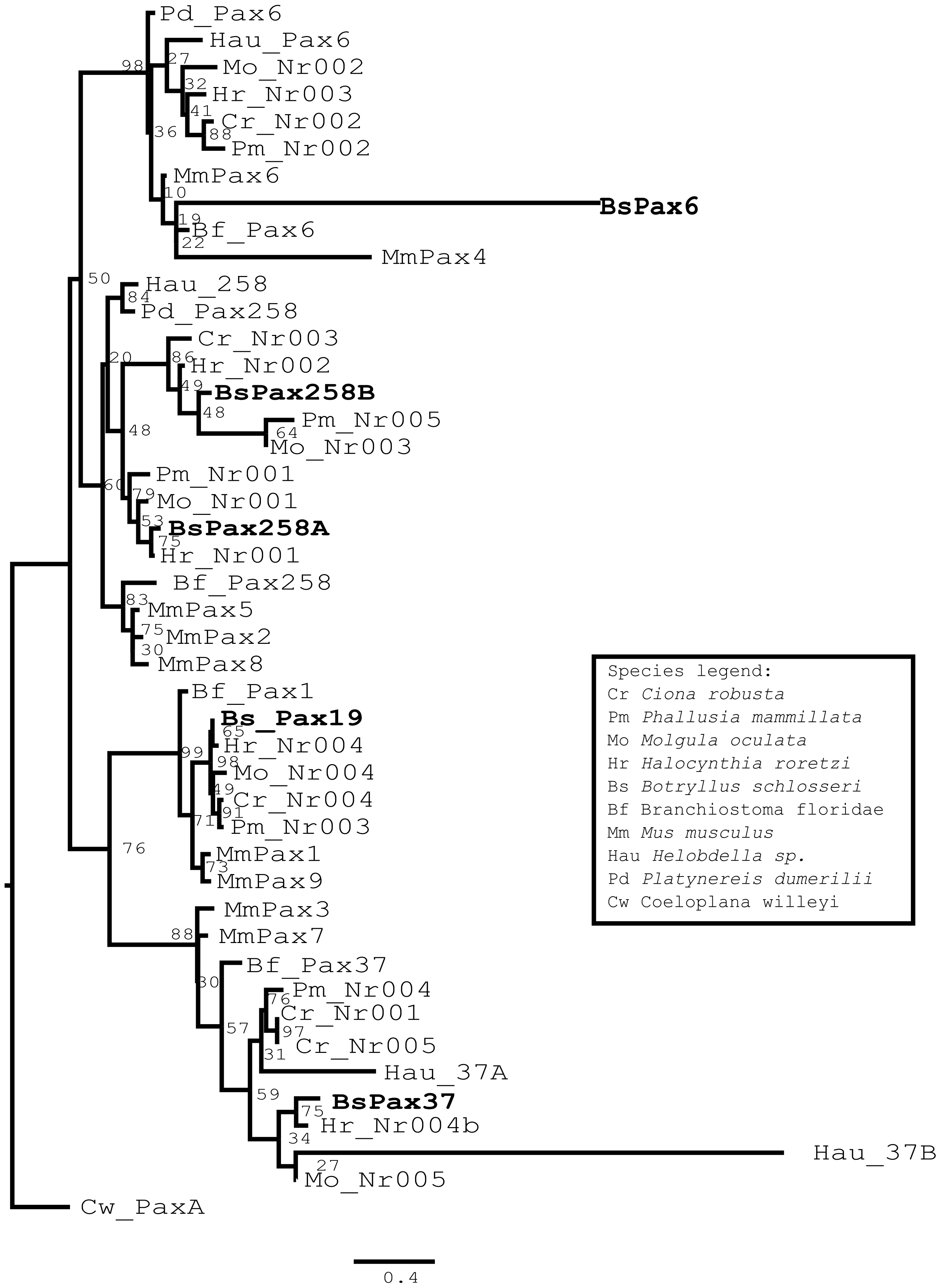

Bs_Nro06
Cr_Nr010
Cr_Nr003
Mo_Nr003
Hr_Nr006
Bs_Nr001
Bf_Hox2
Pm_Nr001
Pm_Nr007
Cr_Nr006
Pm_Nr008
Bs-Hox10
Mo_Nr006
Mo_Nr002
Bf_Hox9
Bl_Hox10
Bf_Hox14
Bf_Hox13
Bf_Hox11
Bf_Hox12
Cr_Nr008
Bf_Hox3
Cr_Nr007
Bs-Hox3
Hr_Nr002
Cr_Nr009
Hr_Nr007
Bs-Nr003
Od_Nr001
Bf_Hox4
Bs_Nr002
Mo_Nr001
Pm_Nr002
Cr_Nr004
Pm_Nr004
Bf_Hox7
Bf_Hox6
Bf_Hox8
Bl_Hox5
Cr_Nr001
Cr_Nr005
Pm_Nr005
Pm_Nr006
Bs_Nr004
Mo_Nr004
Bf_Hox1
Cr_Nr002
Hr_Nr003
Pm_Nr003
Bs_Hox1
Hr_Nr005
Bs_N005
Mo_Nr005
Cr_Nr011
Hr_Nr009
Hr_Nr001
Bf_NOTO-like
Cr_Nr012
Pm_Nr009
Bb_MOX
2.0
Species legend:
Cr Ciona robusta
Pm Phallusia mammillata
Mo Molgula oculata
Hr Halocynthia roretzi
Bs Botryllus schlosseri
Bb Branchiostoma belcheri
Bf Branchiostoma floridae
Bl Branchiostoma lanceolatum

Mafft Alignment: Hox trimal: gappyout

```
Bs_Nr006 LTSNSYSYPN YGSYSSVQHK TKPVSAQ--- ---------T AQTKPHKKIR TSFTKDQVQQ LEDDFLRNNY
Bs_Nr004 MSVHTEFAPT PRNSSDSEIR NGSNPGSTAG KDTAIYPWMN QTTDQLKRPR TAYTRFQTLE LEKEFHFNQY
Bs_Nr002 MAANTSAMPG PGKNGDHEGK LGKCMKSKGE SPPTVYPWMN SGAEQGKRPR TAYTRHQVLE LEKEFHFNRY
Cr_Nr001 LNNWGIMEPR PTSPSDVTER KSDESSDGSD RDAIIYPWMG ETPDPSKRTR TAYTRYQTLE LEKEFHYNRY
Cr_Nr005 LNNWGIMEPR PTSPSDVTER KSDESSDGSD RDAIIYPWMG ETPDPSKRTR TAYTRYQTLE LEKEFHYNRY
Pm_Nr005 MSK------- ---------- ---------- -----FVFSA GDDDPSKRTR TAYTRYQTLE LEKEFHYNRY
Pm_Nr006 VSNAASFLPQ PNSQSGSEVE EKRDKEEKTD SEAKIYSWMA GDDDPSKRTR TAYTRYQTLE LEKEFHYNRY
Mo_Nr004 ITE------- -NRNNDFCLR STKKSAIELT ETLPFYMNLD ERLDPQKRTR TAYTRFQTLE LEKEFHFNRY
Bl_Nr005 MSSTCTSPPS SASQNPSNNN TANNTASTSP SNIPMYPWMS AGTGDNKRTR TAYTRYQTLE LEKEFHFNRY
Pm_Nr002 MSVRGVHTPN DANQNIHEDE FGKDSHGSDS TPPVVYPWM- NGMDPGKRPR TAYTRHQVLE LEKEFHYNRY
Mo_Nr001 MRT------- ---------- ---------- ---------T EQGDTGKRPR TAYTRHQVLE LEKEFHYNRY
Bf_Nr003 MSSESASPPQ SGHP-GHPVQ QANPALNTAQ MTTPIYPWM- -TAPERKRGR QTYTRYQTLE LEKEFHFNKY
Pm_Nr004 MSCRQSSYPY LGHQTAGNAE DVFEADDPEA SQQPIYPWM- -HGPQRRRGR QTYSRYQTLE LEKEFHYNRY
Cr_Nr004 MSI----FPT HGRGKDDKTD EEYQAMGSSN GPPVVYPWM- NGLESGKRPR TAYTRHQVLE LEKEFHYNRY
Bf_Nr011 MSSPPHSRPT SNHSATTVSD GGSPPSNGPP TVPMVYPWMS YNGQDTKRSR TAYTRQQVLE LEKEFHFNRY
Bs-Hox10 VSSADVFSPR STSSSALEYD VDAPYQDKSE DEDPTKHWL- --TANGRKKR VPYTKYQLLE LEKEFHYNQY
Cr_Nr006 QTSKTSTPPA DENRSPSHVT TKEETEDAGE EADPTKHWL- --TASGRKKR VPYTKYQLLE LEKEFHYNQY
Bf_Nr001 MTSTGYSQPY R----DGQMK VKSEEESHGS VVANQPGWM- -NNHSSRKKR CPYTRFQTLE LEKEFLYNMY
Bf_Nr006 ---------- ---------- ---------- ----TSNWM- -SAKSTRKKR CPYTKYQTLE LEKEFLFNMF
Bf_Nr009 ---------- ---------- ---------- ----TDVWW- -KLQSSRKKR CPYSKVQLLE LEKEFLYNMY
Bf_Nr014 MTSGPTTGPV G----HPNTA CSTPPASSPL HVCGATSWM- -APRVGRKKR CPYTKYQILE LEKEFLFNMY
Bf_Nr007 ---------- ---------- ---------- ---------G SLTKPVRPKR RPYSKYQLNE LENEYVQNQY
Bf_Nr008 ---------- ---------- ---------- ---------Q SVARGGRKKR CPYTKYQLSV LEQEYIQNRY
Mo_Nr006 MPK------- -KNEAYTKVQ VQYPTTLLSV ESDPTRHWL- --TANGRKKR VPYTKFQLLE LEKEFHYNQY
Pm_Nr008 MSTDLKTAPT SEEASNDQTS VSPFGADSTD ESDPTKNWL- --TASGRKKR VPYTKFQLLE LEKEFHYNQY
Mo_Nr002 MDSLTNQNSV SNNKLKNKEG VYHPTLTKHE ESDPTRHWL- --TANGRKKR VPYTKFQLLE LEKEFHYNQY
Od_Nr001 MNWVNGAIPQ SHHQDDLGSC DSREKKSDDA KSLKIYPWM- -PGNDNKRIR TAYTRHQTLE LEKEFHFNRY
Bf_Nr004 MSSASHHLPQ YNNTSAQGSP GSTVGAQLPQ PTPPVFPWMT AMGEEKKRGR QTYTRYQTLE LEKEFHFNKY
Bf_Nr002 MSSRDNVSPN CDNQTQQEEK TATMGMAQSQ LAIPFYPWM- -AGPERRRGR QTYSRYQTLE LEKEFHFNKY
Bs-Hox3  MSQAANLFPT GGHERHQNGD TPEMGEGSSP VSKHIYPWMN GGDKSSKRER TAYTNAQLVE LEKEFHFNRY
Hr_Nr007 MTS------- ---------- ---------- ---------F PEVEENKRTR TAYARWQLLE LEKEFHYSRY
Hr_Nr002 MSVSRNDKPT RKNNSRQNGN SAQKHSANCS SSKHIYPWMS NGDKSSKRER TAYTNSQLVE LEKEFHFNRY
Bf_Nr012 INSWPALHPT TSDLS----- -------TQP ANIADFPWVT DQVNSSRRLR TVFTNTQLLE LEKEFHYNKY
Bf_Nr013 MQKTGNVSPT CMKTNDNNHG SAPQSNASIL ANTKIYPWMG LGGAAGKRAR TAYTSAQLVE LEKEFHFNRY
Cr_Nr010 ---HAESYPI QWNQAQEQVN NFSSHQHTQP ESQVSY---T SEVEIRKKER TSFTHDQVRQ LEADFMENHY
Bs_Nr003 YSSSNGYIPT NGQTQNRSTE TDDTESSDAK DKALKFPWMY PEVEENKRTR TAYARWQLLE LEKEFHFSRY
Cr_Nr009 LSAHRASNPR GQKQSHDCSF SGYVAGKLAG TSSVKFQWMF PEVDENKRTR TAYTRWQLLE LEKEFHFSRY
Cr_Nr003 INSRFSYRPQ FAYEEPYRGI PSYPTAHIRD EQCPDFPWMI RPAGASRRLR TAYTNTQLLE LEKEFHYNKY
Bs_Nr001 INSGLNQIPH DSDVANFGNY DPHQSPAYMT TPHPDYPWMV RPPGTSRRLR TAYTNTQLLE LEKEFHFNKY
Pm_Nr001 QTA------- ---------- ---------- ---------I RPPGASRRLR TAYTNTQLLE LEKEFHYNKY
Pm_Nr007 QTA------- ---------- ---------- ---------I RPPGASRRLR TAYTNTQLLE LEKEFHYNKY
Mo_Nr003 ---------- ---------- ---------- ---------L KANGSSRRFR TAYTNTQLLE LEKEFHYNKY
Bs-Hox1  MSSPTSPYPT SAQYPDYGER DKLDNAATNS DLANTYDWMY GYVNQVGNGR TNFSTKQLTE LEKEFHYNKY
Hr_Nr003 ---------- ---------- ---------- ---------Y GYAGTAGNGR TNFTTKQLTE LEKEFHFNKY
Cr_Nr002 MNSINESYPT SSQYNNSAYD TPTASPQTNT LPSNTYDWMY TYGVTGNNGR TNFTTKQLTE LEKEFHFNKY
Pm_Nr003 MNSIASHYPN AASYNNINFS SASEQQQGTT DSANTYDWMS GYGNQLGNGR TNFTTKQLTE LEKEFHYNKY
Bf_Nr010 MNSMPNGDPQ NSSYNHHTGQ DPMISTTAPS PPVATYDWMG FTTSGPNNGR TNFTTKQLTE LEKEFHYNKY
Hr_Nr006 MNF------- ---------- ---------- ---------I RPPGTSRRLR TAYTNTQLLE LEKEFHYNKY
Bb_MOX   LAGHTSGMPE NGQEQNVVND DDRLSCS--- ---------L DLSAKPRKER TAFTKQQIME LENEFRHHNY
Cr_Nr007 VAAEMSKSPA SNSKQ----- -------QQN KDPGIFDWMD KCYSASKRER TAYTNSQLVE LEKEFHFSHY
Cr_Nr008 VQSHSRDFPN NCSERYHRND PSYMVDQGSG SSMQQLPAVD DEQGDRKRMR TAFTGSQLLE LEREFHSDMY
Bs_Nr005 ISSENTADPT NSAASHHVAV NPHPSQPGIG PHLPALDLMG FFNNKARRPR TAFTSQQLLE LESQFKQNKY
Hr_Nr005 LSASSQSQPR QQHNSLASHP PSFPQQPTLP NQQAAMEWNP GILNKSRRPR TAFTSQQLLE LERQFRQNKY
Mo_Nr005 MAADHHVSPT LAAIAPSQHY PSLPQNPTAQ QQVVAMEMLP GLLSRSRRPR TAFTSQQLLE LEKQFKQNKY
Cr_Nr012 MDSNMNTKPT TDEENVVGAA QSSLSPCATS HDDVTKQLLS PQNEIKKRPR TAFTPEQIKR LEAEFHRNKY
Pm_Nr009 LTDSDESAPI WAESN----- -------SVS SVSSDITKVS PPGEVKKRPR TAFTPDQIKR LESEFHRNKY
Hr_Nr001 VSFNSHLQPT IANSTPSSHW PSPIPQGTVG QRRSSMENVG QSEAKAKRLR TIFTPEQLER LEREFLKQQY
Bf_NOTO  MTTKEHHIPA AQAQTPYSYL QTAPKVKPLG YKTKGCPLAD GKSRKGKRIR TIFTPEQLER LEREFSRQQY
Cr_Nr011 MTVFTSYNPT FKSHSDCSYE STSPKSNVIP TSPNAFRISL RKHKPNRKPR TPFSTEQLLS LERRFQDKQY
Hr_Nr009 MVSANSARPK HSSTRLHKQN TSDVDSGATT SLSCSTPPLG GKLKKCRRSR TVFTELQLTG LEKRFENQKY
```

```
                71
Bs_N006     LTRLRRYELA MKLNLSERQI KVWFQNRRMK WKRQSVVAHH SHGSQDKVSP M
Bs_N004     LTRRRRIEIA QDLSLTERQI KIWFQNRRMK WKKSNTLRPS QRYDATFD-- -
Bs_N002     LTRRRRIEIA HSLCLTERQV KIWFQNRRMK WKKDNKLPNT KSRVLQGSGM S
Cr_H001     LTRRRRIEVA HTLCLTERQI KIWFQNRRMK WKKENKLSSL NSVTHVSGPL S
Cr_H005     LTRRRRIEVA HTLCLTERQI KIWFQNRRMK WKKENKLSSL NSVTHVSGPL S
Pm_H005     LTRRRRIEIA HALCLTERQI KIWFQNRRMK WKKGNALRNP LAG------- -
Pm_H006     LTRRRRIEIA HALCLTERQI KIWFQNRRMK WKKENKVPPP PPPPPPPPPP P
Mo_H004     LTRRRRIEIA HTLCLSERQI KIWFQNRRMK WKKDNNVKSL NEVSHISGPA K
Bl_H005     LTRRRRIEIA HALCLTERQI KIWFQNRRMK WKKENKLKSL SQCQQTQGLN P
Pm_H002     LTRRRRIEIA HNLCLSERQV KIWFQNRRMK WKKDHKLPNT KVRNSSSSSS S
Mo_H001     LTRRRRIEIA HSLCLTERQV KIWFQNRRMK WKKDHKLPNT KTNHTQAKKA R
Bf_H003     LTRRRRIEIA HALCLTERQI KIWFQNRRMK WKKENKLESL KQQPAESETS S
Pm_H004     LTRRRRIEIA HALCLTERQI KIWFQNRRMK WKKENKEVTP NALFQDKEVL E
Cr_H004     LTRRRRIEIA HGLCLSERQV KIWFQNRRMK WKKDHKLPNT KVRNPPTSSL P
Bf_H011     LTRRRRIEIA HSLGLTERQI KIWFQNRRMK WKKDNRLPNT KTRSSSASGG S
Bs_Hox10    LSRERRQEVA KAVTLSDRQV KIWFQNRRMK WKKEKKEEKV RDG------- -
Cr_H006     LSRERRLEVA KSVKLTDRQV KIWFQNRRMK WKKERREERQ RDDHHPMSGI H
Bf_H001     LTRERRYEIS QHVNLTERQV KIWFQNRRMK MKKMSKQRQE QQQSPPQ--- -
Bf_H006     VTRERRQEIA RQLNLTDRQV KIWFQNRRMK MKRMKQRAMQ QLMEEKQQQQ Q
Bf_H009     ITREQRGEIA RKVNLTDRQV KIWFQNRRMK MKRMKQRHEE E--------- -
Bl_H014     VSRERRQEIS RHVNLSDRQV KIWFQNRRMK MKRMNKAREE QIRNHQERGH Q
Bf_H007     ISRDKRLQLS QKLNLTERQV KIWFQNRRIK QKKLDRRNSE MCP------- -
Bf_H008     VSRETRLELS QRLNLTDRQV KIWFQNRRMK QKRLEFRSRN QT-------- -
Mo_H006     LTRERRLEVA KSVSLSDRQV KIWFQNRRMK WKKERKEEKM RDG------- -
Pm_H008     LSRERRLEVA KAVKLTDRQV KIWFQNRRMK WKKERREERM RESVPHEQLA G
Mo_H002     LTRERRLEVA KSVHLSDRQI KIWFQNRRMK WKKEKKEDSM KSMLDIASPN F
Od_H001     LTRKRRIEIA TTLQLSERQV KIWFQNRRMK WKKEHKSNGK NA-------- -
Bf_H004     LTRKRRIEIA HLLGLTERQI KIWFQNRRMK WKKENKIPSL NATTINQLTN M
Bf_H002     LTRRRRIEIA HALGLTERQI KIWFQNRRMK LKKEAAMLCP PKAETETEKS S
Bs_Hox3     LCRPRRIEMA QMLNLTERQI KIWFQNRRMK YKKDQKLKVV PMARDDGFFH Q
Hr_H007     ISRPRRIELA AMLNLTERHI KIWFQNRRMK WKKDQAICRR SSNGDDDETL G
Hr_H002     LCRPRRIEMA QMLSLTERQI KIWFQNRRMK YKKDQKLKVM PIHPLTRDNP Y
Bf_H012     VCKPRRKEIA SYLDLNERQV KIWFQNRRMR QKRRDTKGRS EIGTDPGTAA E
Bf_H013     LCRPRRVEMA AMLNLTERQI KIWFQNRRMK YKKEQKVKGG GSGGGSGGMN S
Cr_H010     LTRLRRYELA LKLNLTERQI KVWFQNRRMK LKREKT---- ---------- -
Bs_N003     ISRPRRIELA AMLSLTERHI KIWFQNRRMK WKKD------ ---------- -
Cr_H009     ISRPRRIELA AMLNLTERHI KIWFQNRRMK WKKDQAANSK TGKVRDITAE I
Cr_H003     LCRPRRIEIA TLLDLTERQV KVWFQNRRMK HKRQQLQQSQ KDNDDSDVTS A
Bs_N001     LCRPRRIEIA TLLDLTERQV KVWFQNRRMK HKRQQQQNKH DDQKAGDESA N
Pm_H001     LCRPRRIEIA TLLDLTERQV KVWFQNRRMK HKRQQQQQSQ KDKCGDSDDS S
Pm_H007     LCRPRRIEIA TLLDLTERQV KVWFQNRRMK HKRQQQQQSQ KDKCGDSDDS S
Mo_H003     LCRPRRIEIA TLLDLTERQV KVWFQNRRMK HKRQRKESKN DGNDDSKAEI G
Bs_Hox1     LTRARRHEIA DSLSLNETQV KIWFQNRRMK QKKRDKEAEK ILKANATQLQ P
Hr_H003     LTRARRVEIA AALLLNETQV KIWFQNRRMK QKKRDKEAEK LIQQNRNAVD T
Cr_H002     LTRARRVEIA AALRLNETQV KIWFQNRRMK QKKRDKEAEK LNGNKMSQSK S
Pm_H003     LTRARRVEIA AALHLNETQV ---------- ---------- ---------- -
Bf_H010     LTRARRVEIA AALNLNETQV KIWFQNRRMK QKKREKENGF STPGSGGSPA G
Hr_H006     LCRPRRIEIA TLLDLTERQV KVWFQNRRMK HKRQQQQSKQ DGKDDKDDDE D
Bb_MOX      LTRLRRYEIA VKLDLTERQV KVWFQNRRMK WKRTKGGAAM RDKEALKELL P
Cr_H007     LCRPRRIELA QGLGLTERQI KIWFQNRRMK FKKEQKQKAV LQRNNGDSVM G
Cr_H008     LTRLRRIRIA QMLNLSEKQI KIWFQNRRVK KKKGEKPQQD SREQ------ -
Bs_N005     LTRPRRYEVA RDLCLTETQV KIWFQNRRMK WKRSKAEREE RHKESLTSTS -
Hr_H005     LSRPMRFEVA TRLCLTETQV KIWFQNRRMK WKRNKKEREE SQQGGCDSHN T
Mo_H005     LSRPKRFEVA TNLCLTETQV KIWFQNRRMK WKRSKKENSE KSDDEN---- -
Cr_H012     LSVGKRMELS KALKLTETQI KIWFQNRRTK WKREYLSEWE VWAHQNYYAM H
Pm_H009     LSVGKRMELS KALKLTETQI KIWFQNRRTK WKREYLSEWE VWMHQNYYAM H
Hr_H001     MVGTERFYLA KELNLGEAQV KVWFQNRRIK WRKQKHAGEL NNKSPYTSVT K
BF_H001     MVGSERYHLA ASLNLTEAQV KVWFQNRRIK WRRQNYEQQQ ARLAQLSSSH T
Cr_H011     LSIAERAEFS ASLALSETQV KIWFQNRRAK AKRLHEAEFE KVKLAAAAAA Y
Hr_H009     LSTPDRMELA ETLSLSQLQV KTWYQNRRMK WKKQVMQGGG TEPPTKPKGR P
```

BsPax

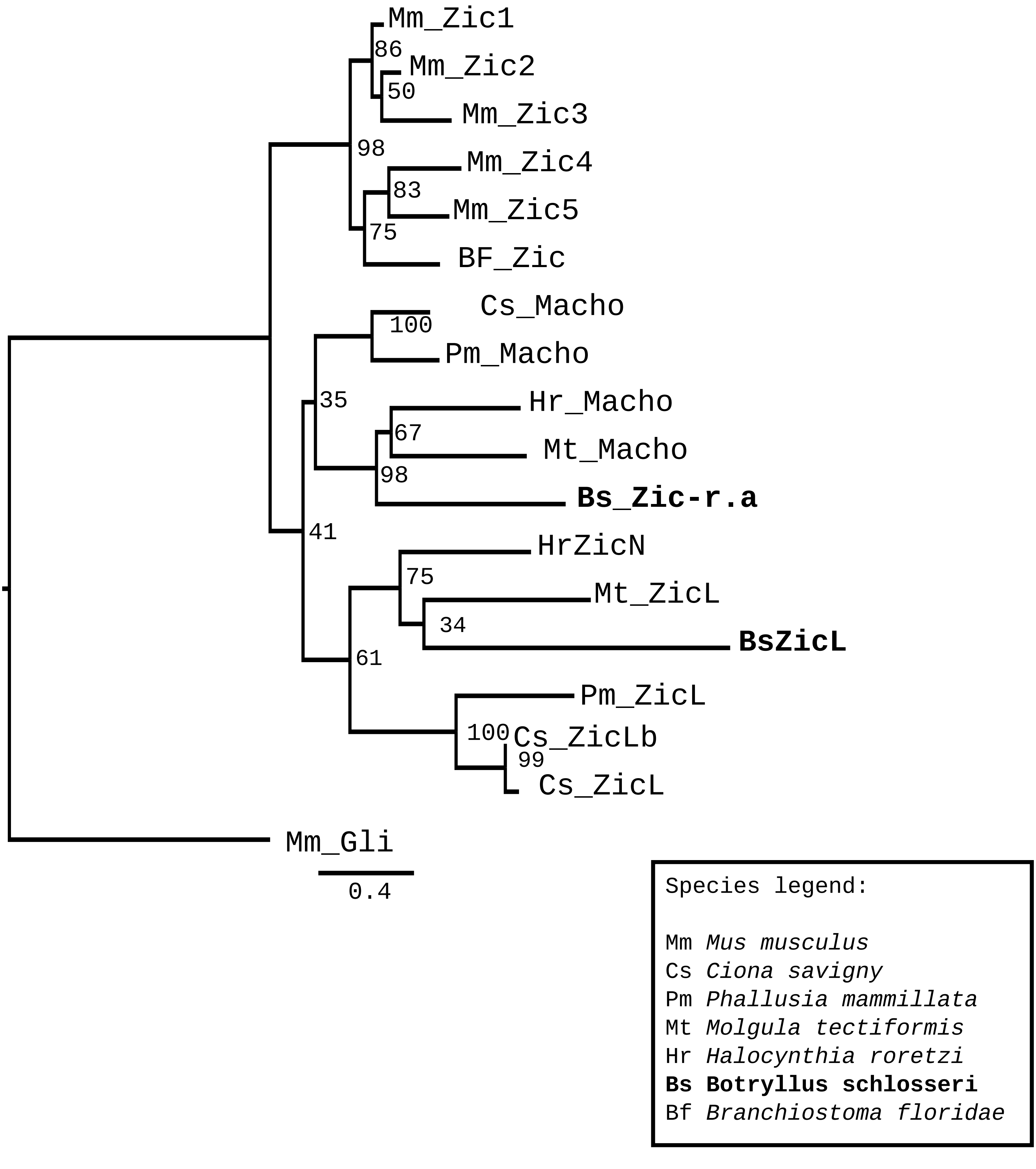

Pd_Pax6
Hau_Pax6
Mo_Nr002
Hr_Nr003
Cr_Nr002
Pm_Nr002
MmPax6
BsPax6
Bf_Pax6
MmPax4
Hau_258
Pd_Pax258
Cr_Nr003
Hr_Nr002
BsPax258B
Pm_Nr005
Mo_Nr003
Pm_Nr001
Mo_Nr001
BsPax258A
Hr_Nr001
Bf_Pax258
MmPax5
MmPax2
MmPax8
Bf_Pax1
Bs_Pax19
Hr_Nr004
Mo_Nr004
Cr_Nr004
Pm_Nr003
MmPax1
MmPax9
MmPax3
MmPax7
Bf_Pax37
Pm_Nr004
Cr_Nr001
Cr_Nr005
Hau_37A
BsPax37
Hr_Nr004b
Hau_37B
Mo_Nr005
Cw_PaxA
Species legend:
Cr Ciona robusta
Pm Phallusia mammillata
Mo Molgula oculata
Hr Halocynthia roretzi
Bs Botryllus schlosseri
Bf Branchiostoma floridae
Mm Mus musculus
Hau Helobdella sp.
Pd Platynereis dumerilii
Cw Coeloplana willeyi
0.4

Pax

```
BsPax37     MQTGQGRVNQ LGGVFINGRP LPNHVRHKIV EMAAAGIRPC MISRQLRVSH GCVSKILCRY QETGSIKPGA
Hr_Nr004b   MDTCQGRVNQ LGGVFINGRP LPNHIRHKIV EMAAHGIRPC VISRQLRVSH GCVSKILCRY QETGSIKPGA
Mo_Nr005    MLEGQGRVNQ LGGVFINGRP LPNHIRHKIV EMAAHGVRPC VISRQLRVSH GCVSKILCRY QETGSIKPGA
Cr_Nr001    MLEGQGRVNQ LGGMFINGRP LPSHVRHKIV EMAAHGVRPC VISRQLRVSH GCVSKILCRY QETGSIKPGA
Cr_Nr005    MLEGQGRVNQ LGGMFINGRP LPSHVRHKIV EMAAHGVRPC VISRQLRVSH GCVSKILCRY QETGSIKPGA
Pm_nr004b   MLEGQGRVNQ LGGMFINGRP LPNHIRHKIV EMAAHGVRPC VISRQLRVSH GCVSKILCRY QETGSIKPGA
MmPax3      MPLGQGRVNQ LGGVFINGRP LPNHIRHKIV EMAHHGIRPC VISRQLRVSH GCVSKILCRY QETGSIRPGA
MmPax7      MPLGQGRVNQ LGGVFINGRP LPNHIRHKIV EMAHHGIRPC VISRQLRVSH GCVSKILCRY QETGSIRPGA
Bf_Pax37    MLPGQGRVNQ LGGVFINGRP LPNHIRHKIV EMAAAGIRPC VISRQLRVSH GCVSKILCRY QETGSIKPGA
BsPax258A   MRSGHGGINQ LGGVYVNGRP LPDPVRQTIV DMAHQGVRPC DIARQLRVSH GCVSKILARY YETGSIKPGI
Hr_Nr001    MRSGHGGINQ LGGVYVNGRP LPDPVRQTIV DLAHQGVRPC DIARQLRVSH GCVSKILARY YETGSIKPGV
Pm_Nr001    -RSGHGGINQ LGGVYVNGRP LPDQIRQQIV DQAHHGVRPC DIARQLRVSH GCVSKILARY YETGSIRPGV
Mo_Nr001    MRSGHGGVNQ LGGVYVNGRP LPDPVRQRIV DMAHQGVRPC DIARQLRVSH GCVSKILARF YETGSIKPGV
Hau_258     MQDSHGGVNQ LGGVFVNGRP LPDMVRQRIV EMAHQGVRPC DISRQLRVSH GCVSKILGRY YETGSIKPGV
Pd_pax258   MKQGHGGVNQ LGGVFVNGRP LPDVVRTRIV ELAHQGVRPC DISRQLRVSH GCVSKILGRY YETGSVRPGV
MmPax2      MHPGHGGVNQ LGGVFVNGRP LPDVVRQRIV ELAHQGVRPC DISRQLRVSH GCVSKILGRY YETGSIKPGV
MmPax5_     MRTGHGGVNQ LGGVFVNGRP LPDVVRQRIV ELAHQGVRPC DISRQLRVSH GCVSKILGRY YETGSIKPGV
MmPax8      MRSGHGGLNQ LGGAFVNGRP LPEVVRQRIV DLAHQGVRPC DISRQLRVSH GCVSKILGRY YETGSIRPGV
Bf_PAx258   MSGGHGGVNQ LGGVYVNGRP LPDVVRHRIV ELAHQGVRPC DISRQLRVSH GCVSKILRRY YETGSIKPGV
BsPax19     MEQTFGEVNQ LGGVFVNGRP LPNAIRLRIV ELAQLGIRPC DISRQLRVSH GCVSKILARY NETGSILPGA
Hr_Nr004    MEQTFGEVNQ LGGIFVNGRP LPNTIRLRIV ELAQLGIRPC DISRQLRVSH GCVSKILARY NETGSILPGA
Mo_Nr004    KDQNYGEVNQ LGGVFVNGRP LPNVIRLKIV ELAQLGIRPC DISRQLRVSH GCVSKILARY NETGSILPGA
Cr_Nr004    MNQTFGEVNQ LGGVFVNGRP LPNALRLRII ELAQLGIRPC DISRQLRVSH GCVSKILARY NETGSILPGA
Pm_nr003    MDQTFGEVNQ LGGVFVNGRP LPNAIRLRII ELAQLGIRPC DISRQLRVSH GCVSKILARY NETGSILPGA
MmPax1      MEQTYGEVNQ LGGVFVNGRP LPNAIRLRIV ELAQLGIRPC DISRQLRVSH GCVSKILARY NETGSILPGA
MmPax9      MEPAFGEVNQ LGGVFVNGRP LPNAIRLRIV ELAQLGIRPC DISRQLRVSH GCVSKILARY NETGSILPGA
Bf_Pax1     MEQTFGEVNQ LGGVFVNGRP LPNAIRLRIV ELAHVGIRPC DISRQLRVSH GCVSKILARY NETGSILPGA
BsPax6      MATGHSGVNQ LGGAFVNGRP LPGDTRQRIV KLAHEGHRPC DISRLLQVSN GCVSKILARY VLAFTAGPYD
Hr_Nr003    MDSGHSGVNQ LGGMFVNGRP LPDSTRQKIV ELAHNGARPC DISRILQVSN GCVSKILARY YETGTIRPRA
Mo_Nr002    MHKGHSGINQ LGGMFVNGRP LPDTTRQRIV ELAHNGARPC DISRILQVSN GCVSKILARY YETGTIRPRA
MmPax6      MQNSHSGVNQ LGGVFVNGRP LPDSTRQKIV ELAHSGARPC DISRILQVSN GCVSKILGRY YETGSIRPRA
Bf_Pax6_    M--GHSGVNQ LGGVFVDGRP LPDSTRQKIV ELAHQGARPC DISRLLQVSN GCVSKILGRY YETGSIRPRA
Pd_pax6_    MKRGHSGVNQ LGGVFVNGRP LPDSTRQRIV ELAHSGARPC DISRILQVSN GCVSKILGRY YETGSIRPRA
Cr_Nr002    MHKGHSGMNQ LGGMFVNGRP LPDSIRQKIV EFAHNGARPC DISRILQVSN GCVSKILARY YETGTIRPRA
Pm_nr002    HHKGHSGVNQ LGGMFVNGRP LPDSIRQKIV EFAHNGARPC DISRILQVSN GCVSKILARY YETGTIRPRA
Hau_pax6    MKRGHSGVNQ LGGMFVNGRP LPDTTRQKIV ELAHNGARPC DISRILQVSN GCVSKILGRY YESGSIRPRT
MmPAx4      MQDGLSSVNQ LGGLFVNGRP LPLDTRQQIV QLAIRGMRPC DISRSLKVSN GCVSKILGRY YRTGVLEPKC
BsPax258B   MDAGGSGVNQ LGGMYVNGRP LPEPIRKRIV DLSHQGVRPC DISRQLRVSH GCVSKILARF YETGSIRPGV
Hr_Nr002_   MSSGGNGVNQ LGGMYVNGRP LPEPIRKRIV DLSHQGVRPC DISRQLRVSH GCVSKILARY YETGSIRPGV
Mo_Nr003_   MDSGGSGVNQ LGGMYVNGRP LPEPIRQRIV DLSHQGVRPC DISRQLRVSH GCVSKILARF YETGSIKPGV
Pm_nr005    EDDGGSGVNQ LGGMYVNGRP LPEPIRQRIV NLSHQGVRPC DISRQLRVSH GCVSKILAR- ----------
Cr_Nr003    MRSSGGGVNQ LGGMYVNGRP LPEPIRQRIV NLSHQGVRPC DISRQLRVSH GCVSKILARY YETGSIRPGV
Hau_37A     MRNGQGKVNQ LGGLFINGRP LPLNIRLKIV EMASQGVRPC VISRTLCVSH GCVSKILQRY QETGSIRPGA
Cw_PAxA_    MDNRRIGVNQ LGGAFKNGRP LPEDIRKRIV DLAHFGVRPC DISRQLRVSH GCVSKILGRY SDTGSILPGN
Hau_37B_    M--SQGRVNQ LGGVFINGRP LPNHVRLKII QLASQGVRPC VISRSLRVSH GCVSKILQRY QETGSIRPGS
```

```
            71
BsPax37     IGGSKPKVSN AEIESKIEQY KKENPTMFSW EIRDQLIKDM LDRASAPTVS AISRILRAKG CDTEN
Hr_Nr004b   IGGSKPKVTN SEIESKIEQY KKDSPSMFSW EIRDQLIKEG LDRSSAPTVS AISRILRSKG CDTSN
Mo_Nr005    IGGNKPKGNN NDIEKKIEEY KKESPSMFSW EIRDKLIKEN IDRSSAPTVS AISRVLRAKG CDTSN
Cr_Nr001    IGGSKPKVAT ADVDNKIEEY KKENPGIFSW EIRERLIKEG IDRSNVPSVS SISRTLRAKG CDVEN
Cr_Nr005    IGGSKPKVAT ADVDNKIEEY KKENPGIFSW EIRERLIKEG IDRSNVPSVS SISRTLRAKG CDVEN
Pm_nr004b   IGGSKPKVAS LDVENKIEEY KNENPGIFSW EIRERLIKDG IERNTVPSVS SISRTLRAKG CDVDT
MmPax3      IGGSKPKVTT PDVEKKIEEY KRENPGMFSW EIRDKLLKDA VDRNTVPSVS SISRILRSKF GKGEE
MmPax7      IGGSKPRVAT PDVEKKIEEY KRENPGMFSW EIRDRLLKDG HDRSTVPSVS SISRVLRIKF GKKED
Bf_Pax37    IGGSKPRVAT PEVEKKIEDY KRDNPGMFSW EIRDRLLKDG MDRSTVPSVS SISRILRGKG HKISD
BsPax258A   IGGSKPKVAT PKVVDKICEY KRQNPTMFAW EIRDRLLLEC IDSENVPSVS SINRIVRNKA AEKSK
Hr_Nr001    IGGSKPKVAT PRVVEKICEY KRQNPTMFAW EIRDRLLVEC IDTENVPSVS SINRIVRNKA AEKSK
Pm_Nr001    IGGSKPKVAT PKVVEKICEY KRQNPTMFAW EIRDRLLGEQ IDQDNVPSVS SINRIVRNKA AENAK
Mo_Nr001    IGGSKPKVAT PRVVEKICDY KRQNPTMFAW EIRDRLLREC IDADNVPSVS SINRIVRNKA AENTK
Hau_258     IGGSKPKVAT PKVVDAICKY KRENATMFAW EIRDRLLAEG VDQENVPSVS SINRIVRNKA AEKTK
Pd_pax258   IGGSKPKVAT PKVVGAICKY KRENPTMFAW EIRDRLLAEG VDQENVPSVS SINRIVRNKA AEKAK
MmPax2      IGGSKPKVAT PKVVDKIAEY KRQNPTMFAW EIRDRLLAEG IDNDTVPSVS SINRIIRTKV QQPFH
MmPax5_     IGGSKPKVAT PKVVEKIAEY KRQNPTMFAW EIRDRLLAER VDNDTVPSVS SINRIIRTKV QQPPN
MmPax8_     IGGSKPKVAT PKVVEKIGDY KRQNPTMFAW EIRDRLLAEG VDNDTVPSVS SINRIIRTKV QQPFN
Bf_PAx258   IGGSKPKVAT PKVVEKIAEY KRQNPTMFAW EIRDRLLAEG IDNDTVPSVS SINRMTRNLV HRIVR
BsPax19     IGGSKPRVTT PSVVNAIKDY KVRDPGIFAW EIRDRLLSDC IDKYNVPSVS SISRILRNKI GNVLH
Hr_Nr004    IGGSKPRVTT PSVVNAIKDY KIRDPGIFAW EIRDRLLSDC IDKYNVPSVS SISRILRNKI GNVLH
Mo_Nr004    IGGSKPRVTT PSVVNAIKDY KLRDPGIFAW EIRDRLLSDA IDKYNVPSVS SISRILRNKI GNVLH
Cr_Nr004    IGGSKPRVTT PGVVNAIKDY KVRDPGIFAW EIRDRLLSDA VDKYNVPSVS SISRILRNKI GNVLH
Pm_nr003    IGGSKPRVTT PGVVQAIKEY KVRDPGIFAW EIRDRLLSDA VDKYNVPSVS SISRILRNKI GNVLH
MmPax1      IGGSKPRVTT PNVVKHIRDY KQGDPGIFAW EIRDRLLADG VDKYNVPSVS SISRILRNKI GSLAQ
MmPax9      IGGSKPRVTT PTVVKHIRTY KQRDPGIFAW EIRDRLLADG VDKYNVPSVS SISRILRNKI GNLAQ
Bf_Pax1     IGGSKPRVTT PEVVKAIKKY KTLDPGIFAW EIRDRLLAEG VDKYNVPSVS SISRILRNKI GNTTQ
BsPax6      FSTIVCRISV TAATAK---- ---------- ---------- ---------- ---------- -----
Hr_Nr003    IGGSKPRVAT PEVVTRIARY KRECPSIFAW EIRDRLLNEG INSENIPSVS SINRVLRNLT TDVVT
Mo_Nr002    IGGSKPRVAT PEVVARIALF KQECPSIFAW EIRDRLLSEK INAENIPSVS SINRVLRNFA NDKLN
MmPax6      IGGSKPRVAT PEVVSKIAQY KRECPSIFAW EIRDRLLSEG VTNDNIPSVS SINRVLRNLA SEKQQ
Bf_Pax6_    IGGSKPRVAT PEVVAKIAQF KRECPSIFAW EIRDRLLSEG ITNENIPSVS SINRVLRNLA SGEKN
Pd_pax6_    IGGSKPRVAT PEVVNKVAQY KRECPSIFAW EIRDRLLSEG VNQDDIPSVS SINRVLRNLA SETQK
Cr_Nr002    IGGSKPRVAT PEVVNKIASY KRECPSIFAW EIRDRLLNEG INNDNIPSVS SINRVLRNLN GDHQT
Pm_nr002    IGGSKPRVAT PQVVNKIAMY KRECPSIFAW EIRDRLLNEA VNAENIPSVS SINRVLRNLN GENGR
Hau_pax6    IGGSKPRVAT MSVVQRIAHY KKECPSIFAW EIRDRLLSEE ISQENIPSVS SINRVLRNLS SNNAK
MmPAx4      IGGSKPRLAT PAVVARIAQL KDEYPALFAW EIQHQLCTEG LTQDKAPSVS SINRVLRALQ EDQSL
BsPax258B   IGGSKPKVAT PPVVVKITEY KRENPTMFAW EIRDRLLSEG VTPDSVPSVS SINRIVRSKT SDFMK
Hr_Nr002_   IGGSKPKVAT PYVVSKITEY KRENPTMFAW EIRDRLLAEN ISSDTVPSVS SINRIVRSKT AEFMK
Mo_Nr003_   IGGSKPKVAT PHVVGKITEY KRENPTMFAW EIRDRLLSDN IINDHVLNIQ NIGHPTESKH SSLER
Pm_nr005    ---------- ---------- ---------- ---------- ---------- ---------- -----
Cr_Nr003    IGGSKPKVAT HDVVMRITEY KRENPTMFAW EIRDRLLADE VSQETVPSVS SINRIVRSKT SEFMK
Hau_37A     IGGSKPRVKM EDVEKKIEEY KKEDPSIFSW EIRERLIKEG IDENSVPSVS SISRLQKQTR TPPNN
Cw_PAxA_    IGGSKPKVAT PEVVAIISEY KRRNPSMFAW EIRDRLLMDG MPSNKVPSAS TINRILRNKV TETNK
Hau_37B_    IGVGKHRSSN KPMVNNDDNN NEDDECKLKV RINNKNNNNN INNKRNININ NNSVTCRKQN LNSNN
```

Zic PhyML,LG, 100 Bootstrap



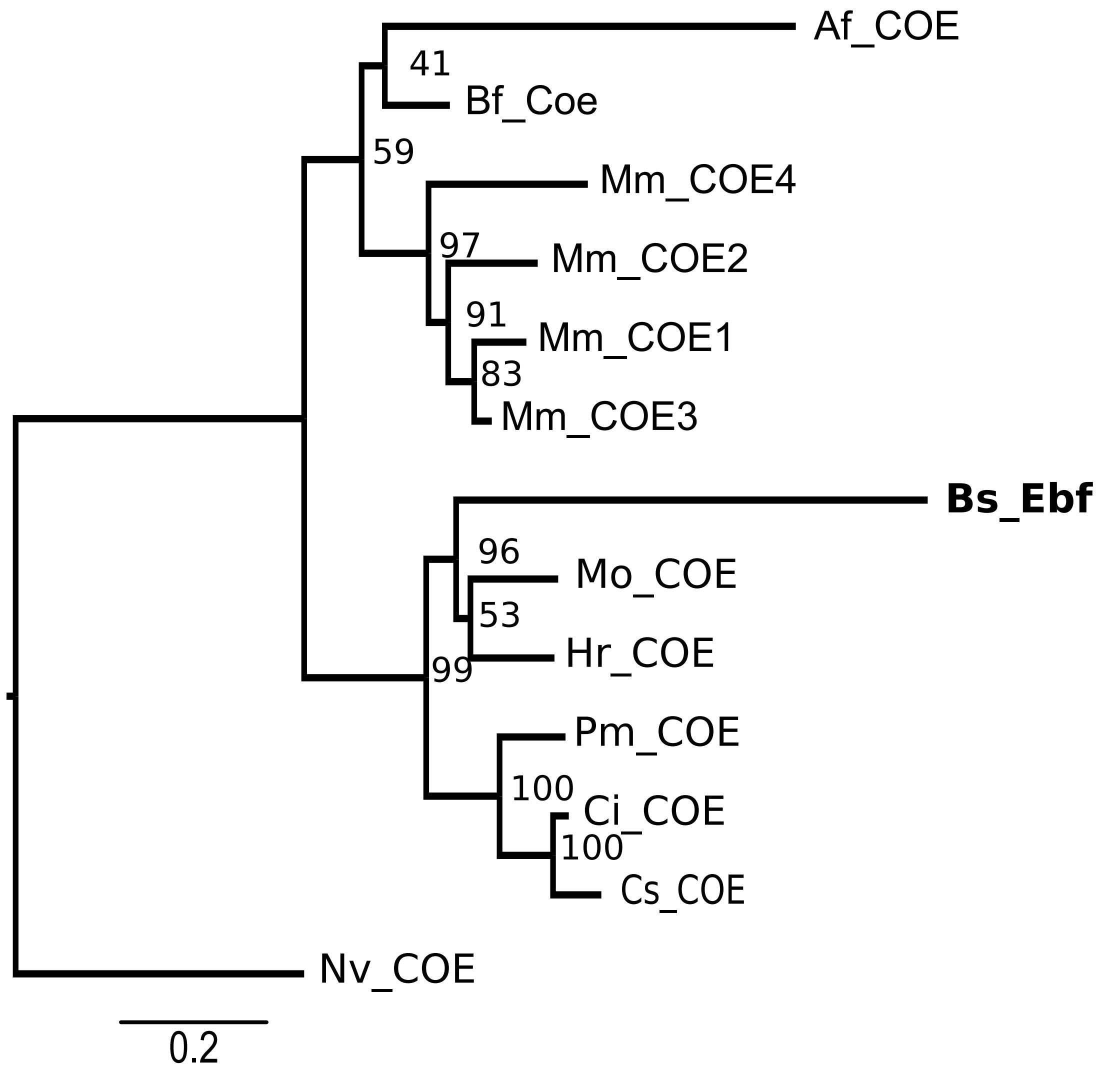
Zic

```
Bs_Macho   QAYPTF---- --QQQYYDGR YTMLSEDFVP VTSDEVIYD- -----SSSPF HSASHNCNYH NCADIYPYVT
Hr_Macho   STYVTTYQSP YEPHATFQSG RGFVASSEDS YQDNGYV--S AQFDSNGKPP TALNFHSFHP NKNDFYTFIH
HrZicN     MAYATAYAFS HQPTAPFPGY TNVLPQPMHG ITSPGSQ--A ALYNPNVAPQ LKTDSQFFTH QPNQYLAYGQ
Mt_ZicL    MAYATANTDA FNCYGPLEGK QTIVAVSERT VAAPFRNVEM TDQQQSYMP- ---------- --QNAISHIH
Cs_MAcho   MAFVGTSSAG YQPHHPYEAK DEVLQVEQQA LSTTQSVWSP SQYKYGAWSA PSLNFNLPVH HPQAFFAYSQ
Pm_macho   MAFATTNNSA FPKTHHYDSK EEVFVFRQSG DASPGSAWPP AQYKYGDLPP SPINFNLPVH NLQSFFTYTQ
AmphiZic   SGYPTATSGN FHGHGHADSS -HLMAHHVNG LSLPGDGYG- --FNQVSSPM APMNM----H GPAAFFRYMR
Mm_Zic3    ---------- --RVAPRGSP GSVQISFACT LAHFGRIFC- -----NTLPY SPMNMNMGVH GPGAFFRYMR
Mm_zin1    GSYSSASAGG FGPHGHTDAA GHLLVNGQMR LGFSGDMYPP EQYGQVTSPY GPMNVNMAAH GAGAFFRYMR
Mm_Zic2    GSYSGPGAGG L-HHAHSDAQ GHLLLNGQMR LGLPGEVFGS EQYRQVASPY GPMNMNMGMH HPGAFFRYMR
Mm_ZIC4    MKYKTSSSCG HHPQLAASSN -PSVLNGLLR LGIPGDMYAS EPFAPGPMAY GGMNLTMNLH GPGAFFRYMR
Mm_zic5    SGYTATASGT YA---GRDGG GSALLNGQMR LGLAAELYGE PPFAPGAVAY GAVNLNLNLH HPGAFLRYMR
Cs_ZicLb   ------YSAL SDPYAAYDPR ASFLQAHQHS IDSKPMQLNS AAYCAGMIPH FPQSLHLATH NP-------R
Cs_ZicL    ------YSAL SDPYAAYDPR ASFLQAHQHS IDSKPMQLNS AAYYAGMIPH FPQSLHLATH NP-------R
Pm_ZICL    SSYVPGSSII ADPYAAYDPR ASFLPQQQPH ASFPQVAFHT SSYLPGVFPQ IPVHYQHPT- -P-------P
Mt_Macho   YDYAEQYCAP FQATQQYDGY RSVFVHSENR YNSPDYYYGQ AQYQTFGFPQ NPPNWSFNYP SALEYYSYVN
BsZicL     MSYALGEEAG KQPHQIRHPM HQVGAQVQYA QNW------P AVYHQNMMP- ---------- --PQRFGHLQ
Mm_GLI     NSYSEPHQAN FMGSQGYGAA RGSLAFINSR CTSPGGSYGP PQMSHGLQPQ LKSELDMMVS SPDHLLGMLR

           71
Bs_Macho   PQVYEEFVCK WRSDCKSSGQ TCNAMLWTQL DLVNHINTNH VGGPEQADHS CYWENCSRRR KTFKAKYKLV
Hr_Macho   QQLQSDYVCK WMYPANKQQE PCNRVFVTLL DLVTHISLNH IGGSEQIDHT CYWQNCSRGK KSFKAKYKLV
HrZicN     AQ-QKGLTCK WIVSGKSGEI TCDREFYSMN QLVDHVTVDH VGGHDQADHT CYWKDCTRE- KSFQAKYKLV
Mt_ZicL    PQINKVLCCK WISKDKANTE PCNKTFYNIA DIVDHLAVDH VGGHDQTDHS CYWKDCSREC KPFQAKYKLV
Cs_MAcho   AEVRVELACK WTKPENQQVE PCNLVFHSML DLVTHVGRDH VGGPEHTDHA CYWQDCSREN KPFKAKYKLI
Pm_macho   AELHQELVCK WTKPVHLHVE PCNLVFHNMV ELVAHVSRDH VGGPEQADHT CYWESCSRKD KPFKAKYKLI
AmphiZic   QPIKQEAVCL WIDPDQ-PEK PCNKSFSTMH EIVTHVTVEH VGGPECTNHA CFWKDCPRDG RAFKAKYKLV
Mm_Zic3    QPIKQELSCK WIEEAQLSRK SCDRTFSTMH ELVTHVTMEH VGGPEQNNHV CYWEECPREG KSFKAKYKLV
Mm_zin1    QPIKQELICK WIEPEQLANK SCNKTFSTMH ELVTHVTVEH VGGPEQSNHI CFWEECPREG KPFKAKYKLV
Mm_Zic2    QQIKQELICK WIDPEQLSNK SCNKTFSTMH ELVTHVSVEH VGGPEQSNHV CFWEECPREG KPFKAKYKLV
Mm_ZIC4    QPIKQELICK WLGDDSPMSR PCSKTFSTMH ELVTHVTVEH VGGPEQANHI CFWEECPRQG KPFKAKYKLV
Mm_zic5    QPIKRELICK WLDPEELAGK PCSKTFGTMH ELVNHVTVEH VGGPEQSSHV CFWEDCPREG KPFKAKYKLI
Cs_ZicLb   QPVSHPVACK WMNPKHGG-R SCDVIFHDMH DLVNHVTRDH VGGMDQTDHT CYWEDCSRKR KCFKAKYKLV
Cs_ZicL    QPVSHPVACK WMNPKHGE-R SCDVIFHDMH DLVNHVTRDH VGGMDQTDHT CYWEDCSRKR KCFKAKYKLV
Pm_ZICL    IPSQEPVTCQ WKSPNQNGGL VCGVVFHDLN DLVGHVTRDH VGGIEQTDHT CYWGSCSRTN KAFKAKYKLV
Mt_Macho   AECRGEFVCK WTLLQNGEYH ICNRVFGNIM DLVSHVSICH IGGPEQVDHT CYWENCQREG KAFKAKYKLV
BsZicL     APVIQPLVCL WV-PSKDREA PCKHTFGCIR DLVEHISNEH VRGHDHTDHS CYWKTCTRE- KGFHAKYKLV
Mm_GLI     EDLVYETDCR W--------D GCSQEFDSQE QLVHHINSEH IHG-ERKESV CHWGGCSREL RPFKAQYMLV

          141
Bs_Macho   NHMRVHTGEK PFACPFPNCG KMFARSENLK IHKRIHTGER PFGCPHPGCD RRFANSSDRK KHSHVHTSDK
Hr_Macho   NHTRVHTGER PFSCPFPGCG KMFARSENLK IHKRIHTGER PFTCTFPGCD RRFANSSDRK KHSHVHTSDK
HrZicN     NHIRVHTGEK PFICLFPNCG KVFARSENLK IHKRTHTGEK PFVCPFDGCD RRFANSSDRK KHTYTHSTSK
Mt_ZicL    NHIRVHTGEK PFICLYPGCG KVFARSENLK IHKRIHTGEK PFVCPFPGCQ RRFGNSSDRK KHTYTHRTQK
Cs_MAcho   NHIRVHTGEK PFQCPYPGCG KVFARSENLK IHKRTHTGEK PFCCEFSGCN RRFANSSDRK KHTHVHTTDK
Pm_macho   NHIRVHTGEK PFLCPYPGCG KVFARSENLK IHKRTHTGEK PFCCDFKGCN RRFANSSDRK KHTHVHTTDK
AmphiZic   NHIRVHTGEK PFPCPFPGCG KLFARSENLK IHKRTHTGEK PFKCEFEGCD RRFANSSDRK KHSHVHTSDK
Mm_Zic3    NHIRVHTGEK PFPCPFPGCG KIFARSENLK IHKRTHTGEK PFKCEFEGCD RRFANSSDRK KHMHVHTSDK
Mm_zin1    NHIRVHTGEK PFPCPFPGCG KVFARSENLK IHKRTHTGEK PFKCEFEGCD RRFANSSDRK KHMHVHTSDK
Mm_Zic2    NHIRVHTGEK PFPCPFPGCG KVFARSENLK IHKRTHTGEK PFQCEFEGCD RRFANSSDRK KHMHVHTSDK
Mm_ZIC4    NHIRVHTGEK PFPCPFPGCG KVFARSENLK IHKRTHTGEK PFRCEFEGCE RRFANSSDRK KHSHVHTSDK
Mm_zic5    NHIRVHTGEK PFPCPFPGCG KVFARSENLK IHKRTHTGEK PFKCEFDGCD RKFANSSDRK KHSHVHTSDK
Cs_ZicLb   NHIRVHTGEK PFLCPYPDCG KMFGRSENLK IHQRTHTGER PFPCKFPGCE RRFANSSDRK KHSYMHNTEK
Cs_ZicL    NHIRVHTGEK PFLCPYPDCG KMFGRSENLK IHQRTHTGER PFPCKFPGCE RRFANSSDRK KHSYMHNTEK
Pm_ZICL    NHLRVHTGEK PFVCNFPGCG KTFGRSENLK IHKRTHTGEK PFPCKFPGCD RRFGNSSDRK KHSYMHNTEK
Mt_Macho   NHLRVHTGEK PFSCPYPNCG KVFARSENLK IHKRIHTGER PFTCTYPGCD RKFANSSDRK KHSHVHTSDK
BsZicL     NHIRIHTGER PFICTNKTCG KRFARSENLK IHLRIHSGKK PFGCPAKGCE RFFANSSDRK KHSYTHLNTK
Mm_GLI     VHMRRHTGEK PHKCTFEGCR KSYSRLENLK THLRSHTGEK PYMCEQEGCS KAFSNASDRA KHNRTHSNEK
```

```
           211
Bs_Macho   PYTCKVGCDK TYTHPSSLRK HMRLHEEKGD --GTSSPDSE LSSSKSGPDT TVTCSQPLCY PASPSDDWY
Hr_Macho   PYTCKIGCDK NYTHPSSLRK HMRLHEAQGD ECKIDASNIP SSIKR----- SVSPGNLTNF NAYSANDWY
HrZicN     PYACKVGCKK SYTHPSSLRK HLKMHEAEGI GGSDSRMTSP TSSSSSNNAE GVSPKSEINS FFRTNNPFP
Mt_ZicL    PYICPVGCGK TYIHPSSMRK HVKSHEEFGR MSSGSSSESP QN-------- -------PTSM MTQNISQMY
Cs_MAcho   PYLCKVGCEK SYTHPSSLRK HMKVHESQGD AISLSGISAP DSISRGEDAS SLSPPAPCSV SHAAQNNWQ
Pm_macho   PYLCKIGCDK SYTHPSSLRK HMKLHESQGD DVQPSEPHSP DTVTQNINTS CVTPSSPASA NVSPQNSWQ
AmphiZic   PYNCKVGCDK SYTHPSSLRK HMKVHGKTSP SFSPSSTSSE THLPRITQHP SATLAP-TPT HATNLSEWY
Mm_Zic3    PYICKV-CDK SYTHPSSLRK HMKVHESQGS ASGYESSTPP AIASANSKDT --TPSATSTS HPPNFNEWY
Mm_zin1    PYLCKM-CDK SYTHPSSLRK HMKVHESQGS ASGYESSTPP TIVSPTTDNP SMSPSS-SAV HSSNFNEWY
Mm_Zic2    PYLCKM-CDK SYTHPSSLRK HMKVHESQGS ASGYESSTPP GLVSPSAEPQ NLSPAAVSAV HSSNFNEWY
Mm_ZIC4    PYMCKVGCDK CYTHPSSLRK HMKVHGR--- PSGYDSAITS ALASPSLESG SVACSAAVVV RGTDVSE--
Mm_zic5    PYYCKIGCDK SYTHPSSLRK HMKIHCK-SP ALGYSSVGTP VGLSPVLDPT SSTLSP---- -VTNLNEWY
Cs_ZicLb   LYTCKYGCDR SYTHPSSLRK HIRMHESNGD -------TSP TTSSDFASSP PMTSQPP--- ---------
Cs_ZicL    LYTCKYGCDR SYTHPSSLRK HIRMHESNGD -------QSP TTVTDCNPLR RMTSSPPMFL F----HEWN
Pm_ZICL    LYTCHYDCDR SYTHPSSLRK HIRMHESNGD STPHSSCSSP AS-SQGQAEP --AHQPPTHN WVTSSSGLQ
Mt_Macho   PYNCKVGCGK SYTHPSSLRK HMRLHETQGD NNSLSSLTNP DDYSDKMSSP RSSEMQPKPV FVEKQNDWY
BsZicL     PYSCK-RCGK KYSHPSSLRK HEKSHNSKQQ TDCDSSGQSP VSADESLPMT PISQQPDCGI VGQQSENLP
Mm_GLI     PYVCKLGCTK RYTDPSSLRK HVKTHRGDGP AQSCSSDHSP AGSTEDLSSL SSSSMSPTSV YPDPPENWG
```

Ebf / COE

PhyML, LG, 100 bootstrap

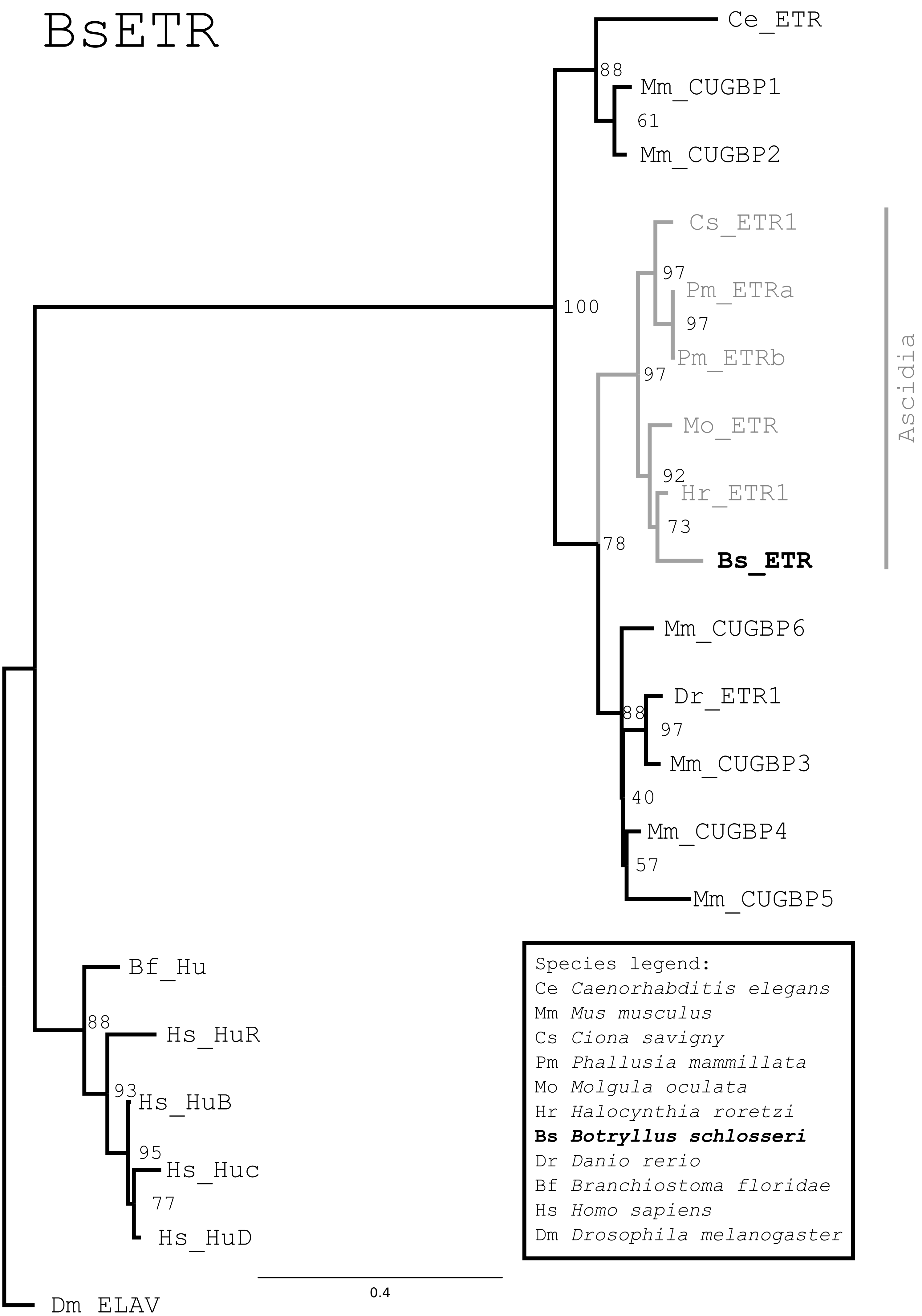


Af *Apis florea*
Bf Branchiostoma floridae
Bs *Botryllus schlosseri*
Ci *Ciona intestinalis*
Cs *Ciona savigny*
Hr *Halocynthia roretzi*
Mm *Mus musculus*
Mo *Molgula oculata*
Pm *Phallusia mammilata*

Ebf

```
Bs_Ebf    TSALRDQPML PITGWMQTNL VEQPGILRIH RAEFSIQPPK NLRKSNFFNF SLLFYDNRGL PARATKTEFK
Ci_COE    MATMAGPQLP AVRGWMQTTL VEPNGNVGLH RAHFEKQPPN NLRKSNFFHF VLALYDRQGQ PVEIERSAFV
Cs_COE    --NMDGPQLS AVRGWMQTSL IEASGNVGLH RAHFEKQPPN NLRKSNFFHF VLALYDRQGQ PVEIERSAFV
Pm_COE    --TMPGQQLP AVRGWMQTSL VESSSSVSLH RAHFEKQPPN NLRKSNFFHF VLALYDRQGQ PVEVERSAFA
M0_COE    MSNLPSQQLA PLRGWVQTGL VDTMPSAGLH RAHFEKQPPG NLRKSNFFHF VLALYDRQGQ PVEVERTAFI
Hr_COE    MSNLSSQQLA PVRGWIQTSL VDALPSVGLH RAHFEKQPPS NLRKSNFFHF VLALYDRQGQ PVEIERTAFI
Mm_COE1   QESILGSGMN AVRTWMQGAV LDAQSGVGLA RAHFEKQPPS NLRKSNFFHF VLALYDRQGQ PVEIERTAFV
Mm_COE3   QENILGSGMN PVRSWMHTAV VDAQSGVGLA RAHFEKQPPS NLRKSNFFHF VLALYDRQGQ PVEIERTAFV
Mm_COE2   QDTLLGAEMD SVRSWVRNVV VDAQSGVALS RAHFEKQPPS NLRKSNFFHF VLALYDRQGQ PVEIERTAFV
Mm_COE4   QDALLPSSLG SVRSWMQSAI LDAQSGVGLA RAHFEKQPPS NLRKSNFFHF VLAMYDRQGQ PVEVERTAFI
Bf_Coe    MDTLLGSQIG SVRNWMQPSI VDTSGSVGLA RAHFEKQPPS NLRKSNFFHF VIALYDRQGQ PVEIERTAFI
Af_COE    MRQI------ VPRAWPQPAL AD-NGTVGVG RAHFEKQPPS NLRKSNFFHF VIALYDRHGQ PIEIERTSFM
Nv_COE    QEQD------ PVVGCLQHRV LESTASGLLD RCHFEKQPPA NLRKSNFFHF VLAFFDHNGQ AVEIERASFI

          71
Bs_Ebf    NFVEEASESG TEKANNGVHY EITFILSNGV SVLQDIYVRL VDAANRERIC YEGQDKNPEM RRVLLTHEIM
Ci_COE    GFVENETEIA GEKTNNGIQY RLQLLYHSGV RTEQDVFVRL IDSATKQSIT YEGQDKNPEM RRVLLTHEIM
Cs_COE    GFIENETEIA GENTNNGIQY RLQLLYQNGL RTEQDVYVRL IDSATKQSIT YEGQDKNPEM RRVLLTHEIM
Pm_COE    GFVENETEVA GEKTNNGIHY RLQLLYHNGI RTEQDVFVRL IDSATKQSIV YEGQDKNPEM RRVLLTHEIM
M0_COE    DFVEGDREAT GEKTHNGIHY RLQLLYHNGV RTEQDLFVRL IDSVTKQAIS YEGQDKNPEM RRVLLTHEIM
Hr_COE    DFVENET--- ---------- ---------- ---------- -------AIC YEGQDKNPEM RRVLLTHEIM
Mm_COE1   GFVEKEKEAN SEKTNNGIHY RLQLLYSNGI RTEQDFYVRL IDSMTKQAIV YEGQDKNPEM CRVLLTHEIM
Mm_COE3   DFVEKEKEPN NEKTNNGIHY KLQLLYSNGV RTEQDLYVRL IDSMTKQAIV YEGQDKNPEM CRVLLTHEIM
Mm_COE2   DFVENDKEQG NEKTNNGTHY KLQLLYSNGV RTEQDLYVRL IDSVTKQPIA YEGQNKNPEM CRVLLTHEVM
Mm_COE4   DFVEKDREPG TEKTNNGIHY RLRLVYNNGL RTEQDLYVRL IDSMSKQAII YEGQDKNPEM CRVLLTHEIM
Bf_Coe    DFVEKDRETA GENTRNGIHY RLQLLYSNGV RTEQDLYVRL IDSMTKQAIV YEGQDKNPEM QRVLLTHEIM
Af_COE    GFVEKDQESE GQKTNNGIQY SLHLLYSNGY ECNRDIFYRC IKNVSRQAIM YEGQDKNPEM CRVLLTHEVM
Nv_COE    DFIEHLPE-- DKNVRNGALY RVLLLYSNGI RTEQDIFVRL IDAATKQIVP YEGQDKNPEM RRILLTHEVM

         141
Bs_Ebf    CSRCCDHKSC GNRNETPSDP T-FDKYDIKF FLKCNQNCLK NAGNPRDMRR FAVAISTDVD MK-FPLAYSE
Ci_COE    CSRCCDKKSC GNRNETPSDP VVIDRYFLKF FLKCNQNCLK NAGNPRDMRR FQVVVSTTVH VDGHVLAVSD
Cs_COE    CSRCCDKKSC GNRNETPSDP VVIDRYFLKF FLKCNQNCLK NAGNPRDMRR FQVVVSTTVH VDGHVLAVSD
Pm_COE    CSRCCDKKSC GNRNETPSDP VVIDRYFLKF FLKCNQNCLK NAGNPRDMRR FQVVVSTTVH VDGHVLAVSD
M0_COE    CSRCCEKKSC GNRNETPSDP VVIDRYFLKF FLKCNQNCLK NAGNPRDMRR FQVVISTTVH VDGHVLAVSD
Hr_COE    CSRCCEKKSC GNRNETPSDP VIIDRYFLKF FLKCNQNCLK NAGNPRDMRR FQVVVCTTVQ VDGHVLAVSD
Mm_COE1   CSRCCDKKSC GNRNETPSDP VIIDRFFLKF FLKCNQNCLK NAGNPRDMRR FQVVVSTTVN VDGHVLAVSD
Mm_COE3   CSRCCDKKSC GNRNETPSDP VIIDRFFLKF FLKCNQNCLK NAGNPRDMRR FQVVVSTTVN VDGHVLAVSD
Mm_COE2   CSRCCEKKSC GNRNETPSDP VIIDRFFLKF FLKCNQNCLK TAGNPRDMRR FQVVLSTTVN VDGHVLAVSD
Mm_COE4   CSRCCDRKSC GNRNETPSDP VIIDRFFLKF FLKCNQNCLK NAGNPRDMRR FQVVVSTTVS VDGHVLAVSD
Bf_Coe    CSRCCDKKSC GNRNETPSDP VIIDRFFLKA FLKCNQNCLK NAGNPRDMRR FQVVIATTVN VDGHVLAVSD
Af_COE    CSRCCDKKSC GNRNETPSDP VIIDRFFLKF FLKCNQNCLK NAGNPRDMRR FQVVISTQVG VEGPLLAVSD
Nv_COE    CSRCCEKKSC GNKNETPSDP VVVDRYCLKF FMKCNQNCLK NAGNPKDMRR FQVHVSTSVD PMYDMIACSD

         211
Bs_Ebf    TMFVHNNSKH GRRAKDLENG KSGPCITSVC PSEGWVTGGT TVIIVGNNFT RDMQVAFDTT IVFTEFVSEH
Ci_COE    NMFVHNNSKH GRRARRVDPS EASPTIKAIN PAEGWTTGGA TVVIVGENFF DGLQVVFGSM VVWSELITQH
Cs_COE    NMFVHNNSKH GRRARRVDPS EASPMIKAIN PAEGWTTGGA TVVIVGDNFF DGLQVVFGSM VVWSELITQH
Pm_COE    NMFVHNNSKH GRRARRVDPA EATPAIKAIN PSEGWTTGGA TVIIVGDNFF DGLQVIFGSM VVWSELITQH
M0_COE    NMFVHNNSKH GRRARRLEPT DATPVIKALS PSEGWTTGGA TVIIVGDNFF DGLQVVFGSM IVWSELVTQH
Hr_COE    NMFVHNNSKH GRRAKRIEPA EATPSIKAMS PSEGWTTGGA TVIIVGDNFF DGIQVVFGSM IVWSELITQH
Mm_COE1   NMFVHNNSKH GRRARRLDPS EATPCIKAIS PSEGWTTGGA TVIIIGDNFF DGLQVIFGTM LVWSELITPH
Mm_COE3   NMFVHNNSKH GRRARRLDPS EATPCIKAIS PSEGWTTGGA TVIIIGDNFF DGLQVVFGTM LVWSELITPH
Mm_COE2   NMFVHNNSKH GRRARRLDPS EATPCIKAIS PSEGWTTGGA MVIIIGDNFF DGLQVVFGTM LVWSELITPH
Mm_COE4   NMFVHNNSKH GRRARRLDPS EATPCIKAIS PGEGWTTGGA TVIIIGDNFF DGLQVVFGNV LLWSELITPH
Bf_Coe    NMFVHNNSKH GRRARRLDPS EATPCIKAIS PSEGWTTGGA TVIIIGDNFF DGLQVVFGTM LVWSELITPH
Af_COE    NMFVHNNSKH GRRTKRLDPS DQTPCIKAIS PNEGWTSGGS TVIIIGENFF DGLQVVFGSM LVWSELITPN
Nv_COE    NMFVHNNSKH GRRTRRADGN EDDPCIKAIC PNEGWTIGGS NVILIGDNFF DGLQVVFGSF IVWSEFITPH

         281
Bs_Ebf    ALKIQSPGYH LTGPVEVTIT IKNKQYCKNQ PGRFTYTALN EPSIEYGFQR LMKMIPRHQA DPLQISKEMI
Ci_COE    AIRVQTPPRH LPGVVEVTLS YKNKQFCSGA PGRFVYTALN EPTLDYGFQR LLKTVPRHPG DPERLPKEII
Cs_COE    AIRVQTPPRH LPGVVEVTLS YKNKQFCSGA PGRFVYTALN EPTLDYGFQR LLKTIPRHPG DPERLPKEII
Pm_COE    AIRVQTPPRH LPGVVEVTLS YKNKQFCGAA PGRFVYTALN EPTLDYGFQR LLKTVPRHPG DPERLPKEII
M0_COE    AIRVQTPPRH VPGVVEVTLS YKNKQFCKGS PGRFVYTALN EPTIDYGFQR LLKAIPRHPG DPERLPKEIV
Hr_COE    AIRVQTPPRH LPGVVEVTLS YKNKQFCKGT PGRFVYTALN EPTIDYGFQR LLKCIPRHPG DPERLPKEIV
Mm_COE1   AIRVQTPPRH IPGVVEVTLS YKSKQFCKGT PGRFIYTALN EPTIDYGFQR LQKVIPRHPG DPERLPKEVI
Mm_COE3   AIRVQTPPRH IPGVVEVTLS YKSKQFCKGA PGRFVYTALN EPTIDYGFQR LQKVIPRHPG DPERLPKEVL
Mm_COE2   AIRVQTPPRH IPGVVEVTLS YKSKQFCKGA PGRFIYTALN EPTIDYGFQR LQKVIPRHPG DPERLAKEML
Mm_COE4   AIRVQTPPRH IPGVVEVTLS YKSKQFCKGA PGRFVYTALN EPTIDYGFQR LQKVIPRHPG DPERLPKEVL
Bf_Coe    AIRVQTPPRH IPGVVEVTLS YKSKQFCKGA PGRFVYTSLN EPTIDYGFQR LAKLVPRHPG DPERLPKEII
Af_COE    AIRVQTPPRQ IPGVVEVTLS YKTKQFCKGA PGRFVYVSLN EPTIDYGFQR LQKLIPRHPG DPEKLPKEII
Nv_COE    ALRVQAPPSP MPGVVKVYLM HKEKQYCKYA PAKFGYTALV EPTIDYGFQR LSKLIPRHPG DPERIPKEIV
```

```
          351
Bs_Ebf    LTRAAELLEA FSVRPYSQLQ SAPPPPLQNS FTSTSSTMMA SAMGGYNNQY FAAQDRMD-- ----------
Ci_COE    LKRAADVMEA VISRPPSQML AAAPCAPPGS FPQSASAAMA VAMNGYASQF GGTPDRFDGS DSGYSRGSVS
Cs_COE    LKRAADVMEA VISRSPSQML GATPCPPSGS FPQSASAAMA VAMNGYASQF VGNPDRFDGS DSGYSRGSVS
Pm_COE    LKRAADLMEA FQYKPHGQML APPPTGNHGS FPQASVAMAA AAMNGYASGF VPATERFDGS DSGYSRGSVS
M0_COE    LKRAADVMEA VMTRPYNQVP APPPAPMHNT FNGSSPAMMA AGVNTYNSQY ITTPDRLDGS DSGYSRGSVS
Hr_COE    LKRAADVMEA FMSRPYNQVT PAPSAPLQGS YNGSTSAMAA AS--GYHNQY ISTPDRLDGS DSGYSRGSVS
Mm_COE1   LKRAADLVEA LYGMPHNNQE ALYSVPRNPA LANTSVHAGM MGVNSFSGQL VNVSEASQAT NQGFTRNSVS
Mm_COE3   LKRAADLVEA LYGMPHNNQE ALYSVPRNPT LGNTPAHTGM MGVNSFSSQL VNVSETSQAN DQGYSRNSVS
Mm_COE2   LKRAADLVEA LYGTPHNNQE ALYSVPRNPA LSSSPAHSGM MGINSYGSQL VSISESTQGN NQGYIRNSIS
Mm_COE4   LKRAADLAEA LYGVPSSNQE ALYSAPRAGP LAPSHPHPAV VGINAFSSPL IAVGDTTP-- EPGYARSSAS
Bf_Coe    LKRAADLAEA IYSMPRNPNL PALTGPRSPA VNNSVGAGGM MGMNSFGNQL VSVPDNSSQT NHGYSRQSVS
Af_COE    LKRAADLAEA LYSMPRSGNA GITGAPRSGH PPAPPTSSSA TAFNSYTGQL VTVQENGSSV TTGGGGGSVG
Nv_COE    LKRAADLAET LYQMPRTPTS TGLATPKSGT LPDSMTNVFT QSITPDNSSH DNASDQTDNV VQSTSNATLS

          421
Bs_Ebf    --GYSPHAGP QHSSAPPYGS PVNGYVGPTF TNMTNSSANM FCNAGLLPQS PNPAMNCLGG TTPGIFSFSP
Ci_COE    PRGYSPQTTP HSLNPAPYAP SFNSYSSPTF TNMNNTSPGL FSGSGIIPPS PHNGMNPLSG TTPGIFSFSP
Cs_COE    PRGYSPQTTP HSLNPAPYPP SFNSYSTPTF TNMNNSSPGI FPGSSIIPPH THNGMNGLSG TTPGIFSFSP
Pm_COE    PRGYSPQGTP NSVTPSAYTQ PFNSYTCPTF TNMTNPSTAL FSGSSILPSS PPNPMNPLSS TTPGIFSFSP
M0_COE    PKGYSPQATP HSTAAPPYCA AMNGYSNPTF TNMTNSSSNM FAGTGLFPTS PNGGMNALTG TTPGIFSFSP
Hr_COE    PKGYSPQSTT HNVLVSHYGT TMNAYSCPTF TNMTNTASNM FCNTGLLPPS PNSAMNGLAG PTPGIFSFSP
Mm_COE1   PHGYVPSTTP QQTNYNSVTT SMNGYGSAAM SNLG--GSPT FLNGAIVPSS PTASSTSLSS SSSGIFSFSP
Mm_COE3   PRGYVPSSTP QQSNYNTVST SMNGYGSGAM ANLGVPGSPG FLNGGIVPSS PTASSVTLSS STHGIFSFSP
Mm_COE2   PRGYSSSSTP QQSNYSTSSN SMNGYSNVPM ANLGVPGSPG FLNGGIMSSS PTS------S STSSILPFS-
Mm_COE4   PR-FAPSPGS QQSSYGSGGA GLGSYGAPGV TGLGVPGSPS FLNGAIMPSS PPASSMSLAA PTTSVFSFSP
Bf_Coe    PRGYG-ASTP HSTNYSTSAS GLNGYGSGNL SNMPVPSSPG FLNGSTIPSS PTPTPSSLTA STPGIFFFSP
Af_COE    GSGGSADAYR QSSSTPH--- -------SHS TNGSYSVANP YTGSPTLYTS PHEQY----- ---GIFYTS-
Nv_COE    GVNMSHHNGV MASS------ SCYGNGSPAL SCMAVPASPG YFTGAFIPQS PSPTPNSVPS SNNSIFSF-P

          491
Bs_Ebf    ANMISAVKQK SAFAPVIRAH NTPSPNHSGV SETTINGYT
Ci_COE    ANMISAAKQK SAFAPVHRPH NSPSPLAPSN GNIALNGYS
Cs_COE    ASMISAAKQK SAFAPVHRPH NSPSPMAPPN SSLVENSES
Pm_COE    ANMISAAKQK SAFAPVIRPH NSPSPLASSN GNLIMNGYS
M0_COE    ANMISAVKQK SAFAPVVRNH NSPSPGPAGA VDQVMNTYT
Hr_COE    ANMISAVKQK SAFAPVIRAH NSPSSSSGTP GPGILQGTY
Mm_COE1   ANMVSAVKQK SAFAPVVRPQ TSPPPTCTST NGNSLQAIS
Mm_COE3   ANVISAVKQK SAFAPVVRPQ ASPPPSCTSA NGNGLQAMS
Mm_COE2   SSVFPAVKQK SAFAPVIRPQ GSPSPACSSG NGNGFRAMT
Mm_COE4   VNMICAVKQR SAFAPVLRPP SSPSQACPRA HREGLRGLQ
Bf_Coe    ANMISAVKQK SAFAPVVRPQ SSPSPAGASS NGTGLQGGW
Af_COE    -------GDG SAFAPVVRPP TTSVPPHWST QHH--TAAQ
Nv_COE    PNMIQAVKQK SAFNPVSRPE STAHPRSPAV VQGVMSTYN
```



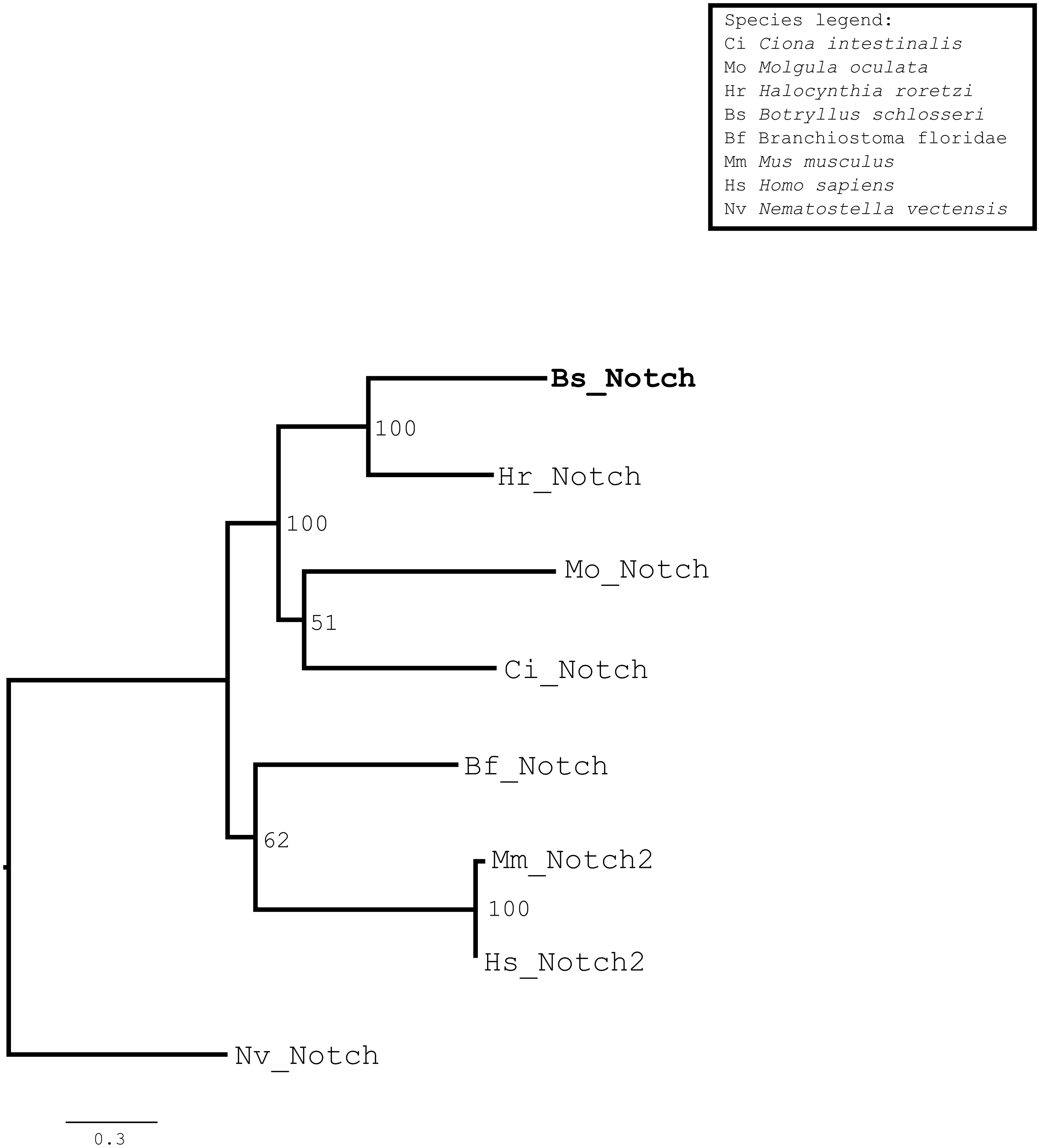
BsETR
Ce_ETR
88
Mm_CUGBP1
61
Mm_CUGBP2
Cs_ETR1
97
Pm_ETRa
100
97
Pm_ETRb
97
Mo_ETR
92
Hr_ETR1
73
78
Bs_ETR
Ascidia
Mm_CUGBP6
Dr_ETR1
88
97
Mm_CUGBP3
40
Mm_CUGBP4
57
Mm_CUGBP5
Bf_Hu
88
Hs_HuR
93
Hs_HuB
95
Hs_Huc
77
Hs_HuD
Dm_ELAV
0.4
Species legend:
Ce Caenorhabditis elegans
Mm Mus musculus
Cs Ciona savigny
Pm Phallusia mammillata
Mo Molgula oculata
Hr Halocynthia roretzi
Bs Botryllus schlosseri
Dr Danio rerio
Bf Branchiostoma floridae
Hs Homo sapiens
Dm Drosophila melanogaster

ETR

```
BsETR      MEVPKDEDAI KLFIGQVPKQ WTEMDLRPIF EPFGEIYELS VLHDKFTRMH KGCAFLTYCK KSSAIQAQDY
HrETR1     MEFDKDDDAV KLFIGQVPKN WTEHELRPIF EPYGEIYELS VLHDKYTGMH KGCAFLTYCK KTPAINAQNF
MoETR      MEMDKEEDAI KLFIGQVPKH WDEKDLRPIF EVYGEIYELS ILHDKYTGMH KGCAFLTYCK KSSALQAQNC
Cs-ETR1    METEKDDDAV KLFIGQVPKN WEEKELRQVF APFGEIFELS VLRDKYTGMH KGCAFLTYTQ KTSAMNAQNS
Pm-ETRa    -DPEKEDDAI KLFIGQVPKN WEEKELRQLF SSFGEIYELS VLRDKYTGMH KGCAFLTFVK KIAALNAQNA
DrETR1     M---KEADAI KLFIGQIPRN LEEKDLKPIF EQFGKIYELT VIKDKYTGMH KGCAFLTYCA RESALKAQNA
BfHu       MTGSTKDSKT NLIVNYLPQT MTQDEIKSLF SSIGEVESCK LIRDKVTGQS LGYGFVNYVK PQDAEKAINT
HsHuB      MSGNTEDSKT NLIVNYLPQN MTQEELKSLF GSIGEIESCK LVRDKITGQS LGYGFVNYID PKDAEKAINT
HsHuD-     MTGTTDDSKT NLIVNYLPQN MTQEEFRSLF GSIGEIESCK LVRDKITGQS LGYGFVNYID PKDAEKAINT
HsHuc      MTNATDDSKT NLIVNYLPQN MTQDEFKSLF GSIGDIESCK LVRDKITGRD LGYGFVNYPD PNDADKAINT
HsHuR      MEDRGDIGRT NLIVNYLPQN MTQDELRSLF SSIGEVESAK LIRDKVAGHS LGYGFVNYVT AKDAERAINT
DmELAV     MNGT--ETRT NLIVNYLPQT MTEDEIRSLF SSVGEIESVK LIRDKSKGQS LGYGFVNYVR PQDAEQAVNV

           71
BsETR      LHEKKTLPGM NHPMQVKPAD TVNKDRKLFV GMLGKKQDEE DVRRIFEKFG TIEECTILRT TDGQSKGCSF
HrETR1     LHEQKTLPGM NHPMQVKPAD TVNKDRKLFV GMLGKRQTEE DIRQLFEKFG HIEECTILRT PDGQSKGCSF
MoETR      LHEKKTLPGM NHPMQVKPAD TVNKDRKLFI GMLGKRQTEE DVKKLFEPFG KIEECTILRL PDGQSKGCSF
Cs-ETR1    LHERKTLPGM NHPIQVKPAD TVSKDRKLFV GMLGKRQNED DVRILFEPFG TIEECTILRT PEGQSKGCAF
Pm-ETRa    LHERKTLPGM NHPLQVKPAD TVNKDRKLFI GMLGKRQTEE DVQALFEPFG NIEECTILRT PEGQSKGCAF
DrETR1     LHEQKTLPGM NRPIQVKPAD SEGRDRKLFV GMLGKQLSDA DVRKMFEPFG SIEECTVLRG PDGASKGCAF
BfHu       LNGLRLQAKT IKVSYARPSS QAIKDANLYV SGLPKTMTQQ DLEGLFEAHG RIITSRILFD PTGQSRGVGF
HsHuB      LNGLRLQTKT IKVSYARPSS ASIRDANLYV SGLPKTMTQK ELEQLFSQYG RIITSRILVD QTGISRGVGF
HsHuD-     LNGLRLQTKT IKVSYARPSS ASIRDANLYV SGLPKTMTQK ELEQLFSQYG RIITSRILVG QTGVSRGVGF
HsHuc      LNGLKLQTKT IKVSYARPSS ASIRDANLYV SGLPKTMSQK EMEQLFSQYG RIITSRILVD QTGVSRGVGF
HsHuR      LNGLRLQSKT IKVSYARPSS EVIKDANLYI SGLPRTMTQK DVEDMFSRFG RIINSRVLVD QTGLSRGVAF
DmELAV     LNGLRLQNKT IKVSFARPSS DAIKGANLYV SGLPKTMTQQ ELEAIFAPFG AIITSRILQN ANDTTKGVGF

           141
BsETR      VKLSSAAAAK AAIEAIHGKT MPGASSSIVV KLADTDRERA VRKMNQIANS -PVALQLSPY GSTIPQGWSP
HrETR1     VKLSTSTGAR AAIDALHGST MPGASSSIVV KLADTDKERA IRKMQQMAQN SPVTLQLGPY GSPLPQGWSP
MoETR      VKLSNADDAK AAINALHGST MPGASSSLVV KLADTDKERA VRKMQQMANN SPVALQLGTY PA----GWSP
Cs-ETR1    VKLSCHQEAK SAMEALHGST MPGASSSLVV KFADTDKERA VRRMQQMAQT SPVPLQIGTY GLPYAQGWNS
Pm-ETRa    VKLSTNQEAR AAIDSLHGST MPGASSSLVV KFADTDKERA IRRMQQMAQT NPVPVQVGTY SLPYPQGWNS
DrETR1     VKYQSNAEAQ AAISALHGST LPGASSSLVV KFADTEKERG IRRMQQVASQ SPMNLHLGAY NAYTQAYLSP
BfHu       VRFDQRVEAE RAITELNGHI PKGATDPITV KFANNPSQNH AKALQQAAYL APARRYLGPM LHQTARRYSP
HsHuB      IRFDKRIEAE EAIKGLNGQK PPGATEPITV KFANNPSQKT NQAILSQLYQ SPNRRYPGPL AQQAQRRFSP
HsHuD-     IRFDKRIEAE EAIKGLNGQK PSGATEPITV KFANNPSQKS SQALLSQLYQ SPNRRYPGPL HHQAQRRFSP
HsHuc      IRFDKRIEAE EAIKGLNGQK PLGAAEPITV KFANNPSQKT GQALLTHLYQ SSARRYAGPL HHQTQRRFSP
HsHuR      IRFDKRSEAE EAITSFNGHK PPGSSEPITV KFAANPNQNK NVALLSQLYH SPARRFGGPV HHQAQRRFSP
DmELAV     IRFDKREEAT RAIIALNGTT PSSCTDPIVV KFSNTPGSTS KIQLPAFLNP QLVRRIGGAM HTPVNKRFSP

           211
BsETR      MTALAAQQIT SIPPQKEGPE GCNLFIYHLP AEFQDQDLAN AFQRFGTVIS AKVFIDRATS QSKCFGFVSF
HrETR1     LATAFGAQIA GIPPQKEGPE GCNLFIYHLP QEFTDADLAN VFQPFGNVIS AKVFIDRATN QSKCFGFVSY
MoETR      IATGTLPTTP GIPPQKEGPE GCNLFIYHLP QEFTDADLAN VFQPFGSVIS AKVFIDRATN QSKCFGFVSY
Cs-ETR1    LASPQPTPMS GVPPQKEGPD GCNLFIYHLP QEFSDADLAN VFQPFGNVIS AKVFIDRATN QSKCFGFVSF
Pm-ETRa    LAAPQPTQLS GIP----GPE GCNLFIYHLP QDFTDADLRN VFQPFGNVIS AKVFIDRATN QSKCFGFVSF
DrETR1     VATV--SSGT STPQQREGPE GCNIFIYHLP QEFTDSEMLQ MFLPFGNVIS AKVFVDRATN QSKCFGFVSF
BfHu       MGADPM---- GVGNIPNGGT GFCIFVYNLA PDTEDSVLWQ LFGPFGAVTN VKVIRDFQTN KCKGFGFVTM
HsHuB      MTIDGMTSLA GINIPGHPGT GWCIFVYNLA PDADESILWQ MFGPFGAVTN VKVIRDFNTN KCKGFGFVTM
HsHuD-     ITIDGMTSLV GMNIPGHTGT GWCIFVYNLS PDSDESVLWQ LFGPFGAVNN VKVIRDFNTN KCKGFGFVTM
HsHuc      IAIDGMSGLA GVGLSGGAAG GWCIFVYNLS PEADESVLWQ LFGPFGAVTN VKVIRDFTTN KCKGFGFVTM
HsHuR      MGVDHMSGLS GVNVPGNASS GWCIFIYNLG QDADEGILWQ MFGPFGAVTN VKVIRDFNTN KCKGFGFVTM
DmELAV     MAGDMLAAAA ATTLASGPGG AYPIFIYNLA PETEEAALWQ LFGPFGAVQS VKIVKDPTTN QCKGYGFVSM

           281
BsETR      DNSGSAQAAI QGMNGYQIGP KRLKVQLKRP KEQS
HrETR1     DNPVSAQTAI QTMNGFQIGM KRLKVQLKRP KEQS
MoETR      DNPLSAQTAI QTMNGFQIGM KRLKVQLKRP KEQS
Cs-ETR1    DNAISSQAAI QTMNGFQIGM KRLKVQLKRP KEQS
Pm-ETRa    DNAISSQAAI QTMNGFQIGM KRLKVQLKRP KEQS
DrETR1     DNPASAQAAI QAMNGFQIGM KRLKVQLKRP KDAN
BfHu       AHYDEAVVAI AQLNGYCLGG RVLQVSFKTN KAHK
HsHuB      TNYDEAAMAI ASLNGYRLGD RVLQVSFKTN KTHK
HsHuD-     TNYDEAAMAI TSLNGYRLGD RVLQVSFKTN KAHK
HsHuc      TNYDEAAMAI ASLNGYRLAE RVLQVSFKTS KQHK
HsHuR      TNYEEAAMAI ASLNGYRLGD KILQVSFKTN KSHK
DmELAV     TNYDEAAMAI RALNGYTMGN RVLQVSFKTN KAKG
```

BsNotch

ML, LG, 100 bootstrap, mafft, gappyout



Notch

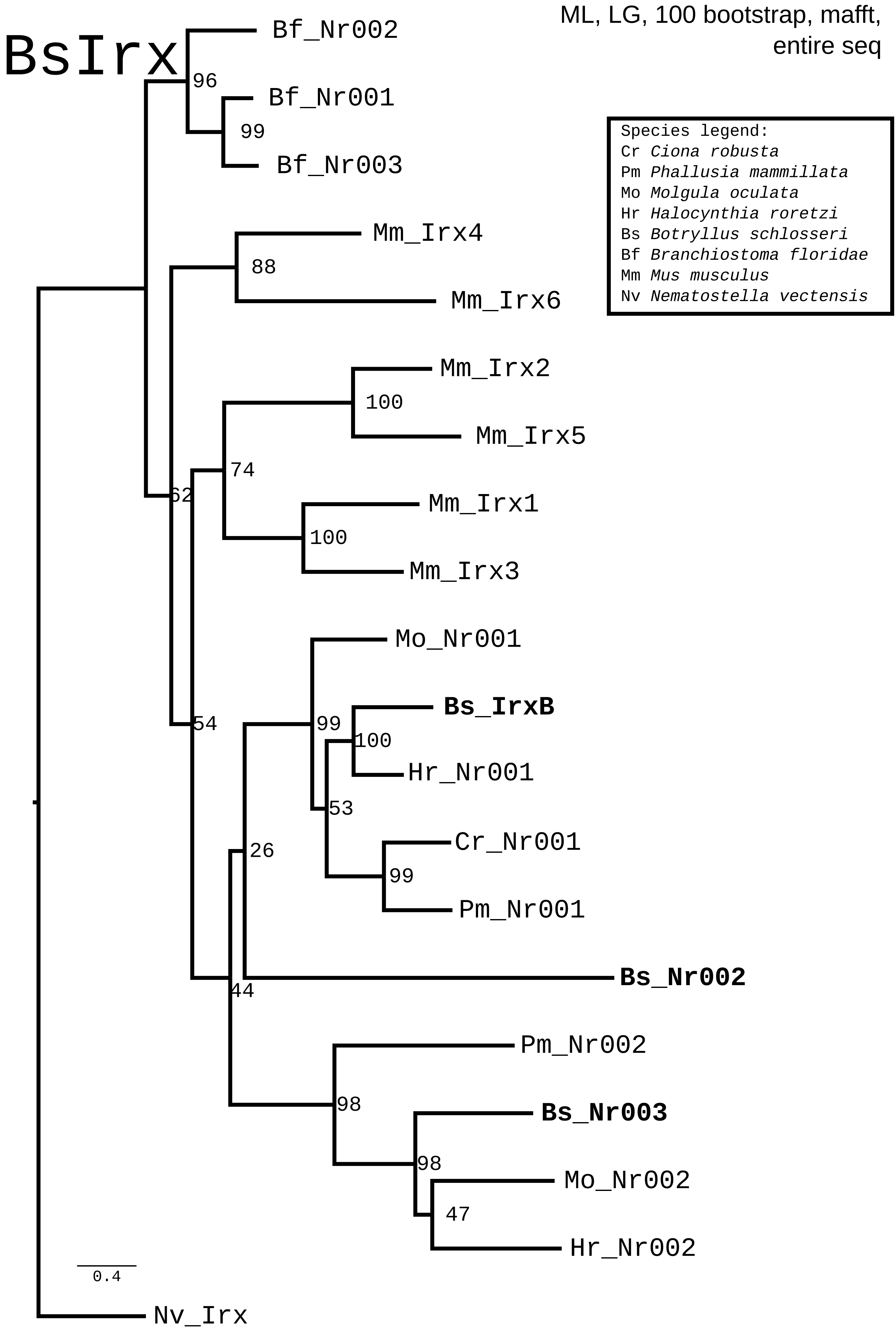
```
Bs_notch    MK-------- ---------- ---------- ---------- ---------- ---------- ----------
Hr_Notch    MLSSIVESLC TSDQCKNGGT CQSNTCSCTS SYVGDTCEVS SPCIPSPCLN SGTCAVTELT QCTCQTGFTG
Ci_Notch    MGFSFAGNCC SLNPCLNGGE CGNDICRCAR GYGGATCNVQ DTCDVSHCNN GGTCHIQDGR VCVCEENYTG
Bf_Notch    MFSCLFLLLC QANPCLNGGV CRQGTCICRD GWVGEYCQEV DPCASNPCRH GGTCVITIDD QCNCQPGYTG
Mm_Notch2   ---------- ---------- ---------- ---------- ---------- ---------- ----------
Hs_Notch2   MPLCCAAPAC RDGPCVNEGM CGTGYCKCPE GFLGEYCQHR DPCEKNRCQN GGTCVMLSRD ECTCQVGFTG
Mo_Notch    MDTGIIQTTC GGDPCENGGT CAGSSCSCQT GYIGTYCQYV DPCPSNPCQN DGSCIALDSD QCSCGTQNYG
Nv_Notch    ---------- ---------- ---------- ---------- ---------- ---------- ----------

           71
Bs_notch    ---------- ---------- ---------- ---------- ---------- ---------- ----------
Hr_Notch    DTCSQVLYCS PNPCSNGAGC EELSNSFKCT CTSGYYGDTC ANDVNECDTP DICQNAGTCS NNDGGYSCSC
Ci_Notch    PRCNNYDYCS SEPCRNGATC TNLPQHFSCT CAEYWFGDTC ENFRSPCISP NICKNNGVCE TTDDHYQCVC
Bf_Notch    DTCDVVEHCY SQPCKNGGTC TSSESGYTCT CLGGFEGSTC QSDIDECAGT NPCQNGGQCS NTMGSYVCHC
Mm_Notch2   ---------- ---------- ---------- ---------- ---------- ---------- ----------
Hs_Notch2   KECQWTDACL SHPCANGSTC TTVANQFSCK CLTGFTGQKC ETDVNECDIP GHCQHGGTCL NLPGSFECNC
Mo_Notch    DRCEYTDHCA SDPCDNGGTC ADTTSTFICT CPSTYFGQTC ANDKDECTDP NLCQNGGACQ NTIGGYDCTC
Nv_Notch    ---------- ---------- ---------- ---------- ---------- ---------- ----------

          141
Bs_notch    ---------- ---------- ---------- ---------- ---------- ---------- ----------
Hr_Notch    VAGFEGNNCE VNIDDCSGHS CQNGATCADA VSTYDCHCPA EWTGQYCTID VDECSLSNNV A-KRRDLQQT
Ci_Notch    LPGFGGTNCE IDIDDCQDDP CHNGATCIDQ ANSPLCQCAP GWTGATCEHG VDECALNPSA CLHGGTCLNS
Bf_Notch    MSGFTGDDCE VNVDDCVDHL CENGAACVDG VNEYTCTCPS QWAGRYCNED VDECMQSPNI CLNCGTCHNT
Mm_Notch2   ---------- ---------- ---------- ---------- ---------- ---------- ----------
Hs_Notch2   LPGFEGSTCE RNIDDCPNHR CQNGGVCVDG VNTYNCRCPP QWTGQFCTED VDECLLQPNA CQNGGTCANR
Mo_Notch    DSGFSGTNCN VNIDDCVSHA CENNGSCVDG VNSYTCDCTS GWIGDYCGVD IDECMINPNQ CQNGGTCTNS
Nv_Notch    ---------- ---------- ---------- ------SCPK NFTGARCEID VDECTTLSQP CQNGGTCSNV

          211
Bs_notch    ---------- ---------- ---------- ---------- ---------- ---------- ----------
Hr_Notch    EGGFTCNCVY GFTRDDCSEN IDDCSNVACF HNARCIDQAG TFECLCTPGN R-ILCHLDDA CISDPCARGA
Ci_Notch    EGSYQCTCVN GWTGEDCGTN EDNCLNKPCF HGATCIDLVG RFECICPLNK TGLLCHVDNA CDPPPCAAGA
Bf_Notch    VGGYSCVCVN GWIGDDCSEN FDDCASAACF DGATCHDRVG FFMCECAPGK TGLLCHLDDA CESSPCNEGA
Mm_Notch2   ---------- ---------- ---------- ---------- ---------- ---------- ----------
Hs_Notch2   NGGYGCVCVN GWSGDDCSEN IDDCAFASCT PGSTCIDRVA SFSCMCPEGK AGLLCHLDDA CISNPCHKGA
Mo_Notch    EGSYKCNCVN GWEGFNCEVN IDDCAQTACF FGSTCHDKVA SFSCECAPGI TGLLCHLKDE CETNPCHEGA
Nv_Notch    YGGYMCRCIT GWDGADCSVN IDECKQPRCQ HGGTCVDKIG SYTCICPPGK TGLVCNFDDE CASNPCSANA

          281
Bs_notch    ---------- ---------- ---------- ---------- ---------- ---------- ----------
Hr_Notch    TCDTNPITGH WMCDCPDGWT DKDCSKDIDE CSLGGNPCEH NGQCNNTDGS FECICVAGYS GPRCETNINE
Ci_Notch    QCDTNPINGQ AICTCPAGYT GVNCSLDLNE CDLGTNPCEY DGTCVNIPGS YKCVCANGYS GPRCEENINE
Bf_Notch    ICDTNPVNGQ PICTCPDGYE GQLCMQDIDE CALGENPCEH DGECNNVPGS FTCTCTDGFT GDRCEVNINE
Mm_Notch2   ---------- ---------- ---------- ---------- ---------- ---------- ----------
Hs_Notch2   LCDTNPLNGQ YICTCPQGYK GADCTEDVDE CAMASNPCEH AGKCVNTDGA FHCECLKGYA GPRCEMDINE
Mo_Notch    KCDTNPMTGK AICTCPDGYT GPVCDIDIDE CSKG-DPCEH GGTCVNLAGS FSCECQAGYT GRTCTMNINE
Nv_Notch    TCVTS-FSGK ASCICNSGWT GKNCDVDIKE CEGDSSPCYH GGTCRDIPGS FVCDCVPGFA GARCEDNINE

          351
Bs_notch    ---------- ---------- ---------- ---------- ---------- ---------- ----------
Hr_Notch    CEPNPCRNDA TCLDMIGNFN CVCMPGFTGI ICDEDIDECE SNPCANGGTC IDEVNAYTCS CALGFTGDDC
Ci_Notch    CESNPCVNDA TCLDKIGTFL CLCLPGYTGV TCSEEIDECE SNPCLYDGEC IDQVHGYICN CAAGYTGTNC
Bf_Notch    CASNPCQNQG TCIDDIGEFR CACMPGFAGD LCETDVDECA SSPCLN-GLC RDGINKYECE CDPGFEGTTC
Mm_Notch2   ---------- ---------- ---------- ---------- ---------- ---------- ----------
Hs_Notch2   CHSDPCQNDA TCLDKIGGFT CLCMPGFKGV HCELEINECQ SNPCVNNGQC VDKVNRFQCL CPPGFTGPVC
Mo_Notch    CEDMPCKNDA TCLDDIGQYQ CVCMAGYSGV NCTEDIDECA SSPCLNSGGC THGINSYTCT CDAGFTGKNC
Nv_Notch    CESNPCVH-G LCLDYQNKFE CACSKGYTGR LCDVEINECD SNPCLNGGQC HDGLGNYSCT CQVGYVGEIC

          421
Bs_notch    ---------- ---------- ---------- ---------- ---------- ---------- ----------
Hr_Notch    SQNIDECAST PCMNKATCID KANAYECECA PGYTGVHCET NIDDCVINPC HYGSCRDGVN TFYCDCLLGY
Ci_Notch    QTNIDDCALA HCQNHIECVD DVDDYTCICL PGYGGRNCET DIDECASSPC HYGTCVDGIN RYTCNCELGY
Bf_Notch    ENNINECANG PCRNGAHCSD LVTTYACTCL EGFTGTDCEI NIDDCQSNQC QHGTCVDGVA SFTCSCEPGY
Mm_Notch2   ---------- ---------- ---------- ---------- ---------- ---------- ----------
Hs_Notch2   QIDIDDCSST PCLNGAKCID HPNGYECQCA TGFTGVLCEE NIDNCDPDPC HHGQCQDGID SYTCICNPGY
Mo_Notch    DFNINECTSN PC-SHGTCQD GANQFVCFCD EGYTGPTCDI NIDDCAIKPC FHGTCQDEVN GYSCVCDGGY
Nv_Notch    QTNYDDCASN PCQNDGTCQD GIAQYTCLCP LGFTGKSCET NTNECAGNPC NMGTCIDGIN SFRCQCPLGF
```

```
            491
Bs_notch    ---------- ---------- ---------- ---------- ---------- ---------- ----------
Hr_Notch    EGTKCQTDTN ECASSPCENG GTCTDEIGYY TCTCPTGTSG SSCEINPDDC VGNPCQYGTC VDGVDDYSCS
Ci_Notch    GGTLCDTPIN ECDSNPCQNG GTCHDLVASY SCDCPLGTAG SNCETNIQEC ASNPCQHGQC VDGINKYTCD
Bf_Notch    NGPLCESPVD ECDSDPCQNG GTCEDLVNGY RCNCLAGTSG SNCEVNQDDC TGNLCVHGVC QDGLNDYTCQ
Mm_Notch2   ---------- ---------- ---------- ---------- ---------- ---------- ----------
Hs_Notch2   MGAICSDQID ECYSSPCLND GRCIDLVNGY QCNCQPGTSG VNCEINFDDC ASNPCIHGIC MDGINRYSCV
Mo_Notch    TGEKCKDEID ECDSTPCNYG GICTDLVNSY SCACPTGTSG PSCEVNPDDC IAVNCGSGTC IDGLNDYTCQ
Nv_Notch    TGNRCETEID ECESSPCQNG GTCKDKINGY VCICPPGASG THCENDPNDC PANACQNGVC IDGMNTYSCK

            561
Bs_notch    ---------- ---------- ---------- ---------- ---------- KFSEDMFSP- ----------
Hr_Notch    CTPGYTGEHC DTDINECDSN PCMNGATCQN EVNNFVCQCP PGIMGTQCSS DIQECSSNPC LHEYARRDQH
Ci_Notch    CDPGYSGLHC GADINECLSN PCYNGGTCEN RLNSFVCHCP AGYNDPRCYS NIDECMSNPC VNGDCQDDIN
Bf_Notch    CDGGYEGNNC EREIDECASS PCHNGGICHD LVNAFSCECP PGYHDQLCYS NVNECESSPC AHGTCQDGIN
Mm_Notch2   ---------- ---------- ---------- ---------- ---------- ---------- ----------
Hs_Notch2   CSPGFTGQRC NIDIDECASN PCRKGATCIN GVNGFRCICP EGPHHPSCYS QVNECLSNPC IHGNCTGGLS
Mo_Notch    CPKGYEGELC NKEINECNSN PCQNGGVCID ENDGFRCDCP DGYTSPTCND NFNECSSNPC QHGTCTDGVN
Nv_Notch    CHPGFTGFSC GVPVNECASS PCRNGGTCED GVAQYYCRCR DGYTSKTVLL TLNLQMKNA- ----------

            631
Bs_notch    ---------- ---------- ---------- ---------- ---------- ---------- ----------
Hr_Notch    VHCICDAGYQ GENCETEINE CASNPCQH-G ACENKVAQFV SHCDAGYTGT ACEIDINECA TQPCQNGGTC
Ci_Notch    KYCNCYHGYV GVNCEYDINE CESSPCVNGG SCLDLVNSYS CRCKTGYEGV NCETDIDECS PSPCLNGGSC
Bf_Notch    DYCTCENGYE GKNCDVNIDE CASNPCQHEG QCDDGIGRYE CQCLPGYEGV NCDINTDECA SNPCQNGGRC
Mm_Notch2   ---------- ---------- ---------- ---------- ---------- ---------- ----------
Hs_Notch2   GYCLCDAGWV GINCEVDKNE CLSNPCQNGG TCDNLVNGYR CTCKKGFKGY NCQVNIDECA SNPCLNQGTC
Mo_Notch    HFCSCVDGFT GLACENEMDE CDSDPCQNDA TCVDFINEYD CLCKDGFSGV NCENEIDECA STPCVNGGLC
Nv_Notch    ------HSFS GKNCEVRIDS CIDHTCQNGA SCVSSTYAYS CQCKPGYTGQ YCETDVDECA ARPCVNGA--

            701
Bs_notch    ---------- ---------- ---------- ---------- ---------- ---------- ----------
Hr_Notch    TSGINSYNCA CPAKYTGVNC ETELSPCVPN PCENGATCQE SADYLAYVCQ CPEGFRGPTC ATDINECVNS
Ci_Notch    ANLIGRYVCT CPLGFTGSEC ETALTPCDPI PCNNGGSCQP NADYLGYTCV CTPGFQGPTC ANDINECMSP
Bf_Notch    LDGVNNYVCD CDLPFVGTNC QTELAPCRPN PCENLGACIP SADYQTFTCN CADGFEGETC ADDINECQSN
Mm_Notch2   ---------- ---------- ---------- ---------- ---------- ---------- ----------
Hs_Notch2   FDDISGYTCH CVLPYTGKNC QTVLAPCSPN PCENAAVCKE SPNFESYTCL CAPGWQGQRC TIDIDECISK
Mo_Notch    TDGTNEYTCV CKAQFTGKNC QHELTPCDPN PCQNSAVCSA SPSYLTFQCN CGTGFKGTTC SEDINECDTS
Nv_Notch    ---------- ---------- ---------- ---------- ---------- ---------- ----------

            771
Bs_notch    ---------- ---------- ---------- --DD------ ---------- ---------- ----------
Hr_Notch    PCKNGGGCTN LVPGYQCTCS QGFTGKDCDT DIDDCSSNPC LNGGQCLDDV GSYKCLCLPG FEGNNCQEEV
Ci_Notch    PCKNGGKCRN REPGYFCECL DGYSGVNCEE NVDDCASDPC MNGGTCLDDV NSYKCLCKRG FDGNQCQNDV
Bf_Notch    PCKNGAPCVN LEGDFRCDCL MGFAGELCSV NIDDCDPDPC HNGGTCNDGI NSYTCSCMPG FGGTNCEEDI
Mm_Notch2   ---------- ---------- ---------- ---------- ---------- ---------- ----------
Hs_Notch2   PCMNHGLCHN TQGSYMCECP PGFSGMDCEE DIDDCLANPC QNGGSCMDGV NTFSCLCLPG FTGDKCQTDM
Mo_Notch    PCLNSGICAN TEPGYQCQCM DGYSGTNCAV DIDDCASSPC LNGATCEDDV NQYKCVCKKG FDGIRCESEV
Nv_Notch    ---------- ---------- ---------- ---------- -----CVDGV NGFICRCDPG FTGDRCQINV

            841
Bs_notch    ---------- ---------- ---------- ---------- ---------- ---------- ----------
Hr_Notch    NECASFPCKN GGICTDYVNS YVCTCLSGFY SLDCEKNIED CSSSSCMNGG TCVDGINSYS CSCTANFTGD
Ci_Notch    NECENEPCKN GATCTDYVNS YACTCPPGFR GTTCMENIDE CNIGSCLNGG TCVDGINSYS CNCMAGFTGA
Bf_Notch    DECYSNPCQN GGQCIDAVNG YACDCVVGFI GTNCQTNKDD CTSSSCFSGG TCIDGINTFT CHCPSGFTGS
Mm_Notch2   ---------- ---------- ---------- ---------- ---------- ---------- ----------
Hs_Notch2   NECLSEPCKN GGTCSDYVNS YTCKCQAGFD GVHCENNINE CTESSCFNGG TCVDGINSFS CLCPVGFTGS
Mo_Notch    NDCDPNPCQN GATCTDYFNS YVCSCLVGYI GVNCETNVND CTDSSCFNGG TCVDGVNTYT CQCAPGYEGY
Nv_Notch    DDCQSSPCVH GGSCIDSINT YTCQCPKGFT GPRCEIHINE CSSDPCQHGG TCSDRIGSYS CYCRPGYTGS

            911
Bs_notch    ---------- ---------- ---------- ---------- ---------- ---------- ----------
Hr_Notch    KCQNAVNNCA SLQCQNGGTC YYDSGDPKCA CVHGYTGTHC ESLQNLCTPN ICKNGSCVQT SNTVSCNCLG
Ci_Notch    NCERDIDECV SSPCKNGAPC IHGINTFTCQ CLTGYTGPTC AQMVDLCQNN PCRNGQCSQT GTTSKCLCTS
Bf_Notch    NCQHEINECD SNPCQNGATC VDQTGYFSCI CTYGYEGVTC QSQKDLCADD PCRNGTCTQS GDRYECLCED
Mm_Notch2   ---------- ---------- ---------- ---------- ---------- ---------- ----------
Hs_Notch2   FCLHEINECS SHPCLNEGTC VDGLGTYRCS CPLGYTGKNC QTLVNLCSRS PCKNGTCVQK KAESQCLCPS
Mo_Notch    KCESDVNECD SNPCKNGATC KNHIASYECH CLAGYTGTNC ETLINLCSPN PCEHGTCAQQ YTEFTCTCNS
Nv_Notch    NCQHPLDRCA NDPCRNGATC RRTGEDLS-- ---------- ---------- ---------- --DFHCECPL

            981
Bs_notch    ---------- ---------- ---------- ---------- ---------- ---------- ----------
Hr_Notch    GYEGTDCAVP QVSCTVGASL LGIAVSDLCL NGGTCHDTST AHECSCVAGF TGSYCDIDID ECASVPCKNG
Ci_Notch    SYSGVYCDVP RLSCSAAATW QGVEETSLCQ HGGQCINSGS THYCSCRAGY VGSYCETDED DCASYPCKNG
Bf_Notch    EWTGLICDMT KVSCAAAASG RNVSLANLCQ NGGTCVDTGN SHNCNCAAGY RGSYCSEEID ECASSPCQNG
Mm_Notch2   ---------- ---------- ---------- ---------- ---------- ---------- ----------
Hs_Notch2   GWAGAYCDVP NVSCDIAASR RGVLVEHLCQ HSGVCINAGN THYCQCPLGY TGSYCEEQLD ECASNPCQHG
Mo_Notch    GYTGKTCGNQ LQTCEVQAAS QGTSVANVCN GNGTCINSGV SHMCTCNAGY TGSYCDTKIN HCDPDPCINA
Nv_Notch    GYKGTICDVK EVSCAVAGT- -------ICA NGGTCFDSNG VQSCTCKPGF TGSYCRTNID ECAKGPCKYG
```

```
          1051
Bs_notch  ---------- ---------- ---------- ---------- ---------- ---------- ----------
Hr_Notch  ATCNDLINSY SCICALGYEG ATCLTDKDEC ASSPCKNGGT CIDRINSFYC SCLAGTEGVL CEINEDECEI
Ci_Notch  ATCTDYPGSY TCTCMDGFQG TRCETELNEC ESNPCLNNGQ CVDLINAYKC LCQVGTSGDR CQDNENECTS
Bf_Notch  AECRDGLGTY TCACRPGYQG VNCEQEINEC ISNPCQNGGT CIDMVNEYRC SCPPGTQGLL CEINNDNCFA
Mm_Notch2 ---------- ---------- ---------- ---------- ---------- ---------- ----------
Hs_Notch2 ATCSDFIGGY RCECVPGYQG VNCEYEVDEC QNQPCQNGGT CIDLVNHFKC SCPPGTRGLL CEENIDDCAR
Mo_Notch  VGCTSGINGY ECECEPGYQG VICQLDIDEC TSNPCLNGGT CLNAINGFQC SCPRGTLGLL CEVVISLCDP
Nv_Notch  ATCHDAVANY TCTCTAGFTG KNCDININEC ASNPCQR-GS CLDLVNGYLC SCPKGYIGKH CEVNADDCFV

          1121
Bs_notch  ---------- ---------- ---------- ---------- ---------- ---------- ----------
Hr_Notch  NICLNGGVCI DGIGGFSCQC PSGYEGRRCQ GDVNECLSNP CSSPGSLACI QGSNSYQCVC DADYTGSECQ
Ci_Notch  TSCFNSGNCV DGIGGFTCTC VPGYAGPRCE GDVNECVPNP CNTEGSTNCV QGINKFTCTC KQGYTGTRCE
Bf_Notch  GACYHDGTCV DGIGEFTCRC RPGYVGQRCE GDVNECLSNP CDAEGTLDCV QLENDYSCDC KPGYTGGRCE
Mm_Notch2 ---------- ---------- ---------- ---------- ---------- ---------- ----------
Hs_Notch2 GPCLNGGQCM DRIGGYSCRC LPGFAGERCE GDINECLSNP CSSEGSLDCI QLTNDYLCVC RSAFTGRHCE
Mo_Notch  NPCQNNGHCT SGVGSFSCQC KPGYEGDLCN GDINECLPNP CNPGGSVDCV QGINEFACLC KDGYYGDTCS
Nv_Notch  NACFNGGSCV DGIAEFKCTC PLGFSGSRCE VDVDECASSP CSALGTEKCI NNIGAYHCQC KQGFLGRHCD

          1191
Bs_notch  ---------- ---------- ---------- ---------- ---------- ---------- DM--------
Hr_Notch  IRIGSCDINP CLNDGICTDN SQDITGYKCQ CTWGYYGKKC ENSYSMCSAN FFCYHEAPCQ DG---TCICT
Ci_Notch  KKLTSCSTMP CFNGATCTDA PYEPTGYRCT CSNGYRGTNC EEVMSRCRDG LVCYNGGTCH ENGGEMCXCP
Bf_Notch  RTVDSCESDP CLNGGACSQS GYEDGGFQCS CTRDYAGDRC QFHESNPCFS LPCYNNGDCQ QTDSYRCMCP
Mm_Notch2 ---------- ---------- ---------- ---------- ---------- ---------- ----------
Hs_Notch2 TFVDVCPQMP CLNGGTCAVA SHTASGPRCF CP---SPRDC ESG----CAS SPCQHGGSCH PQPYYSCQCA
Mo_Notch  SQASSCSKNP CLNGATCTDN SLEPIGYFCS CTNDYRGKNC EIEFNTCSLD MICYNDGQCV DGSSPYCKCP
Nv_Notch  LNINECLSYP CRNGGECKDG ADGYGSYSCK CPLGFAGANC EESINEC-LS QPCKNGGSCR DINGYKCDCP

          1261
Bs_notch  ---------- ---------- ---------- ---------- ---------- ---------- ----------
Hr_Notch  ENYYGTNCEK TTAQTCSNTC ELVKGDGICD KECNVHECEF DGGDCSLGIS LGGKCDKELR CWQLYQDGHC
Ci_Notch  ADYNGDNCIY YIPGVCDPTC AALKGNGECD AKCNTHGCEY DGGDCSYGNT PWAKCPASLQ CYNYFHDGAC
Bf_Notch  EDYNGLVCEI YIPDSCLDSC ASKARDGHCD EECNIHECDW DGTDCALGNH PWANCTSSLQ CWNYFADGNC
Mm_Notch2 ---------- ---------- ---------- ---------- ---------- ---------- ----------
Hs_Notch2 PPFSGSRCEL YTAPTCSQYC ADKARDGVCD EACNSHACQW DGGDCSLTME PWANCSSPLP CWDYI-NNQC
Mo_Notch  YGYTGTQCMS NTNTQCSQYC QSRAGNGICD NSCNTHGCDY DGGDCTLGVK LWSECRKELE CYLRNGDGHC
Nv_Notch  SKMIGKNCET VVKDQCVPNC AKKFDGGKCN PKCNTHECNW DGTTCSLGIE PWSNCTSGKA CYQVFANGVC

          1331
Bs_notch  ---------- ---------- ---------- ---------- ---------- -----FKPPP R------VI-
Hr_Notch  DEHCNNEDCL HDGMDCATPV GDCNPEKYCN EYYNNDYCDQ GCNNANCGWD GADCIKDYPA QYARGFLVIR
Ci_Notch  DEACNNADCL FDGFDCETPH EACPAQDYCD KNYKNNKCDS GCNTFGCGWD GLDCDKDHPQ QPAAGELIMI
Bf_Notch  DQQCNTEECF YDGRDCEEPH PTCNPDTYCG NFFGDGNCDS GCNNMECGYD GGDCDEEYPP VTAKGYLVMI
Mm_Notch2 ---------- ---------- ---------- ---------- ---------- ---------- ----------
Hs_Notch2 DELCNTVECL FDNFECQGNS KTCK-DKYCA DHFKDNHCNQ GCNSEECGWD GLDCAADQPE NLAEGTLVIV
Mo_Notch  DEQCNSEVCG YDGFDCQPPV GNCPF-DYCA AHYADGNCDQ GCNNIECNYD GLDC-EEYPP QLIVGILVLY
Nv_Notch  DRECNTGGCL FDGFDCKPSV PKCGADKYCA ARFANAECDA ICNNVACQND GLDCSFKKP- EIVEGTLVLV

          1401
Bs_notch  ---------- ---------- ---------- ---------- ---------- ---------- ----------
Hr_Notch  MEITPEEMRE QEMIFIIEFG KILRSGIEI- --ISIEEIFL ESSVEMSKTR HRRATSATSE VKVRLDNRNC
Ci_Notch  VLIPPEELIS RQAEFLRFIG TVLHSVVEFK LTPYIEAISS QLSHNKTHTR LRRASNTFSK AHLLVNSQKC
Bf_Notch  VAAPPSVLLN NSVPFLRMLS IILRSSMSFK IDPMVYPWYG ESNFRALTGR YRRATEVQSI VYLEMDNSKC
Mm_Notch2 ---------- ---------- ---------- ---------- ---------- ---------- ----------
Hs_Notch2 VLMPPEQLLQ DARSFLRALG TLLHTNLRIK RDLMVYPYYG EKSAAMKKQR MTRRSEVASK VFLEIDNRQC
Mo_Notch  SYGDVDEINR LKSEFERDLS RLVRGIIFQM VD--VQLHTA PVSSSSGNGR FRRASESQSM MRFTLNKQMC
Nv_Notch  LLVVPEAFMN GSRVFMRELS RTLNTIAFIK KDELVKVYPL PPSAPVPAER RKRSAEKIVQ VQINLDNRGC

          1471
Bs_notch  ---------- ---------- ---------- ---------- ---------- ---------- ----------
Hr_Notch  LLNDDSCFWE ASSAASYVSA SASTGTLQVS FDVSTATSQP ETSDPETPLV YILSGSVGGV LVLLVAVVLV
Ci_Notch  KQTSSECIDT TDNAASVISA RASTGTLEAE FPINTVASTT PTPPPQ-DYT YIVGAAGGVV LVIVIIGVLT
Bf_Notch  YRAHDDCFET AKEAADFLAA LLHDGLSQLS YPVKSLYSGE PAPPTTQNPL YVVVAAVGIL LVMVGLGVLM
Mm_Notch2 ---------- ---------- ---------- ---------- ---------- ---------- ----------
Hs_Notch2 VQDSDHCFKN TDAAAALLAS HAIQGTL--S YPLVSVVSES LTPE-RTQLL YLLAVAVVII LFIILLGVIM
Mo_Notch  VKLDLICPSN VKNVQNLLGE QADTRTLSSV FRIETEESTE --SGGSSVIT YIAAGGGIMV LIIVIVGVLV
Nv_Notch  ---ETDCFQS TEQAAKYLGA QQSTGKLNLP YPVYSEKPTV PETGFQPEPL WIAILCVGVP LFIV--GVL-

          1541
Bs_notch  ---------- ---------- ---------- ---------- ---------- ---------- --SKY--EEA
Hr_Notch  QKRKRENGTL WLPEGFGVPK KKRREPIGQD LNLKKLSKSS QLDPNSSTPF ISHWEDGNKH KQSKYLAQDQ
Ci_Notch  RKRKRETSTL WAPEGFNVTK KQRREPIGQD -DLNGLNGSG ELTQTAGTPF LNRWENTSPQ KSNHYHVQYS
Bf_Notch  TKRRRESGQL WYPEGFKLTK ERRRDPVGQE VGMKSFTPPA NVDDNETVRD SSAWEDDMPP EAKRMKTQQG
Mm_Notch2 ---------- ---------- ---------- ---------- ---------- ---------- -PKKAKA---
Hs_Notch2 AKRKRKHGSL WLPEGFTLRR DKRREPVGQD VGLKNLS--V QVSELIGTGT SEHWVDDEGP QPKKVKA---
Mo_Notch  QKRKRENGNL WVPEGFQLFK KRRKE----- LNLNNLSKA- --DMNAQTPF LPHAAEA--- QASKYSAET-
Nv_Notch  AGGKRVYTKL WLPEGF-VRR PLRRDPVGQE HSMRSMNKSS DLEE------ EGAVGGDLPP QEARDAK--E
```

```
            1611
Bs_notch    EARKPTPE-F EVQKDKRKWT PGHYEAAR-- -TPSREATTP LMRDN--HQD IDACGPDGIT PLMVATTMGG
Hr_Notch    SKPNSALSDR EAHPDVRKWT PQHLQAGGLS E---LLPSSP LTENGQDLND IDIRGPDGIT PLMVASTRGG
Ci_Notch    SAQKSSSRDD MEPTDNRKWT PQHLEAADLS RTPVTDLTPP -PHIDVDEDD VNARGPDGVT PLMVASIRGG
Bf_Notch    DATADQTLSD YDDTDHRQWT QQHLHAADIR NLAQVALTPP QTEGEADSVD VNVRGPDGLT PLMLASFRGG
Mm_Notch2   --EDEALLSE DDPIDRRPWT QQHLEAADIR HTPSLALTPP QAEQEVDVLD VNVRGPDGCT PLMLASLRGG
Hs_Notch2   --EDEALLSE DDPIDRRPWT QQHLEAADIR RTPSLALTPP QAEQEVDVLD VNVRGPDGCT PLMLASLRGG
Mo_Notch    ---------C KAKEDKRQWT PHHLEAANNS TACQIMNTPP -QSDCPESED INARGPDGYT PLMIASVRGG
Nv_Notch    VDESQRVKVS EKEKDTRQWT RLHREAADVT VTA-LALTPP -QEGESEGID VDARGPGGFT PLHLASCRGT

            1681
Bs_notch    GIEVLEEEEI GEEGSENMIA SLLMQGASLQ SQTERTGETP LHLAARYARA DAAKRLLDAG ADANMKDNTG
Hr_Notch    GIEVVSDEES AEEGSENIIA NLLTQGASLS AQTDRTGETA LHLAARYARA DAAKRLLDAG ADANMKEQTG
Ci_Notch    GVDHGISDDE SQEGSDSMIA GLLLQGASLS AQTDRSGETA LHLAARYARA DAAKRLLDAG ADANMKDHSG
Bf_Notch    GLEYVEDQED GEDESANVIT DLLMQGANIN AQTDRTGETS LHLAARYARA DRAKRLLDAG ADANARDNTG
Mm_Notch2   SSDLSDEDED AEDSSANIIT DLVYQGASLQ AQTDRTGEMA LHLAARYSRA DAAKRLLDAG ADANAQDNMG
Hs_Notch2   SSDLSDEDED AEDSSANIIT DLVYQGASLQ AQTDRTGEMA LHLAARYSRA DAAKRLLDAG ADANAQDNMG
Mo_Notch    GIEIGLSEED SGEGSDNMIA GLILQGASLS ATTDRTGETA LHLAARYARA DAAKRLLDAG ADANMKEQTG
Nv_Notch    LVDGCSIDDD KEDSGGAMVS DLLALGASYG ARTDIEKETP LHLAARHSRA DAAKRLLHAG ADPNARDKLG

            1751
Bs_notch    RTPLHAAVAA DALGVFQILI RSRATDLDAK TMDGTTPLML AARHAVEGMV DDLINAHAEV DTVDNHGKTA
Hr_Notch    RTPLHVAVAA DAQGVFQILL RNRATDLNAS TNDGTTPLIF AARLAVEGMV EELINAQADI NAVDSHGKSA
Ci_Notch    RTPLHSAVAA DAQGVFQILL RNRATDLDAR TNDGTTPMIL ASRLAVEGMV EELISANADV NAVDDHGKSA
Bf_Notch    RTPLHAAIAS DAQGVFQILL RNRATDLDAR TNDGTTPLIL AARLAVEGVV EELLNSHADA NACDDNGKSA
Mm_Notch2   RCPLHAAVAA DAQGVFQILI RNRVTDLDAR MNDGTTPLIL AARLAVEGMV AELINCQADV NAVDDHGKSA
Hs_Notch2   RCPLHAAVAA DAQGVFQILI RNRVTDLDAR MNDGTTPLIL AARLAVEGMV AELINCQADV NAVDDHGKSA
Mo_Notch    RTPLHSSVAA DAQGVFQILL RNRATDLDAK TNDGTTPLIL ASRLAVEGMV EDLITAHADV NAVDNHGKSS
Nv_Notch    RTPLHLAVGA DAQGVFQILL RNRTTDLEAA MEDGTTPLIL AARLDLLDIV KDLIKASCKV NNVDAQGKSA

            1821
Bs_notch    LHWAAAVNNV EALSSLLRAG AKKDAQTERE ETALFLAARE GSYEAVKLLL EFQANRDVTD HMDRLPRDIA
Hr_Notch    LHWAAAVNNI EALTALLRAG ANRDAQTERE ETPLFLAARE GSYEAVRVLL DHFANRDVTD HMDRLPRDIA
Ci_Notch    LHWAAAVNNV DAVSTLLRAG CNRDAQTERE ETPLFLAAKE GSYEAIRILL DHYANRDITD HMDRLPRDIA
Bf_Notch    LHWAAATNNE EAVISLLSHG VNKDVQDNKE ETPLFLAARE GSFQAAKILL DHYANRDITD HMDRLPRDIA
Mm_Notch2   LHWAAAVNNV EATLLLLKNG ANRDMQDNKE ETPLFLAARE GSYEAAKILL DHFANRDITD HMDRLPRDVA
Hs_Notch2   LHWAAAVNNV EATLLLLKNG ANRDMQDNKE ETPLFLAARE GSYEAAKILL DHFANRDITD HMDRLPRDVA
Mo_Notch    LHWAAAVNNN DAIRALLRAG ANHDAQTNKE ETPLFLAAKE GSYEAVKLLL DNFANRDITD HMDRLPRDIA
Nv_Notch    LHWAAAVNSH EVTSELCKNG AKKDMQDDKG QTPLFLGARE GSLEAVRILL LSYANRMIAD NMDKTPEEVA

            1891
Bs_notch    SEKLHSDIIH LLDDFNIVRS PN-------S MDNSPLMY-- -NMGLSKAQP AKNK--RRPK G--DAKAARR
Hr_Notch    NERLHTDIVQ LLDEYNLVRS PNLP----SH MDNSPLVHSN SNYISHMKAT NKGK--RRPN GKNDTKSSRR
Ci_Notch    QERLHTDIVQ LLDEYNLVRS PPLPANMTSH MNGHAMMSRD PYLTSQMKQH TTGKKPRRQS NRADGKPSRR
Bf_Notch    QERMHSDIVK LLDEYNLVRS PSQAANCQSN FGGHHFMNGP A----TQQKR AKPK--RPNK GMADASKMRR
Mm_Notch2   RDRMHHDIVR LLDEYNVTPS PP--GTLTSA L--SPVLCGP NRSFLSLKHT PMGKKARRPN TKKDAKGSRR
Hs_Notch2   RDRMHHDIVR LLDEYNVTPS PP--GTLTSA L--SPVICGP NRSFLSLKHT PMGKKSRRPS AKKDAKGSRR
Mo_Notch    QDRLHIDIVQ LLDEYN-VRS PNMTGS---- -DGSPMMHPG SYM--HMKQH -NGKKNRRSG GK-SEKARRN
Nv_Notch    RQRAHNDIVE LLSDWSICNS PKAAPAPTSP PDQRSPLNGT ASPP-SMDQI SGNKVTTRPK TN-NENGAKR

            1961
Bs_notch    TKKKASSNNA PPNMNQSRPS DGQSP--SMS SP-STYSDPT HSPPITN--- -PANLPPYYP RGMPQG-P--
Hr_Notch    GKKKSASMES PPGMNSSMPS DGQSPSNSLS SPATTYSDST HSPPISN--- --AGMNPYYQ RIISQGNP--
Ci_Notch    VKKTPTSSQG PPSSVTHVGE STLSPNNSLP SPATTYSDST LSPPSSMNGT YHPHLHPHQQ GQFSNMEERM
Bf_Notch    KKKMEGTGTM PRHMD----S TALSPDGSLG GSPHGYNGET -PPPNSMYGV SSQHLHPNLH GALTSGTPQV
Mm_Notch2   KKCLNEKVQL SES------S VTLSPVDSLE ---------- ---------- ---------- ----------
Hs_Notch2   KKSLSEKVQL SES------S VTLSPVDSLE SP-HTYVDTT SSPPNPMTAA PPAPVFSNLH EAHGASTPSV
Mo_Notch    TKKKQAEGHK SPNGVVGHP- DGLSPGGSMT SPATSFSVES LS-PSDMNYP PSSHMSMHQQ RQHSGSMPVD
Nv_Notch    KRKRRK---- -PHGNASQGS STYSPYGKEL SPCSS--GAN FSPPQGSTGI SPSNSSPQQH SSLDANSPQF

            2031
Bs_notch    SMEKHTHPMH HQSQMSMH-- ---------- ---------- ---------- ---------- ----------
Hr_Notch    NGIMENHINI HTPQMTGHGR G-ISGLGLNI GQL------- --------QT RYNETAMHHQ HSQMAAFQSL
Ci_Notch    GTPPRLSHPL YPARSHGHLA GTHTGLGFSL GQIPNGQVPQ EWIQNQQNQQ MYNASGVH-H PSQFAAYRAL
Bf_Notch    NGQPMPRLSH HTADSNSHLR GGETGLGLHL D--PPVRVPN DWLMPQQSQS PPSGPHSHGS PAQMAAFQAL
Mm_Notch2   ---------- ---------- ---------- ---------- ---------- ---------- ----------
Hs_Notch2   ---------- SQLLSHHHIV SPGSGSAGSL SRLHPVPVPA DWMRPPEGKH ITTPIVTFQS IAQPAAPQPQ
Mo_Notch    GVNPRMRQMQ QQQRSMPHLA GSNTGLGLNL GQIPSSQVPQ VVMYNSQSQM RYQP--THQH HNHVTAYHAL
Nv_Notch    DGYLDLDIFE DLTDQDLHLR GSTSDLHDSI SMLPETTVPQ DSSVPTTSAC MYLPAPFGHA PRLYSA----

            2101
Bs_notch    ---------- ---------- ---------- ---------- ---------- ---------- ----------
Hr_Notch    GYRLPDGVRN GEMPFYQPGQ PQPSYQPNMQ QHYGQQITSV SSSN-MEKFP SPPS-NSLSP PSDHFSPSVS
Ci_Notch    GSASSDQRRY SEVPHTQPQL GYTTIQGVPH QHPV-----V QTGTVLEKYP TPPSAHTMSP SSENLSPGPL
Bf_Notch    SNHHNNHRLY HDQPPGE--- --------LA AHPHQTVLP- ----PAEQLP TPPSQHSHGN PSHECTPTHY
Mm_Notch2   ---------- ---------- ---------- ---------- ---------- ---------- ----------
Hs_Notch2   S--------- ---PAVAGIP EMARLPSVAF PTAMMPQQDV AQTISVGKYP TPPSQHSYAS SAAERT----
Mo_Notch    GLSSSEQLRY HDSPLVNSSQ QQRQMHQNMP YHPMHPQQQM QTSTMPDNFP TPPMHNNASP ESDNFSPNAI
Nv_Notch    -HSSPNLCAD NESPSNQSSI QACKLSNRLE EHVMMETTLV ESGMFREKHH TPPSAHSYGT SSYDSSPQKL
```

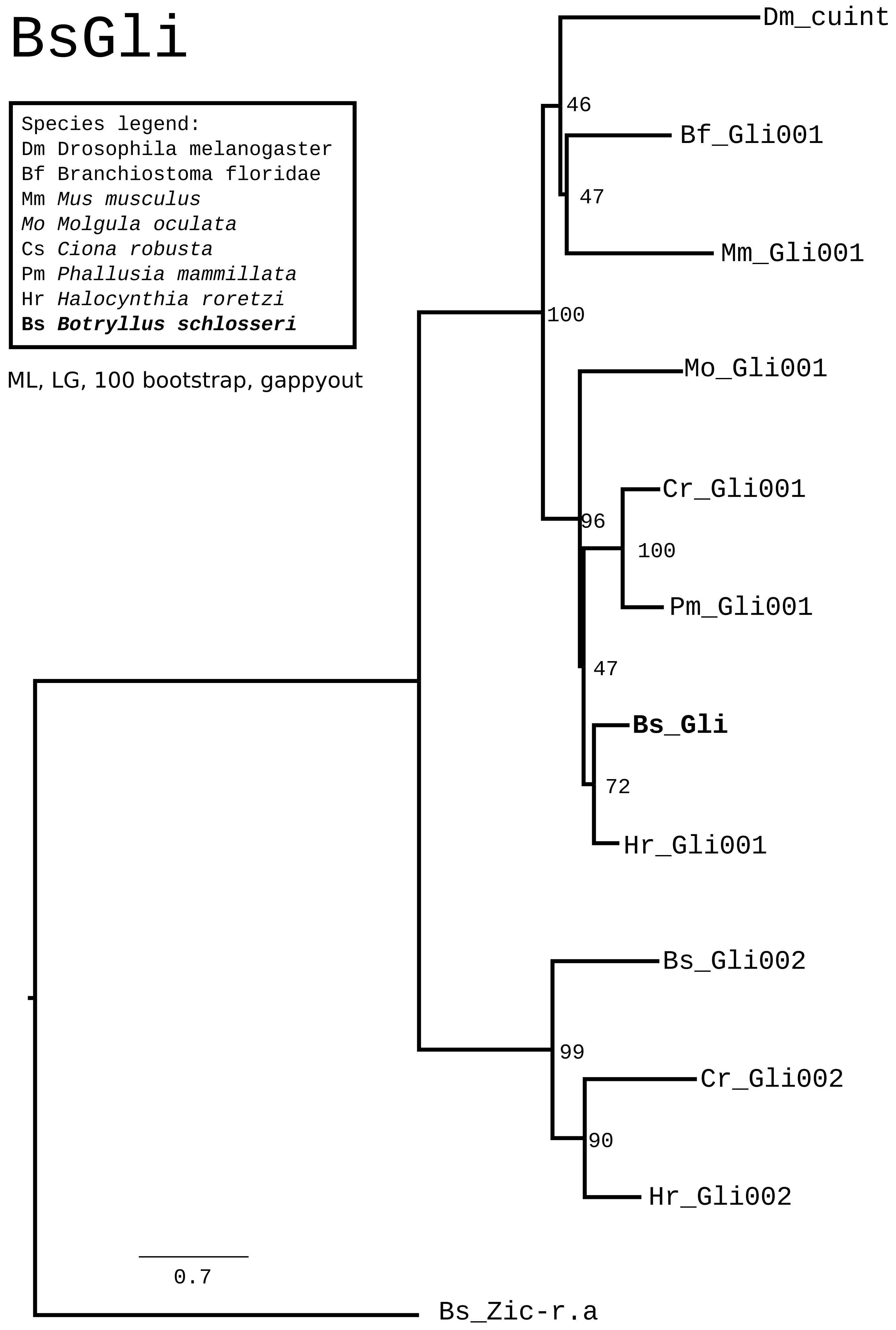
```
             2171
Bs_notch     ---------- ---------- ---------- ---------- ---------- -------
Hr_Notch     ASHVPRTMQE TCYETPSPDS VEWSSSPRSD YWSDSIHSPN GPSSGPQMHH SGVQMYV
Ci_Notch     PPAPKHLVHD LSYETPSPDS PPWSSSSSNS DWSEGIRSPP GPMQGHHHHA PQMRVYA
Bf_Notch     LSHAETIVPE -NYLTPSPDS PQWSSSSATS DWSDGISSPP AAQSHRQPYI PELPVFI
Mm_Notch2    ---------- ---------- ---------- ---------- ---------- -------
Hs_Notch2    -SHSGHLQGE HPYLTPSPES PQWSSSSSAS DWSDVTTSPT PGGAGGGQRG PGMQVYA
Mo_Notch     SQKVGAPMPD SLYETPSPDT SEWSPG-AIS DMSEGVRSPN DIPKTQIVNN HQLPGYA
Nv_Notch     P--------- ISYLTPSPES PDWSSSPSHS DWSN------ ---------- -------
```

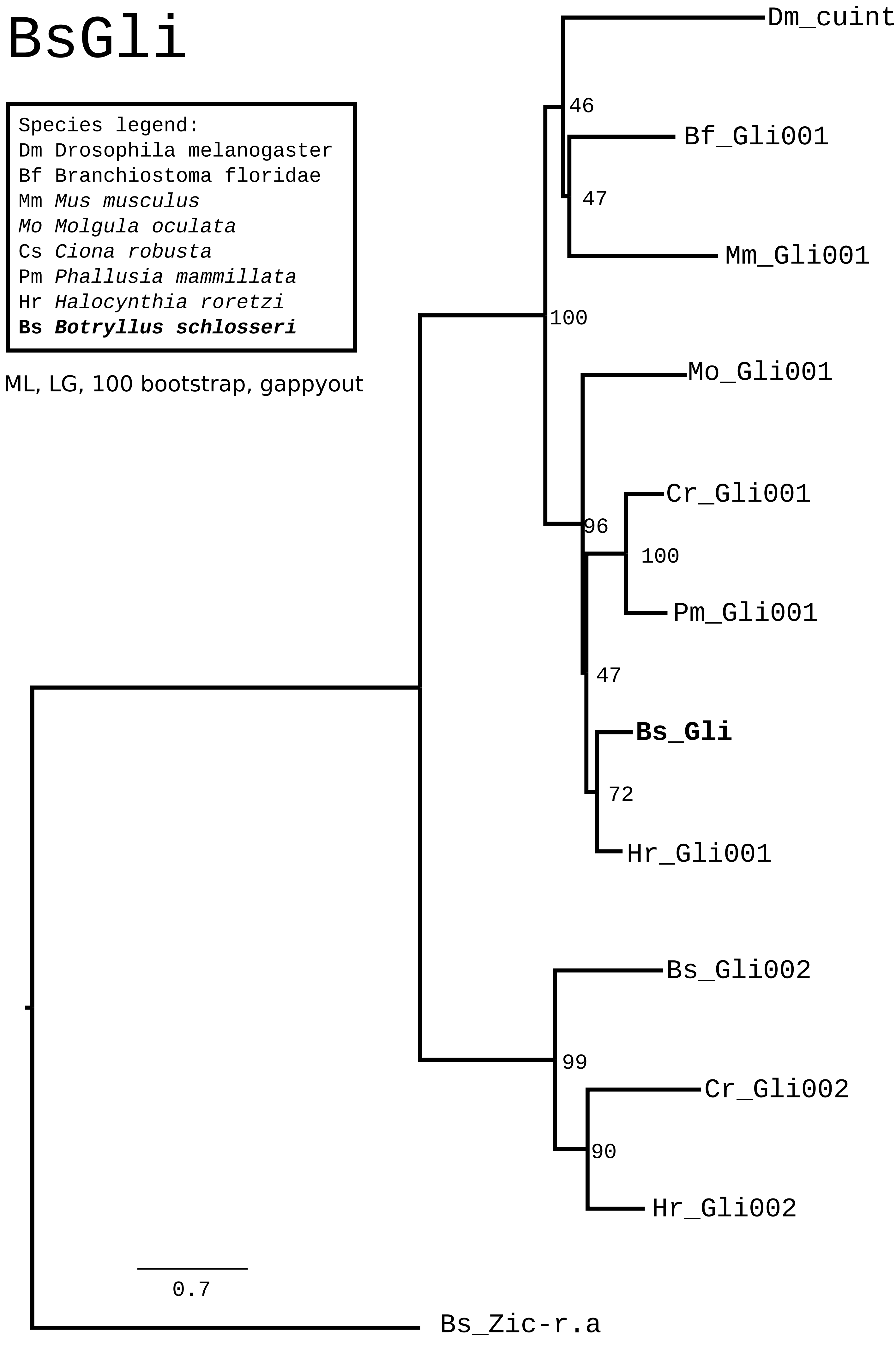

BsIrx
ML, LG, 100 bootstrap, mafft, entire seq
Species legend:
Cr Ciona robusta
Pm Phallusia mammillata
Mo Molgula oculata
Hr Halocynthia roretzi
Bs Botryllus schlosseri
Bf Branchiostoma floridae
Mm Mus musculus
Nv Nematostella vectensis
Bf_Nr002
96
Bf_Nr001
99
Bf_Nr003
Mm_Irx4
88
Mm_Irx6
Mm_Irx2
100
Mm_Irx5
74
62
Mm_Irx1
100
Mm_Irx3
Mo_Nr001
Bs_IrxB
99
54
100
Hr_Nr001
53
Cr_Nr001
26
99
Pm_Nr001
Bs_Nr002
44
Pm_Nr002
98
Bs_Nr003
98
Mo_Nr002
47
Hr_Nr002
0.4
Nv_Irx

Irx

```
            1
Bs_IrxB     MDGRLVAPRI PAIPMYGQPF NTSPFPLPYG PGASAFPQLT IGRAGLQAAA AAYDPYGYGS M-IGDGGRRK
Hr_Nr001    MDARLVASRL PGIAMYGQPF NAPPFPLPYG PTASAFSPLA --RAGLQAAA AYYDPYGYGP MGLSDGARRK
Mo_Nr001    MDARLVASRM PGLAMYGAPY AAQPFPLPYG SPPSAFGPVA AGRAGLQAAA AAYEPYGYGP MGLADGARRK
Cr_Nr001    MDARLVAARL PGFPMYQPPY ATPTFPLHYN PAPAAFPPIG PGRAGLQAAA AAYDPYGYGP MGLSDGARRK
Pm_Nr001    MDARLVAARL PGLAMYGTPY QSPPFQLPYG PAPSAFSHMG PGRAGLQAAA --YDPYGYGP MGLTDGARRK
Bs_Nr003    MLAPKMQSDL PGFHMHAAAA VVTGHPFPLN SPPTQYPSLA PGNGGYQPPH TYYEQYRYFP GDINDGGRRK
Mo_Nr002    MLAPRMQSDV SGFNMYAG-- -AAGPAGAIP SPHSPYPALG TGAGGYQAAP PYYDNYRYCG MDVNDGVRRK
Hr_Nr002    MLATKMQSDI PGFNMYAAAA AATGQTSALN SSHSPYSPLG SGTADYQPSP AYYDSYRYCG MDMNDGVRRK
Pm_Nr002    ---------- ---------- -MHGWGMAGG ---------- PNTGGYQPPP GYYDNYRY-S MEMNDGVRRK
Bf_N001     MSAGLVPPRV GEMSVYGSPY AGQGYITSFG ADPSAFSPLA NPSAGLQQPT AYYDPYPYGS YGAMDGTRRK
Bf_Nr003    MSAGLVPPRV GEMSVYGSPY AGQGYIASFG ADPSAFSPLA NPSAGLQQPT AYYDPYPYGG YGPMDGTPRK
Bf_Nr002    MSAGLVPPRV GEMSVYGSPY AGQGYIAAFG GDPSAFSPLG NPRANLQQPT AYYDPYPYGS YAAMDGARRK
Mm_Irx4     MSLATARHEL NSAGVYGSPY GSQGYGVTYG SEASAFSL-- NSHAGLTAAA AYYEPYPYDR YGTVDGTRRK
Mm_Irx6     MASRLLGSAR PELGIYGAPY AAQSYPLTYG PEPPTLGALN PQAPSLTQPG AYYETYRYGG VELSSAGRRK
Nv_Irx      LPDRVLQHRV ASVVLPRPAF GEAGFSLGSG TGLSAFSPLA RARDSWQASR APFDPFRFSP IDLAAGARRK
Mm_Irx1     MSGAGAAAVT SVLGMYAAPY AAPNYSLPYA ADLSLFSQMG SQPATFHTTP AY---YPYGQ FQYGDPGRPK
Mm_Irx3     MS---ASGSL SNVSVYGAPY AAQGYGLPYA TELPIFPQLG AQHPTAHPHP AF---YPYGQ YQFGDPSRPK
Mm_Irx2     MSSALAAPRS EELGSAFSPY PATGFGLQYS ADAAGFSYVG SPTTGMTGAI SYYGSY---P YQLNDPAYRK
Mm_Irx5     MSSVISGPRT DELGSAFSPY ASPGYNLQYG ADAAAFSYVG SPTPGMAGSL GYYAA-PLGS YPYGDPAYRK
Bs_Nr002    MSSALQPAEL PPHSLYNSPY AGQGSGTTSG SHPNSFPVMV --RR------ --YLSYGLRS LEAVV-GPRT

           71
Bs_IrxB     NATRETTSTL KAWLNEHRKN PYPTKGEKIM LAIITKMTLT QVSTWFANAR RRLKKENKMT WSPRNRCGED
Hr_Nr001    NATRETTSTL KAWLNEHRKN PYPTKGEKIM LAIITKMTLT QVSTWFANAR RRLKKENKMT WSPRNRCGED
Mo_Nr001    NATRETTSTL KAWLNEHRKN PYPTKGEKIM LAIITKMTLT QVSTWFANAR RRLKKENKMT WSPRNRCGED
Cr_Nr001    NATRETTSTL KAWLNEHRKN PYPTKGEKIM LAIITKMTLT QVSTWFANAR RRLKKENKMT WSPRNRCGED
Pm_Nr001    NATRETTSTL KAWLNEHRKN PYPTKGEKIM LAIITKMTLT QVSTWFANAR RRLKKENKMT WSPRNRCGED
Bs_Nr003    NATRESTNTL KAWLQEHKKN PYPTKGEKIM LAIITKMTLT QVSTWFA--- ---------- ----------
Mo_Nr002    NATRESTNTL KAWLQEHKKN PYPTKGEKIM LAIITKMTLT QVSTWFANAR RRLKKENKMT WVPKNRTNDS
Hr_Nr002    NATRESTNTL KAWLQEHKKN PYPTKGEKIM LAIITKMTLT QVSTWFANAR RRLKKENKMT WVPKNRANES
Pm_Nr002    NATRESTNTL KAWLQEHKKN PYPTKGEKIM LAIITKMTLT QVSTWFANAR RRLKKENKMT WVPKNRSNEN
Bf_N001     NATRDATSTL KAWLNEHRKN PYPTKGEKIM LAIITKMTLT QVSTWFANAR RRLKKENKMT WSPRNRCGDE
Bf_Nr003    NATRDATSTL KAWLNEHKKN PYPTKGEKIM LAIITKMTLT QVSTWFANAR RRLKKENKMT WSPRNKSGDE
Bf_Nr002    NATRETTATL KAWLMEHRKN PYPTKGEKIM LAIITKMTLT QVSTWFANAR RRLKKENKMT WSPRNRSGDS
Mm_Irx4     NATRETTSTL KAWLQEHRKN PYPTKGEKIM LAIITKMTLT QVSTWFANAR RRLKKENKMT WPPRNKCADE
Mm_Irx6     NATRESTSAL KAWLHEHRKN PYPTKGEKIM LAIITKMTLT QVSTWFANAR RRLKKENKMT WAPKNKGGEE
Nv_Irx      NATRETTSTL KAWLFEHRKN PYPTKGEKIM LAILTKMTLT QVSTWFANAR RRLKKENKMT WSPRNRCGEK
Mm_Irx1     NATRESTSTL KAWLNEHRKN PYPTKGEKIM LAIITKMTLT QVSTWFANAR RRLKKENKVT WGARSKDQED
Mm_Irx3     NATRESTSTL KAWLNEHRKN PYPTKGEKIM LAIITKMTLT QVSTWFANAR RRLKKENKMT WAPRSRTDEE
Mm_Irx2     NATRDATATL KAWLNEHRKN PYPTKGEKIM LAIITKMTLT QVSTWFANAR RRLKKENKMT WAPRNKSEDE
Mm_Irx5     NATRDATATL KAWLNEHRKN PYPTKGEKIM LAIITKMTLT QVSTWFANAR RRLKKENKMT WTPRNRSEDE
Bs_Nr002    KAPREATSIL KSWLNDHKDN PYPTKSEKMT LSIVTHMTLT QVSTWFANAR RRIKKENKMT WTPKTRSGER

          141
Bs_IrxB     DDDDIIEGED ENNHRSAPED QDVDLESPTK PKWSLAECAA SPTPHLHPAM NPHATLRNWM DMHQMSGPQG
Hr_Nr001    DDDDIMDGDD DNNSSNADGE EEIDLESPLK PKWSLADVAT SPIPSHQTVE GFQATLRNWM DMQHMSGPHG
Mo_Nr001    DDDDMMDDGE GNSAKTSHED EEIDLESPTK PKWSLADVAT SSSRNPNMLA NPQEAIRNWV DMQQMAGPYG
Cr_Nr001    DDEDDLADDA GNHGAFSDGR DERRAGKPPK PKWSIADVAT S---SSQSDD KDMSNVRSWV GLHHLMGPQR
Pm_Nr001    DDDEDMDDGN HSIDKPSENN DTAQSEAPTK PKWSLADVAT SARRTEINFQ NDTGAVRSWM DLHQLSGPTG
Bs_Nr003    ---------- ---------- ---------- ---------- ---------- ---------- ----------
Mo_Nr002    DDRRESDDKD ADEQDAMMDN QHAPSDSPTN VRYSTANVSG SDRRNECRFT SPVTNLQNWV NFHGLQNHQE
Hr_Nr002    EDRRISDDKD IDESQNTSDK SQIKPGFPPT RRPSCNNSYT GTRRSECRFS SPVTSLQNWV NFQAMSGPNS
Pm_Nr002    NDKKEGDDED SQNDNSVFED ERVDSFTAPS SECSIGHSAS GPIDSTSPYQ APYGRNNANT DYYQAAGGVA
Bf_N001     RREDGSDNED DDDREGNSDD DNHGGLDENK PKWSLADTAT CQFSHHRPFD SAVTNLRQWV DFHRVSGTTF
Bf_Nr003    RREDGSDSEN ENDGGEDRNQ EDDRGLEENK PKWSLADTAT SQFPGHRPFD S-VASLRHWV NFSQLAGTAV
Bf_Nr002    EGGEGSDIEG DEDHGSDKDV EDVDLDDAPK PKWSLADTAT SLFSRHRPYD RPVANFQRWV SFSQLSGSST
Mm_Irx4     KRPYGEGEEE EAGGKDDKEL ELSDLEDETK PRWSLAHTAT ASGALDRHQD SPVTSLRNWV DFHDLLDPGR
Mm_Irx6     RKADSGEDSL GCLSQEARGL RLSDLEDGHK PRWSLAHTAA APRDSEGEED SP--AAKAFG NF--LQGPY-
Nv_Irx      KDGVDSDIDD DGDSDSTRDR ESPPREDAKR PNTSLVDTTN QSTAQNHVLD SPVKSLRKWV DFYTVAGSTV
Mm_Irx1     GALFGSDTEG DPEGDQSNED E-------HSK PKWSLAETAT SFLPSHGLYT CHIGKFSNWT NFLAVSAPQQ
Mm_Irx3     GNAYGSEREE EDEAGRRRRD EEIDLENLQK PKWSLAETAT SAAAAHRLVS APLGKFPAWT NFPG----PHM
Mm_Irx2     DEDEGRSKEE SSDSGSECKD KFEDLEDASK PKWSLAEIAT SLLGRHLYYT SPYGNYTNYG NLNALHAASK
Mm_Irx5     EEEENKNDED EPQAGKETEG SLSDSDFLAK PKWSLAEIAT SVLSRPLYYT APYPGYTNYG SFGHLHGPSG
Bs_Nr002    DDNS-SNDED HDRSRDSTDG SLPSFDGNAN PD-----IRL SARAAHHPVT RDADAITSWM KFQHLAAAAT
```

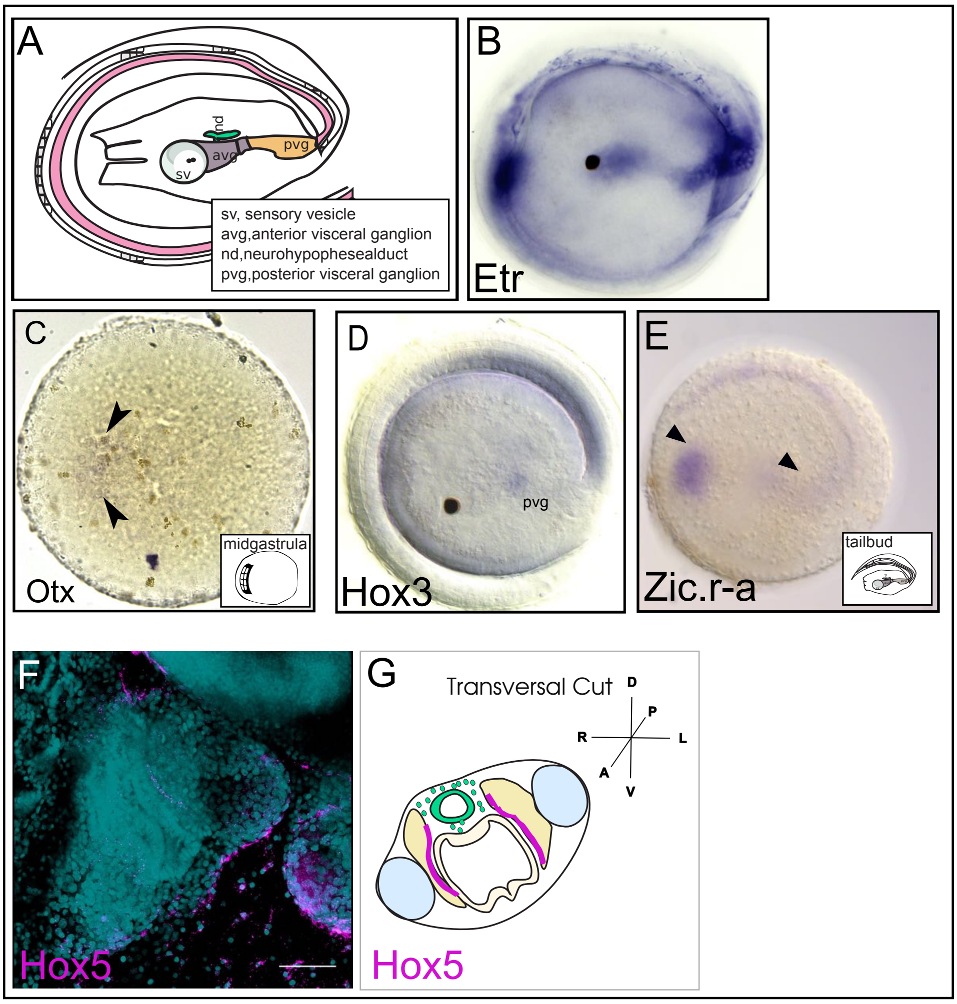
```
            211
Bs_IrxB      FATHALHAHP E
Hr_Nr001     FAHIAHHMQS S
Mo_Nr001     FAAHAMSMHN S
Cr_Nr001     FAMREKPHHS D
Pm_Nr001     FMQQVLNQRK D
Bs_Nr003     ---------- -
Mo_Nr002     SKTQSLDHTN S
Hr_Nr002     MGMRSYDTIS G
Pm_Nr002     ATTNAASMYA S
Bf_N001      VSTPAIALRT H
Bf_Nr003     HAEHAGVART H
Bf_Nr002     SPEHTGAPRP S
Mm_Irx4      NLVPQPPAGA S
Mm_Irx6      ---------- -
Nv_Irx       ITQSAPERRT S
Mm_Irx1      VALHGASARS S
Mm_Irx3      LGHPAAAAYA R
Mm_Irx2      AGSHSSHYRP P
Mm_Irx5      PGLAKKMLRS Q
Bs_Nr002     YPTHKRADRQ S
```


BsGli
Species legend:
Dm Drosophila melanogaster
Bf Branchiostoma floridae
Mm Mus musculus
Mo Molgula oculata
Cs Ciona robusta
Pm Phallusia mammillata
Hr Halocynthia roretzi
Bs Botryllus schlosseri
ML, LG, 100 bootstrap, gappyout
Dm_cuint
46
Bf_Gli001
47
Mm_Gli001
100
Mo_Gli001
Cr_Gli001
96
100
Pm_Gli001
47
Bs_Gli
72
Hr_Gli001
Bs_Gli002
99
Cr_Gli002
90
Hr_Gli002
0.7
Bs_Zic-r.a


BsGli
Species legend:
Dm Drosophila melanogaster
Bf Branchiostoma floridae
Mm Mus musculus
Mo Molgula oculata
Cs Ciona robusta
Pm Phallusia mammillata
Hr Halocynthia roretzi
Bs Botryllus schlosseri
ML, LG, 100 bootstrap, gappyout
Dm_cuint
46
Bf_Gli001
47
Mm_Gli001
100
Mo_Gli001
Cr_Gli001
96
100
Pm_Gli001
47
Bs_Gli
72
Hr_Gli001
Bs_Gli002
99
Cr_Gli002
90
Hr_Gli002
0.7
Bs_Zic-r.a

Primerlist

Primer	seq
Otx Fw	CAGCAGCAACAGAACCAAAA
Otx Rev	GCACCAGAGTACCCGTTGAT
IrxB-Fw	GACGAAGACGCCAAGCAATC
IrxB-Rev	CTGGGACCCTCGGAAAAGAC
Gli-Fw	AGCGGAGGAGACAAATGCTA
Gli-Rev	GAATGCGAGGGAGATATGGA
Pou3 Fw	Published in Ricci et al 2016
Pou3 Rev	Published in Ricci et al 2016
Hox3 Fw	TCAGCCGAGGACTCATTTCT
Hox3 Rev	GGATTGTGATGATGCTGGTG
Pax3/7 Fw	GCCCAATCCGAATATGCTGG
Pax3/7 Rev	ACTCTGGTGAAGTGCGATGA
Zic-r.a Fw	CAACTGCAGCAAAGTGTCT
Zic-r.a Rev	ATGGCCTTTCACCGGTATGA
Ebf	TGCTGCGACCATAAAAGTTG
Ebf	GCCGACTGCAACTGACTGTA
Etr	CAACGTAATTCCGGGACCTTC
Etr	GCGTTAAGTGCGCCAGTGTA
Notch	GAAGAGGAGTCCTGCCTGTG
Notch Rev	GCTCGTACGGTGTGGAAAAT
Pou4	AGCTCAGCAATCCGTCTCAT
Pou4	ATGCTTCGAGTGATCGCTTT


A
nd
avg
pvg
sv
sv, sensory vesicle
avg,anterior visceral ganglion
nd,neurohypopheseal duct
pvg,posterior visceral ganglion
B
Etr
C
Otx
midgastrula
D
pvg
Hox3
E
Zic.r-a
tailbud
F
Hox5
G
Transversal Cut
D
P
R
L
A
V
Hox5