bioRxiv preprint doi: https://doi.org/10.1101/237388; this version posted December 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

CHARACTERIZATION OF THE bHLH FAMILY OF TRANSCRIPTIONAL REGULATORS IN THE ACOEL *S. roscoffensis* AND THEIR PUTATIVE ROLE IN NEUROGENESIS

Perea-Atienza, E[1]., Sprecher, S.G.[3]., and Martínez, P[1, 2, *]

[1]Departament de Genètica, Universitat de Barcelona, Av. Diagonal 643, 08028 Barcelona, Spain.

[2]Institut Català de Recerca i Estudis Avancats (ICREA), Barcelona, Spain.

[3] Department of Biology, University of Fribourg, 10, ch. Du Musée, 1700 Fribourg, Switzerland

*Corresponding author (pedro.martinez@ub.edu)

KEYWORDS: bHLH, Acoela, Neurogenesis, Embryo, Patterning

ABSTRACT

Background. The basic Helix loop helix (bHLH) family of transcription factors is one of the largest superfamilies of regulatory transcription factors and are widely used in eukaryotic organisms. They play an essential role in a range of metabolic, physiological, and developmental processes, including the development of the nervous system (NS). These transcription factors have been studied in many metazoans, especially in vertebrates

bioRxiv preprint doi: https://doi.org/10.1101/237388; this version posted December 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

but also in clades such as the cnidarians and sponges. However, nothing is known about their expression in the most basal bilaterian group, the xenacoelomorphs. Recently, our laboratory has characterized the full complement of bHLH in the genome of two members of the Xenacoelomorpha, the xenoturbellid *X. bocki* and the acoel *S. roscoffensis*. Understanding the patterns of bHLH gene expression in members of this phylum (in space and time) would provide us with new insights into the conserved roles of the bHLH and their putative specificitiesin this group. Our focus is on deciphering the specific roles that these genes have in the process of neurogenesis.

Results: Here, we analyze the developmental expression of the whole complement of bHLH genes identified in the acoel *S. roscoffensis.* The members of bHLH class A seem to have specific conserved roles in neurogenesis, while other members of this and other classes have taken on more generalized functions. All gene expression patterns are described in embryos and early juveniles.

Conclusion: Our results suggest that the main roles of the bHLH genes of *S. roscoffensis* are evolutionarily conserved, with a specific subset dedicated to patterning the nervous system: SrAscA, SrAscB, SrHes/Hey, SrNscl, SrSrebp,SrE12/E47 and SrOligo.

INTRODUCTION

The basic helix-loop-helix (bHLH) proteins constitute a superfamily of transcription factors that are widely present in eukaryotes and play an important role in metabolic, physiological and developmental processes (Gyoja, 2017; Ledent and Vervoort, 2001; Massari and Murre, 2000; Simionato et al., 2007). Members of the bHLH superfamily

bioRxiv preprint doi: https://doi.org/10.1101/237388; this version posted December 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

encode for proteins containing a characteristic bHLH domain 60 amino acids long that includes a N-terminal DNA binding basic (b) region followed by two α-helices connected by a loop region (HLH) of variable length. The HLH domain promotes dimerization, allowing the formation of homo- or heterodimeric complexes between different family members. Some bHLHs also include additional domains involved in protein–protein interactions such as the 'leucine zipper', PAS (Per–Arnt–Sim) and the 'orange' domains (Ledent and Vervoort, 2001; Simionato et al., 2007).

The complex group of metazoan bHLH transcription factors have been classified, using molecular phylogenetic analysis, into 48 orthologous families (45 different families *sensu* Simionato et al., 2007 and three new ones *sensu* Gyoja et al., 2012. Based on phylogenetic affinities and general biochemical properties, the orthologous families are organized into six "higher-order" groups named A to F (Atchley and Fitch, 1997). Proteins of group A, only represented in metazoans, have roles mostly related to developmental processes, including neurogenesis and myogenesis (e.g. Achaete-scute or MyoD proteins). Two important monophyletic families are included in group A: the Atonal-family and the Twist-family. The members of groups A and B have bHLH domains with the ability to bind to a consensus hexanucleotide, the "E boxes", CANNTG (CACCTG or CAGCTG (Group A) and CACGTG or CATGTTG (Group B)) (Dang et al., 1992; Ledent and Vervoort, 2001; Murre et al., 1989; Van Doren et al., 1991). The sequences of the group B proteins are characterized by the presence, in addition to the bHLH domain, of a leucine-zipper domain (e.g., Myc, Max, USF, SREBP, MITF). They are widely represented in plant, fungal and metazoan genomes. The different members of this group have roles in a huge variety of metabolic, cell cycle and developmental processes (among others). Group C is

bioRxiv preprint doi: https://doi.org/10.1101/237388; this version posted December 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

characterized by the inclusion of a PAS domain and/or a PAS-3 domain (domain named after the first three proteins identified from this group: Per, ARNT and Sim). HLHs from this group are involved in environmental homeostasis and developmental signaling (Jones, 2004; Ledent and Vervoort, 2001; Simionato et al., 2007). Recently, a member of this group has been described in the filasterean *Capsaspora owczarzaki* (a protist),suggesting that representation of this group is not only restricted to the diverse metazoans (Sebé-Pedrós et al., 2011). Group D proteins are not able to bind to DNA because they lack the bHLH basic region. Instead, these proteins form heterodimers with the basic HLH domain of members from group A, acting as negative regulators of group A proteins (e.g. Emc or Id) (Jones, 2004; Ledent and Vervoort, 2001; Simionato et al., 2007). Group E bHLH in metazoans contain only two protein families, Hairy and enhancer of split (HES and HEY), the sequences of which include the 'orange' domain and the peptides WRPW or YRPW. They are known as developmental regulators. The last group, F, is constituted by proteins that lack the basic domain but contain a COE (Collier/Olf-1/EBF) domain involved in both dimerization and DNA binding. Phylogenetic analysis of this group is always difficult because the HLH sequences are highly divergent (Jones, 2004; Ledent and Vervoort, 2001; Simionato et al., 2007).

As stated above bHLH transcription factors have been described in many species from several metazoan clades. In the last ten years, since the pioneering study of the origin and the diversification of the bHLH carried out by Simionato et al. in 2007, new full sets of bHLH have been identified via the thorough analysis of many metazoan genomes, including those of the coral *Acropora digitifera*, the planarian *Schmidtea mediterranea*, the placozoan *Tricoplax adhaerens* and the calcisponge *Sycon ciliatum* (Cowles et al., 2013;

bioRxiv preprint doi: https://doi.org/10.1101/237388; this version posted December 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Fortunato et al., 2016; Gyoja, 2014; Gyoja et al., 2012). Although much has been learnt on the composition and evolutionary pattern of diversification over this last decade, knowledge of patterns of gene expression for individual members (in time and space) remains scant, and mostly focused on a few members of, fundamentally, the A and B superfamilies. In this context, the characterization of full complements of the bHLH and analyses of their expression domains in different metazoan phyla will give us the necessary insights required to fully understand how these gene superfamilies and their functions have changed over evolutionary time. With this general task in mind, our laboratory recently identified the gene members of this group of bHLH genes in two species from the monophyletic group Xenacoelomorpha: the xenoturbellid *Xenoturbella bocki* and the acoel *Symsagittifera roscoffensis* (Perea-Atienza et al., 2015), for which genomes and transcriptomes have recently been sequenced (unpublished).

Xenacoelomorpha is a phylum constituted by small, mostly marine, benthic worms divided into three clades: Xenoturbellida, Nemertodermatida and Acoela (Phillipe et al.,2011). The latter two are grouped into the clade Acoelomorpha, the sister group of Xenoturbellida. For a long time the phylogenetic affinities of these clades have been a matter of intense debate see (Hejnol and Lowe, 2015; Hejnol et al., 2009; Philippe et al., 2011; Ruiz-Trillo and Paps, 2016). However, the latest phylogenetic analysis carried out by Canon and colleagues (2016) seemed to resolve this conflict, proposing, based on analysis with strong support, a monophyletic Xenacoelomorpha as the sister group to all remaining bilaterians. All xenacoelomorphs share a relatively “simple” morphology (reviewed in (Achatz and Martínez, 2012; Haszprunar, 2016): they are bilaterally symmetrical; their bodies are surrounded by a ciliated epithelium with a mouth being the only digestive

bioRxiv preprint doi: https://doi.org/10.1101/237388; this version posted December 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

opening to the exterior. Common features found in other metazoans such as the presence of circulatory or excretory systems, anus and coelom are completely absent in this clade (Achatz et al., 2013; Hejnol and Pang, 2016; Nakano, 2015). The different studies carried out on members of this phylum suggest the urgent need to develop systematic molecular and developmental studies, given the poor understanding that we have regarding the molecular underpinnings of their development. Needless to say, a better understanding of the genetic control of developmental processes in xenacoelomorphs should provide us with key new insights into the origins and diversification of the bilaterians (Cannon et al., 2016; Gavilán et al., 2016; Perea-Atienza et al., 2015). In this context we carried out a systematic characterization of a well-known superfamily of transcriptional regulators, the bHLH factors, a family with many roles in the development and patterning of bilaterian tissues. There are several reasons why xenacoelomorphs are an interesting set of biological systems in which to carry out comparative molecular and developmental studies. The first one, as stated above, is the fact that they all belong to a monophyletic group (sharing a common ancestor) and the second one is that, strikingly, within this clade the members show a high diversity in the complexity of many regulatory families and also in the organization of anatomical architectures (i.e. nervous system, arrangement of musculature, position of the mouth, the morphology of copulatory apparatus, etc.), which allow us to compare different ways of "constructing" and patterning organ systems within a set of interrelated animals. Moreover, what is learnt from the study of xenacoelomorphs is of evolutionary significance, since their phylogenetic position within the metazoans, as the most basal bilaterian animals, provides us with a unique window to the events happening at the origin of Bilateria. We should point out that in addition, a significant practical advantage of studying the xenacoelomorphs is the fact that they possess fewer cell types and organs than

bioRxiv preprint doi: https://doi.org/10.1101/237388; this version posted December 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

most bilaterian animals, making their system more amenable to our (comprehensive) research efforts. Hence, for these evolutionary and practical reasons, xenacoelomorphs are an ideal system in which to study the bHLH genes and their expression patterns, as a model for understanding the evolution of gene regulatory factor families and their involvement in the regulation of tissue development (in particular the nervous system, see below).

One of the key innovations linked to the emergence of Bilateria is the origin of centralized nervous systems (brains and cords). How these compact brains are assembled from simpler nerve nets remains a matter of debate (but see: Martín-Durán et al., 2017) for a recent reassessment), a debate that is mostly grounded in the lack of knowledge that we have on the molecular mechanisms that differentially control the development and assembly of nerve nets, cords and compact brains (see, for instance, (Gavilán et al., 2016; Hejnol and Lowe, 2015; Martín-Durán et al., 2017; Northcutt, 2012; Parker, 1919). The comparative approach should provide an answer. In this context the use of members of the Xenacoelomorpha is ideal, since they have different nervous system morphologies (with variable degrees of condensation) all derived from a single, common, ancestor. Briefly, in the most basal group, the xenoturbellids, a unique basiepithelial nerve net surrounds the animal body while acoels (last divergent group) have an anteriorly concentrated nervous system (Achatz and Martínez, 2012; Perea-Atienza et al., 2015) for a comprehensive review, see: Martínez et al., 2017). The nervous system architecture of the most divergent class of acoels (Crucimusculata) represents the first instance of the acquisition of a compact brain in bilaterian evolution.

The fact that many members of the bHLH families are involved in the development and patterning of the nervous system in many bilaterians (but also in cnidarians) makes the

bioRxiv preprint doi: https://doi.org/10.1101/237388; this version posted December 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

study of this superfamily particularly relevant in this context. We have learned over recent years that several bHLH genes, mostly from group A, have a specific role during neurogenesis (so-called proneural genes, belonging to Atonal, Neurogenin and Achaete-scute families) (see for instance: (Guillemot et al., 1993; Jarman et al., 1993; Sun et al., 2001) also reviewed in Bertrand et al., 2002 and Gyoja, 2017). These genes were cloned and analyzed in the context of a thorough characterization of the bHLH in the xenacoelomorph genomes. However we have been lacking key information on where and when these different members are expressed during development and which ones are involved in the development of the nervous system. In order to understand the roles of bHLH genes in xenacoelomorphs we conducted a thorough analysis of their expression patterns in one member of the Acoela, the species *Symsagittifera roscoffensis*, using both colored *in situ* hybridization (ISH) in embryos and juveniles and double fluorescence *in situ* hybridization (FISH) in juveniles. We provide here a description of all of the bHLH gene expression patterns, in space and time, and discuss their role in neurogenesis. Our data is also discussed in the context of the evolution of neural system patterning.

Materials and Methods

bHLH sequence identification

Symsaggitiffera roscoffensis DNA sequences were extracted from the genome assembly and the embryo (mixed stages) transcriptome and were published in Perea-Atienza et al., 2015. In the mentioned analysis, sequences were extracted using HMMER (v.3.0) (Eddy, 2011) and classified by phylogenetic methods.

bioRxiv preprint doi: https://doi.org/10.1101/237388; this version posted December 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Sampling

Adult specimens of *S. roscoffensis* where collected in Carantec (Brittany, France) during one of their reproductive periods (in April 2016). They were cleaned and placed in petri dishes with fresh sea-water until they spawned the cocoons in the media. Every cocoon contains several embryos, mostly synchronized (between ten and twenty). We collected the cocoons at different time intervals, with the time when we detected the fresh spawn being the 'zero' time. The other times corresponded to the number of hours elapsed between the detection of fresh spawn and fixing the cocoons for in situ analysis. For instance, a sample from 0 to 6 hours contained embryos that developed within the first 6 hours. Embryos were separated from the adults by filtering and then treated using 0.01% Pronase (Sigma) and 0.1% thioglycolate (Sigma) in seawater to permeabilize the egg shells. After cleaning them with sea-water, embryos were fixed in 4% formaldehyde (methanol free) overnight at 4ºC. A large proportion of cocoons were not fixed, with the aim of obtaining hatchlings later on. These hatchlings were relaxed using 7% magnesium chloride and fixed, as were the embryos. Some specimens were allowed to develop to older juveniles. After fixation all samples were cleaned three times in 1x PBS and dehydrated progressively with a methanol series (25% – 50% – 75% methanol in PBS).

Gene cloning and in situ hybridization

Dig-labeled and fluorescein-labeled RNA probes were synthesized using the Dig-RNA labeling kit from Sigma, following the manufacturer's instructions (Id_number11277073910, Sigma-Aldrich). After precipitation, the riboprobes were diluted in hybridization buffer to a final working concentration of 2-1 ng ⁄ µl (depending of the probe). In situ hybridization on whole embryos and juveniles was performed following the

bioRxiv preprint doi: https://doi.org/10.1101/237388; this version posted December 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

protocol published by Semmler et al., 2010 with a few modifications: 1) the main solvents, PB-Tween and PB-Triton were replaced with TNT (0.1M TRIS-HCl, pH7.5+ 0.15M NaCl+ 0.05% Tween-20 detergent in RNAse-free water); 2) the proteinase K and glycine steps were suppressed, in order to reduce the damage produced to the samples; 3) before the pre-hybridization overnight, we included a step in which we increased the temperature to 80º C during one of the two last washes with hybridization buffer (HB), to reduce the background. We incubated the sample with the corresponding probe for at least 3 days. The hybridization temperatures were between 55ºC and 61ºC depending on the probe. The specimens were mounted in 70% glycerol and analyzed using a Zeiss Axiophot microscope (Carl Zeiss MicroImaging GmbH) equipped with a Leica DFC 300FX camera.

Colorimetric *in situ* protocols highlight the domains of highest gene expression. For detailed aspects of the patterns, we rely on the more sensitive fluorescence *in situ* alternatives.

Double fluorescence *insitu* hybridization

Juveniles of *S. roscoffensis* used in FISH analysis needed a photo-bleaching treatment step after the re-hydration. In order to reduce background due to auto-fluorescence, as much as possible, we immersed the specimens in a solution of 1.5% hydrogen peroxide, 5% formamide and 0.5% 0.5xSSC in water (RNAse free), during 15 min and under a white light. Samples were washed twice in PBT (1x PBS+0.1%Triton X-100). The following protocol is a short version of the ISH protocol that it is described in detail in Perea-Atienza et al., 2018). For FISH all the hybridization probes where diluted to 1 ng/µl. Afterwards, the samples were incubated in anti-Dig-POD 1:500, overnight at 4ºC (Sigma-Aldrich, Id_number11207733910). More than four washes, over 2 to 3 hours, in MAB-TritonX-100

bioRxiv preprint doi: https://doi.org/10.1101/237388; this version posted December 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

0.1%were used to eliminate the rest of the antibody. The signal was developed in TSA red 1:300, in the so-called TSA Buffer (solution of 2M NaCl + 100mM Borate buffer, pH 8.5), over 2 to 6 hours. To stop the development of the signal, samples were washed in PBT. The antibody quenching was made using 1% H_2O_2 in PBT 0.1% for 45 min at RT. After washing 2 times with PBT, a 2nd quenching step was done with 2xSSC, 50%Formamide and 0.1%TritonX-100 for 10 min at 56ºC. The samples were washed twice in PBT and blocked again, previous to the incubation with the second antibody: anti-DNP-HRP 1:200. After the antibody wash, with MAB-TritonX-100 0.1%, the signal was developed in TSA green 1:300 in TSA Buffer. After the double or single FISH protocols, and in the cases required, we proceeded to combine this procedure with the immunostaining of the samples (using the species-specific anti-synaptotagmin antibody; as a reference, pan-neuronal marker), as explained in the following section.

It is important to note that the general absence of well-defined tissues and the presence of nuclear intermingling in most of the "parenchymal" (external-mesodermal; internal-digestive) tissues of the Acoela makes specially difficult to perform *in situ* hybridization and interpret detailed patterns.

Immunohistochemistry

Immunostaining was performed using the protocols outlined in Achatz and Martínez, 2012. *S. roscoffensis* specimens were incubated in primary anti-synaptotagmin (dilution 1:500) antibodies (previously pre-absorbed) and reacted with the secondary antibody [Alexa Fluor goat anti-rabbit 532 (Molecular Probes, Eugene, OR)]. The anti-synaptotagmin antibody was raised in our laboratory using the specific *S. roscoffensis* sequence from a transcriptome analysis (see: Perea-Atienza et al., 2015). More recently, we raised a new

bioRxiv preprint doi: https://doi.org/10.1101/237388; this version posted December 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

antibody against the whole SrAscA protein (synthesized from the sequences in our transcriptomes). This protein is the ortholog of the proneural, mammalian, gene MASH1 or the *Drosophila* Achaete-Scute. For all antibodies a test of specificity was done using ELISA.

Embryo cell counting

Given the difficulty of staging the embryos in the laboratory, we relied, as a good approximation for developmental time, on the number of cells. Embryo cell counting was performed by incubation of all samples for 10 minutes with Dapi. A sample of 12 individuals from a pool of 12 to 24 hours' post-fertilization embryos were scanned completely using a Leica SP2 confocal laser microscope and their stacks were processed using the software Fiji (Schneider et al., 2012; Schendelin et al., 2012) with the plugin 'cell counter' (author: Kurt de Vos; see: https://imagej.net/Cell_Counter).This procedure allowed us to obtain a precise count of the total number of cells per embryo when performing *in situ* procedures.

Results and Discussion

We carried out a detailed study of the expression of the whole complement of *S. roscoffensis* bHLH genes during development, using early embryos (12 to 24hours post fertilization; average 250 cells/embryo) and hatchlings (time window: 12 to 24 hours post hatch). Taking into account the timing of development of S. *roscoffensis* (and that of other acoels; see Henry et al., 2000; Ramachandra et al., 2002; Semmler et al., 2010 for reference time-frames), we selected these stages as good starting points for the characterization of the

bioRxiv preprint doi: https://doi.org/10.1101/237388; this version posted December 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

nervous system's development and hence they should provide us with some initial insights into the roles played by the different bHLH transcription factors in this acoel species. Before discussing further details of the expression patterns, it is first necessary to note here the experimental limitations associated with the acquisition and staging of the embryo samples, which led us to consider time windows of 12 hours instead of exact time points. A few samples were collected within the first few hours after spawning, however the staining patterns for all analyzed genes were either absent or very faint (irrespective of the probe concentration or staining/developing color time). It is interesting to note that no expression was visible in many other embryos in which bHLH genes were analyzed at the earliest stages (Li et al., 2014; Reitzel et al., 2013; Richards and Rentzsch, 2014). For these reasons we decided not to focus our analysis on earlier embryos (most of the patterns were visible, or more intense, in the biggest embryos at later stages). bHLH expression domains in *S. roscoffensis* were revealed by whole mount colorimetric in situ hybridization (ISH), using probes from 17 different genes, in juveniles and post grastrula embryos. Of these 17 genes, we were not able to obtain expression patterns for NeuroD, PTFa1, PTFa2, PTFb1 or PTFb2 in any of these stages (with the exception of PTFb1 which was expressed in embryos, although the expression levels were very weak). The absence of expression of PTF family was due, most probably, to their relatively low expression levels (also correlated with the low numbers of these transcripts in our EST database). In the case of NeuroD, was impossible to clone the sequence, most probably due to an annotation problem in the genome. Recently, this gene expression has been analyzed in nemertodermatids, however, no pattern has been reported on acoels (Martín-Durán et al., 2017). All the other genes, which showed clear *in situ* patterns, are presented in Fig. 1. For most of the expressed genes, the patterns were always stronger in post gastrulation embryos

bioRxiv preprint doi: https://doi.org/10.1101/237388; this version posted December 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

than in juveniles (probably also due to the easier accessibility of probes to the interior of the embryos and/or higher expression levels). Many of the analyzed genes were expressed in restricted patterns in the embryos, whereas others were expressed more widely within the embryos and/or juveniles. Detailed descriptions of each gene's pattern is given below.

Some bHLH transcription factors are known to be specifically involved in the process of neurogenesis, especially those in group A. Given the bHLH expression patterns previously obtained by colorimetric ISH experiments, and taking into account their neurogenic role in metazoans, we analyzed some of the bHLH genes in more detail, using double fluorescence in situ hybridization (FISH). This latter approach is more sensitive and allows us a higher resolution analysis of the pattern. The FISH experiments were all carried out in juveniles, with the aim of verifying co-expression domains within the nervous system (FISH did not work properly in embryos). To determine the domain of the nervous system, we use a reference marker, the nervous system pan-neuronal gene α-synaptotagmin (see also, in Fig.2, the expression of mRNA and antibody of synaptotagmin specifically in the nervous system). However, as α-synaptotagmin mRNA encodes for a synaptic protein (a terminal differentiation marker), we detected no expression in embryos prior to 24hoursafter fertilization. The first clear signs of the differentiated nervous system were detected, using this marker, in 36–48 hour post-fertilization embryos, showing a bilateral pattern representing the future two anterior brain lobes (Fig. 2) (Gavilán et al., 2015; Perea-Atienza et al., 2015). In the following paragraphs we discuss the bHLH expression patterns obtained by ISH in the acoel *S. roscoffensis* with a special focus on all genes with a putative function in the nervous system (as reported in other animals).

<u>bHLHgene developmental expression patterns</u>

bioRxiv preprint doi: https://doi.org/10.1101/237388; this version posted December 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Resemblance between patterns in different model animals suggests similarities in gene functions. Although not all *S. roscoffensis* bHLH genes share similar patterns with their homologs in bilaterians or metazoans, some of them seem to share similar expression domains, hence pointing to a conservation of roles over evolutionary time. In the next section we analyze these patterns by following the grouping established for the bHLH genes. We describe here the expression patterns of 13 genes as detected by ISH.

Genesfamiliesfrom group A: Achaete-scute, E12/E47, Twist, Net, Nscl and Oligo

Of the whole complement of *S. roscoffensis* bHLH genes, nine were classified within group A: SrAscA, SrAscB, SrAsc_like, SrE12/E47, SrNet, SrNscl, SrOligo/Beta (finally classified as an Oligo ortholog) and SrTwist (see (Perea-Atienza et al., 2015) for general classification of bHLH genes). The expression patterns belonging to: SrAscB, SrAsc_like, and SrE12/E47 were clearly defined and were all located in the anterior part of the juvenile body and also most likely in the animal pole of the post gastrula embryo (future anterior part of the juvenile; see also: (Ramachandra et al., 2002), with the exception of the SrASC_like gene in juveniles (an unclassified member of the Achaete-scute family, see (Perea-Atienza et al., 2015), which was not expressed(or the expression was really weak to be detected (Fig. 1)).

The Achaete-scute family transcription factors are divided in two subfamilies: Achaete-scute A and B. The genes belonging to family A (also named in other clades as “proneural genes”) possess a highly conserved neurogenic role across a wide range of metazoans. It is well known that the Achaete-scute complex members perform a proneural function during embryogenesis and the development of adult sense organs in *Drosophila melanogaster* (Dambly-Chaudiere and Ghysen, 1987). This role is preserved in other metazoans such as

bioRxiv preprint doi: https://doi.org/10.1101/237388; this version posted December 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the beetle *Tribolium castaneun,* where TcAsh is also necessary for the formation of the neural precursor (Wheeler et al., 2003), and in mice, where the gene Mash-1 is essential for the generation of autonomic and olfactory neurons (Guillemot et al., 1993). Moreover, knockout of Mash-1 in mice leads to severe defects in neurogenesis in the ventral telencephalon and in the olfactory sensory epithelium (Casarosa et al., 1999; Horton et al., 1999). In addition, studies in the cnidarian *Nematostella vectensis* have shown that the homologous gene NvashA is specifically expressed in a differentiating subset of neural cell types of the embryonic ectoderm (Layden and Martindale, 2014; Layden et al., 2012; Richards and Rentzsch, 2015). In more recent studies, two members of the Achaete-scute family, Spol-AscA and Spol-AscB, in the planarian *Schmidtea polychroa* showed expression in some putative neural progenitors, though only Spol-AscB was detected in developing embryos and hatchlings (Monjo and Romero, 2015). Given the crucial function in the neural differentiation of the SrAscA orthologs (as exemplified in the Drosophila case: García-Bellido and de Celis, 2009) and the weak expression of this gene obtained by ISH, we decided to raise a specific antibody against this (acoel) bHLH protein, using the whole coding sequence. Interestingly, the antibody recognizes sets of clustered cells in the ectoderm of the animal (Fig. 3), whereas the mRNA pattern observed by fluorescence *in situ* hybridization shows expression in an extensive part of the nervous system, revealing the CNS (brain and cords) (Figs. 4A; 4C), the peripheral tracks and the nerve net (Figs.4B; 4D). The AscA expression domain includes the most posterior part of the cords, the area where they converge (Fig. 4E)

Different functions have been associated with the Achaete-scute gene orthologs belonging to subfamily B. These genes play diverse roles and in most cases, they are not directly

bioRxiv preprint doi: https://doi.org/10.1101/237388; this version posted December 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

involved in the development of the nervous system. In fact, most of them regulate the expression of downstream genes in different tissues; for example, the human and rat bHLH Hash-2/Mash-2 (respectively) are crucial for development of the placenta (Koide et al., 2003), and also, Ascl-2 in mice has been implicated in the regulation of intestinal stem cell fate (van der Flier et al., 2009) and the generation of a specific cell type within the immune system (Liu et al., 2014). In our study, we found that the expression pattern of SrAscB was restricted to a circular area during embryonic post grastrulation development (Fig. 1), while in the juveniles it was localized specifically in the half anterior region of the animal, in two bilaterally symmetric domains of cells that were connected by a track crossing the anterior-posterior body axis (Fig.1). The cell populations that express SrAscB mRNA clearly form part of the anterior neural cords (as can be observed by FISH, Fig. 4A, 4C and Fig.5B, 5C) and some commissures lightly inter-connected, which constitute part of the brain. At the posterior end of the expression domain a thick ventral track crossing the anterior-posterior body axis surrounds the area of the mouth (supplementary material 1). The structures that express SrAscB suggest a domain location lightly ventral (supplementary material 1). The location of expression in the embryo's prospective animal pole and in the anterior region of the acoel juvenile, in the brain area, suggests that SrAscB is indeed expressed in a neuronal population. Among metazoans, the huge variety in expression of the AscB homologs is remarkable. However some patterns seem to be clearly associated with the neural tissue. For instance, some studies reported expression of Mash2 (mammalian ortholog) in Schwann cells of the peripheral nervous system, where the authors define the role of Mash2 as a negative regulator of the proliferation of this glial cell type (Küry et al., 2002). Another example linking AscB expression with the nervous system is the aforementioned case of the neural homolog Spol-AscB in the planarian *S. polychroa* (Monjo and Romero, 2015).

bioRxiv preprint doi: https://doi.org/10.1101/237388; this version posted December 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

These examples show the putative involvement of members of this B group in neurogenesis, and reinforce the possibility that our homolog actually works specifically in neural cells.

Among metazoans, the E proteins play critical roles in cell growth, specification and differentiation, including the neurons. The bHLH Sr_E12/E47 gene showed a high level of expression in two circular domains of the embryo that later on became two anterior-medial regions located immediately posterior to the statocyst (Fig. 1). At first sight, our results seem to conflict with those obtained for the E12/E47 gene orthologs in, for instance, *H. sapiens* and *D. melanogaster* (Daughterless), in which expression is found in most tissues. However, we should point out that different studies have found an increment in the mRNA expression of E12/E47 (and other E2A mRNAs) in some areas of rapid cell proliferation and differentiation in several tissues, including neural tissue. These levels decrease progressively during neurogenesis, and become almost undetectable in the adult nervous system (Roberts et al., 1993). This is consistent with our observations, with the strongest expression pattern in the post grastrula embryo, and lower levels in the anterior body of the juvenile. The expression domain in juveniles also coincides with part of the brain (Fig. 6B, 6C, 6D). The pattern is consistent with an early role in the development of neurons, perhaps in populations of non-terminally differentiated neurons (neuroblasts). E12/E47 genes form heterodimers with other group A bHLH factors (Lassar et al., 1991; Massari and Murre, 2000). In fact, some authors have suggested a specific role of E-proteins in early neural differentiation (Heng and Tan, 2003) and, in the same vein, recent studies have confirmed that E-proteins orchestrate neural stem cell lineage progression (Fischer et al., 2014). bHLH genes encoding other E proteins with a similar role are seen in, for instance, the planarians,

bioRxiv preprint doi: https://doi.org/10.1101/237388; this version posted December 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

in which the gene e22/23 is expressed in the central nervous system (Cowles et al., 2013). What seems clear is that both genes Sr_AscB and Sr_E12/E17 are most probably associated with some bilaterally symmetrical structure or cell population that is undergoing differentiation.

In contrast to the mRNA expression patterns found in the above members of group A, the other members of this group (SrTwist, SrNet, SrNscl and SrOligo) showed diverse expression patterns mostly located in the middle body region. SrTwist was expressed in a discrete spatial domain well delimited in the embryo's animal pole and organized as a pair of bilateral bands on both sides of juvenile specimens. This is not new data, as previous studies from our laboratory that analyzed bilaterian mesodermal gene expression have already revealed the expression pattern of twist orthologs in embryo, juvenile and adult stages of *S. roscoffensis* (M. Chiodin, PhD thesis) and adults of the acoel *Isodiametra pulchra* (Chiodin et al., 2013, suggesting its expression in part of the gonads, the male copulatory organ (only in *I. pulchra*) and neoblasts, all of which are mesodermal derivatives. Furthermore, the mesodermal role of twist homologs, for instance, in *Drosophila*, is well known (Baylies and Bate, 1996).

The expression patterns of the remaining three genes, SrNet, SrNscl and SrOligo, showed a homogeneous distribution throughout the body and throughout the body in similarly dispersed, but slightly different, cell populations the expression domain of SrNet gene being slightly larger than that of SrOligo and SrNscl (Fig. 1). Different expression domains have been described for SrNet orthologs. For example, in the jellyfish *Podocoryne carnea*, Net is expressed in the entocodon, a mesoderm-like structure that gives rise to the striated and smooth muscle of the bell (Müller et al., 2003). Also, studies of the vein patterning of

bioRxiv preprint doi: https://doi.org/10.1101/237388; this version posted December 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

D. melanogaster wing indicate that DmNet acts as a transcriptional repressor and is required to maintain the inter-vein regions during development (Brentrup et al., 2000). In the case of *S. roscoffensis* SrNet was expressed in a region located in the middle of the embryo. In juveniles, the SrNet expression signal covered the statocyst, forming different lines in the anterior territory of the body, and being more intense than the SrOligo expression in the same domain (Fig.1). In agreement with this SrNet expression pattern, and taking into account the roles described for Net homologs in other animal models, we suggest that the acoel Net homolog may have a mesodermal role, without any significant participation in neurogenesis, although clearly further studies are required.

The mammalian transcription factors Olig1 and Olig2 are involved in the specification of the oligodendrocyte lineage. They are not only expressed in oligodendrocyte progenitors, as their expression has also been detected in other progenitor populations that first produce motor neurons and later oligodendrocytes. In several animal models, Olig genes act in combination with patterning and proneural genes to promote neurogenesis and/or oligodendrogenesis (reviewed in Bertrand et al., 2002; Lu et al., 2002; Mizuguchi et al., 2001; Novitch et al., 2001; Sugimori et al., 2007; Zhou and Anderson, 2002). Expression of SrOligo in 12–24 hour post-fertilization embryos of the acoel occurred in a scattered pattern covering most of the embryo, however we detected the highest level of expression in a smaller localized region (Fig. 1). Expression was mostly limited to the center of the juvenile body, being very low in the zone of the brain and with no signal in the posterior part of the organism (Fig. 1). FISH experiments shown that the low expression detected in the brain zone, corresponds to the area where the brain cords converge anteriorly (Fig. 6C, 6D, 6E). Additional to the central expression found in juveniles, SrOligo seems to be more

bioRxiv preprint doi: https://doi.org/10.1101/237388; this version posted December 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

expressed dorsally and most probably present in the nerve net and in a low range in the cords (Fig. 6C, 6E). Due to the conserved role of this gene family among different clades, and taking into account the absence of expression in some of the anterior part (specifically in the region anterior to the statocyst) where most of the nervous system of the acoel is located, more experiments (histological sections)will be necessary to confirm the specific cell populations that express SrOligo. Noteworthy similarities were found between the pattern in *S. roscoffensis* and that in the planarian *S.mediterranea*; in both cases a small anterior population of cells, separated from the rest of the domain, expressed the respective Oligo orthologs. Also, similar expression patterns have been observed in the gene expression of Hes, Srebp and the "low-expression" Ptf family in planarians (see Cowles et al., 2013).

Among metazoans, the contribution of Nscl gene orthologs to the development of sensory cells and neurons has been well-described (Begley et al., 1992; Kim, 2012). The bHLH gene SrNscl in the juvenile of *S. roscoffensis* showed ubiquitously low expression, in a domain that seemed to be entirely posterior to the statocyst. Interestingly this gene showed strong expression in an extensive region in our embryos. Nscl mouse's orthologs show strong expression during post gastrulation, most likely due to their role, among others, in early neurogenesis and in the specification of neurons (Göbel et al., 1992). Functional analysis of the Nscl-1 mouse ortholog demonstrated the need for this gene for correct neural cell differentiation, and in combination with Nscl-2, it is required to control the migration of neuronal precursor cells (Krüger et al., 2004; Schmid et al., 2007). Actually with the data obtained, we cannot conclude the expression of SrNscl within the nervous

bioRxiv preprint doi: https://doi.org/10.1101/237388; this version posted December 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

system. To determine whether there is any relation between SrNscl and neurogenic processes further experiments are needed.

Gene families from group B: Max and Srebp

The bHLHs genes belonging to group B and present in *S. roscoffensis* are SrMax and SrSrebp. The metazoan orthologs of Max are involved in cellular proliferation, development and differentiation. Several studies have shown that the MAX protein forms heterodimers with other transcription factors such as the bHLHs MYC and MAD (Gallant, 2006). MYC-MAX heterodimers are involved in the transcriptional activation of different target genes (Eilers and Eisenman, 2008; Hartl et al., 2010). In *S. roscoffensis*, MAX is the only known member of these complexes, since MYC, MAD and other related bHLH transcription factors such as MLX were not found in the genomes or transcriptomes (Perea-Atienza et al., 2015). This is interesting in itself since it suggests a developmental function for Max that is independent of heterodimerization. Expression of SrMax mRNA in juveniles is completely ubiquitous; however, this was not the case in embryos, where, even though the signal covered a large part of the embryo a differential, asymmetric expression domain was still observed (Fig. 1). Expression patterns of the related Hydra genes, Myc1 and Max, are located in the interstitial stem cell lineage in the gastric region. In addition, Max is expressed, but at a lower level, in the epithelium throughout the entire body column (Hartl et al., 2010).

The other member of group B in our species is SREBP. Srebp orthologs are involved in animal homeostasis, in the regulation of sterol metabolism. They regulate the gene

bioRxiv preprint doi: https://doi.org/10.1101/237388; this version posted December 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

expression of most of the enzymes involved in cholesterol biosynthesis. In the case of mice and rats, the Srebp gene orthologs are expressed ubiquitously, and at a low level, in several cell types, including astrocytes, oligodendrocytes and Schwann cells, and are very active in lipid metabolism (Camargo et al., 2009). Moreover, in the adult planarian *S. mediterranea* this gene shows intense expression over the whole body, where it is expressed in different cell types (Cowles et al., 2013). In this context the SrSrebp expression in *S. roscoffensis,* which was widespread in both stages analyzed (Fig. 2) would be consistent with having a similar role in metabolic regulation, perhaps including most supporting glial cells (see Bery et al., 2010 for evidence of glial cells in *S. roscoffensis*).

Gene families from group C: ARNT

The only representative of group C in our acoel species is ARNT (aryl hydrocarbon receptor nuclear translocator). ARNT can form heterodimers with several bHLH proteins and its function depends on its dimerization partners. For this particular reason, members of the ARNT family tend to be widely expressed (McIntosh et al., 2010). For instance, one of the most common complexes of the mammalian ortholog of ARNT is the heterodimeric complex with the protein AHR (aryl hydrocarbon receptor), a complex that mediates numerous biological and toxicological effects through the regulation of various AhR-responsive genes. Specifically, the AHR/ARNT system controls processes such as oxidation/anti-oxidation, epidermal barrier function, photo-induced response, melanogenesis, and innate immunity (Denison et al., 2011). This complex is thus present in many cell types, being detected ubiquitously in most metazoans (Cowles et al., 2013; Reitzel et al., 2013). In our genomic/transcriptomic analysis, we have found only the gene

bioRxiv preprint doi: https://doi.org/10.1101/237388; this version posted December 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

encoding for ARNT transcription factor, and not the gene for its partner AHR, contrary to what has been described in most metazoans (Perea-Atienza et al., 2015). The receptor AHR is also absent from the genome of the xenoturbellid *Xenoturbella bocki*, suggesting that this gene has probably been lost in all xenacoelomorphs, since both members of the AHR/ARNT system are present in cnidarian genomes (Gyoja et al., 2012; Simionato et al., 2007). SrArnt mRNA was expressed in the entire postgrastrula embryo with small areas inside the domain showing higher levels of expression (Fig. 1). In juveniles, probably due to the low level of signal, we detected a very faint expression, and widespread in the middle of the body (Fig. 1). This expression is consistent with that of other ARNT orthologs. For instance, studies carried out in the anthozoan *N. vectensis* described the expression of NvArnt, with an initial phase detected in the gastrula ectoderm, higher expression in the planula and widespread expression in different tissues in the juvenile (Reitzel et al., 2013). This suggests that ARNT in *S. roscoffensis* is probably expressed, though at low levels, in a variety of cell populations.

<u>Gene families from group E: HES/HEY</u>

As for the previous class, *S. roscoffensis* group E comprises a single HES/HEY subfamily member. The SrHes/Hey gene was expressed in the anterior-medial region of the embryo and later on, in the juvenile stage, its expression become more intense to an area in the central part of the body, posterior to the statocyst (Fig. 1; Supplementary material 2). A detailed expression showed by FISH locates the main domain of SrHes/Hey as a part of the cords (supplementary material 2; Fig. 5D) and at lower level in the brain (Fig. 5).Itis well known that HES proteins act by inhibiting proneural bHLH protein functions through a mechanism that involves the repression of proneural gene's expression (reviewed in

bioRxiv preprint doi: https://doi.org/10.1101/237388; this version posted December 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Guillemot, 2007). A reduction in the number of neural progenitor cells occurs in the absence of mouse Hes genes, in parallel with the premature neural differentiation of neuroblasts (Ishibashi et al., 1995; Kageyama et al., 2008). Moreover, knockdown of its planarian ortholog hesl-3 during regeneration leads to a reduction in the neural population and a miss-patterned brain (Cowles et al., 2013). The expression pattern obtained for the SrHes/Hey gene in juveniles of *S. roscoffensis* resembles that of the three Hes gene orthologs in planarians (hes-1, hes-2 and hes-3), which are not expressed in the most anterior part of the body but are otherwise highly expressed in the central part of the adult body, although only hes-3 is expressed in a clear domain in the CNS (Cowles et al., 2013). However, the expression of Hes genes in neural progenitor cells seems to be well conserved among metazoans.

We have described in the above paragraphs the patterns of expression of all genes encoding bHLH transcription factors in the acoel *S. roscoffensis* genome, and have contrasted those domains with those reported for their homologs from other metazoans, with the exception of NeuroD and the PTF family orthologs for which we don't have data. Owing to the similar expression profiles and the relevance of some of these genes in the development of the nervous system in metazoans, we conclude that they show also conserved roles in the acoels. This statement, although perhaps unsurprising, is worth attention since acoels, in particular *S. roscoffensis*, are known for their long branches in all molecular phylogenetic analyses (Cannon et al., 2016; Hejnol et al., 2009; Philippe et al., 2011), suggesting that they might have highly modified/derived body plans. Needless to say high rates of molecular change do not always correlate with extreme morphological change (Subramanian et al., 2009).

bioRxiv preprint doi: https://doi.org/10.1101/237388; this version posted December 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

In the following section, we discuss in more detail the patterns of some bHLH that are specifically relevant to the development of the acoel nervous system.

Developmental expression of neural bHLH transcription factors in *S. roscoffensis*

Acoel embryos possess a unique early development program that is known as "duet spiral" cleavage (Henry et al., 2000). The first three micromere duets give rise to the ectodermal layer including epidermal and neural progenitors. The formation of the organs' anlage starts in mid-embryonic stages and the symmetrical brain primordium can be observed at the anterior pole, subepidermally, at early stages (Ramachandra et al., 2002).

One of the "most studied" acoel species is *Symsagittifera roscoffensis.* The nervous system of this acoel is arranged in an anterior domain forming a compact brain with neural cell bodies surrounding a neuropil, divided into two lobes and connected by three commissures (Bery et al., 2010; Semmler et al., 2010; reviewed in Martínez et al., 2017). The ventral part of the neuropil projects anteriorly to a commissural ring that surrounds the frontal organ. In addition, three pairs of cords arise from the brain and run along the anterior-posterior body axis, in a specific dorso-ventral distribution. Two dorsal cords arise from the posterior part of the brain (specifically from the third commissure), while the remaining four nerve cords lie more ventrally (a ventral central pair and a latero-ventral pair). The most prominent sensory organ is the acoel statocyst, located in the anterior part of the body and surrounded by the brain neuropil. Anterior to the statocyst are a pair of ocelli consisting of several sensory cells and a pigment cell (Bery et al., 2010; Gavilán et al., 2016; Perea-Atienza et al., 2015; Semmler et al., 2010). Juveniles of *S. roscoffensis* are about 220 µm after

bioRxiv preprint doi: https://doi.org/10.1101/237388; this version posted December 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

hatching and the brain occupies more than a third of their body length, a striking difference when compared to the adult specimens, where the brain occupies only a small anterior region (approximately an eighth of the animal's length) (Bery and Martínez, 2011; Bery et al., 2010; Gavilán et al., 2016; Perea-Atienza et al., 2015; Semmler et al., 2010).

Several studies in many animal systems have demonstrated the involvement of different bHLH family members in neurogenic processes, with the characteristic that most of them (but not all) belong to the so-called group A. *Symsagittifera roscoffensis* possesses an interesting set of bHLH genes, some of which are expressed in domains clearly overlapping the anterior part of the nervous system: SrAscA, SrAscB, SrE12/E47, SrHes/Hey and SrOligo. Their relative expression domains are represented here in a schematic model of the juvenile acoel (see Fig. 7). The genes SrAscA and SrAscB were expressed most anteriorly, in the brain area. Additionally to the CNS, the domain of SrAscA included a good part of the PNS (the tracks which connect the CNS with the surface of the organism and the nerve net) while the SrAscB domain extended from the frontal part of the animal (frontal organ) to the mid-central part of the animal, posterior to the statocyst and surrounding the mouth (Fig. 4, 5, 7; Supplementary material 1). Both gene expression patterns surround the statocyst and cover, most probably, the majority of the centrally located neuropil. Slightly overlapping with the expression pattern of SrAscB we located the domain of SrE12/E47, which was symmetrically divided into two areas, weakly expressed in the frontal brain of the juvenile (Fig. 1; Fig. 6). More extensive was the expression of SrOligo, which partially overlapped with the domains of all the four genes abovementioned, including a small region in front of the statocyst and a region that extended posteriorly to the medial-posterior part of the juvenile's body (Fig. 1; Fig. 6). Most probably, this gene is also expressed in the

bioRxiv preprint doi: https://doi.org/10.1101/237388; this version posted December 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

nerve net, together with the members of the achaete-scute family (Fig. 1,4,5,6. See Fig.8 for a schematic model).

SrNscl from group A and were not strongly expressed in the most anterior brain area. SrHes/Hey from group E was lower expressed in the brain than the SrAscA and SrAscB, however its domain was more extensive, including the major part of the nervous system. This gene was strongly expressed in the medium region of the cords (immediately posterior to the statocyst) and its domain includes a great part of the nervous system (Fig. 5; Supplementary material 2). Therefore, we cannot rule out the possibility that they have conserved the roles of their metazoan orthologs in the formation of neural structures.

Other bHLH genes that seem to possess less relevant roles in the nervous system (due to their low expression and/or the widespread dispersion of their expression patterns) are SrSrebp and SrArnt, from families B and C, respectively. SREBP transcription factor is crucial in cell metabolism, controlling the synthesis of fatty acids, triglycerides, and cholesterol, whereas SrARNT is involved in a huge variety of processes such as environmental responses to toxicity or light, all roles performed with its heterodimerization partner AHR. Ahr (not present in *S. roscoffensis*) has been associated with the migration and morphology of sensory and motor neurons (Huang et al., 2004; Qin and Powell-Coffman, 2004).According to their expression patterns, both genes (SrSrebp and SrArnt) have roles compatible with those observed for their metazoan orthologs.

The differential patterns of expression of these bHLH genes in different parts of the nervous system suggest the role that these genes have in the patterning of the nervous

bioRxiv preprint doi: https://doi.org/10.1101/237388; this version posted December 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

system and the development of its final architecture. These findings are consistent with those of other studies, which point to the importance of the combined expression of some bHLH transcription factors in patterning the neural tissue. They act in concert (or downstream) with other patterning genes that provide positional identity along the major body axis of animals, for instance in the dorso-ventral axis (Pax, Nkx and Irx) and along the anterior-posterior body axis (Otx, Gbx, En, and Hox families) (reviewed by F. Guillemot in 2007; Martín-Durán et al., 2017;our unpublished acoel results). Moreover, some studies have reported the combinatorial activity of orthologs of bHLH proneural proteins (Mash1, Hes1, Olig2) with other patterning proteins such as Pax6 and Nkx2.2 in promoting cell type specification in mice (see for instance: Sugimori et al., 2007).

The complement of bHLH detected in the acoel *S. roscoffensis* is reduced in comparison with other metazoan organisms studied. Most of the described roles of the bHLH orthologs in other organisms could not be performed with the small complement of bHLH present in *S. roscoffensis*. There are several different possible reasons for the detection of a reduced set of bHLHs in this clade. First, there is the obvious possibility that some bHLH appear to be missing because their sequences are too divergent and therefore, difficult to identify (see Fig. 8 and supplementary material 3, for a schematic representation through the different phyla). This is not unlikely in acoels, as their genomes have clearly changed a lot because of the high rate of sequence evolution. This has generated some clear genomic modifications, among which genetic losses are common. A reduction in the number of protein complements has been reported in several families previously (see Gavilán et al., 2016; Perea-Atienza et al., 2015). Other possible factors are the highly divergent sequences of members of other families, for instance the sequences belonging to the Wnt family of

bioRxiv preprint doi: https://doi.org/10.1101/237388; this version posted December 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

ligands identified in *S. roscoffensis* and the acoel *Hosftenia miamia*. They are clearly derived and this has made it impossible to classify them into the well-known metazoan families (Gavilan et al., 2015; Srivastava et al., 2014). An alternative possibility for the scarce number of bHLH relatives in the acoel genome is that we are using a newly sequenced genome, which we assume is almost complete, but could still be missing some fragments of the genome. In this context we should mention that the parallel use of transcriptomes, from adults and embryos, has not provided any new sequences that were not present in the genome.

Conclusions

The genome of *S. roscoffensis* contains a relatively low number of bHLH transcription factors. However a careful phylogenetic analysis of the sequences has revealed that this clade has bHLHs from various subfamilies. Acoels possess, at least, a bHLH belonging to five of the six classical high order groups. The conservation of the bHLH sequences analyzed is also paralleled by the conservation of the different gene expression domains. Clear correlations can be made between the expression domains in acoels and those reported for other metazoans, suggesting conservation of roles over evolutionary time. Our study detected the presence of a pool of bHLH genes (SrAscA, SrAscB,, SrE12/E47 SrOligo and SrHes/Hey)with expression patterns specific to territories within the nervous system, most probably in different cell populations of neurons and/or neural precursors. Their expression, which starts in most cases during embryogenesis, suggests that they are involved in the early specification of neural precursors and the later formation of the nervous system (or subdomains thereof). This analysis, together with that of other genes

bioRxiv preprint doi: https://doi.org/10.1101/237388; this version posted December 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

studied previously in our laboratory, such as the Hox family members Cdx and SoxB, constitutes a first step towards a clear understanding of how the nervous system is assembled in acoels. The study of many other regulatory genes in our laboratory (unpublished) suggests a complex gene network controlling the development of the acoel nervous system. The further characterization and detailed knowledge of the genes involved in neural development and patterning will help to address specific developmental issues in the future. Having access to the sequence of the *S. roscoffensis* genome has been the key factor in this study.

ACKNOWLEDGEMENTS

We would like to acknowledge the support offered by Drs. Cristian Cañestro (UB; Barcelona) and Anna Aragay (IBMB-CSIC; Barcelona), which was critical in order to finish this research project. E. Perea has been a recipient of an APIF (UB) doctoral fellowship. We should also acknowledge the continuous input provided by Brenda Gavilán during most of the duration of this project.

PM would like to acknowledge Prof. William McGinnis (UCSD), in which laboratory he learned the use of basic tools for multiplex in situ hybridization.

Achatz, J. G. and Martínez, P. (2012). The nervous system of Isodiametra pulchra (Acoela) with a discussion on the neuroanatomy of the Xenacoelomorpha and its evolutionary implications. *Front. Zool.* **9**, 27.

Achatz, J. G., Chiodin, M., Salvenmoser, W., Tyler, S. and Martinez, P. (2013). The Acoela: On their kind and kinships, especially with nemertodermatids and xenoturbellids (Bilateria incertae sedis). *Org. Divers. Evol.* **13**, 267–286.

Atchley, W. R. and Fitch, W. M. (1997). A natural classification of the basic helix–loop– helix class of transcription factors. *Evol. Contrib. by Walter M. Fitch* **94**, 5172–5176.

Baylies, M. K. and Bate, M. (1996). twist: A Myogenic Switch in Drosophila. *Science (80-.).* **272**, 1481 LP-1484.

Begley, C. G., Lipkowitz, S., Göbel, V., Mahon, K. a, Bertness, V., Green, a R., Gough, N. M. and Kirsch, I. R. (1992). Molecular characterization of NSCL, a gene encoding a helix-loop-

bioRxiv preprint doi: https://doi.org/10.1101/237388; this version posted December 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

helix protein expressed in the developing nervous system. *Proc. Natl. Acad. Sci. U. S. A.* **89**, 38–42.

Bertrand, N., Castro, D. S. and Guillemot, F. (2002). Proneural genes and the specification of neural cell types. *Nat. Rev. Neurosci.* **3**, 517–30.

Bery, A. and Martínez, P. (2011). Acetylcholinesterase activity in the developing and regenerating nervous system of the acoel Symsagittifera roscoffensis. *Acta Zool.* **92**, 383–392.

Bery, A., Cardona, A., Martinez, P. and Hartenstein, V. (2010). Structure of the central nervous system of a juvenile acoel, Symsagittifera roscoffensis. *Dev. Genes Evol.* **220**, 61–76.

Brentrup, D., Lerch, H., Jackle, H. and Noll, M. (2000). Regulation of Drosophila wing vein patterning: net encodes a bHLH protein repressing rhomboid and is repressed by rhomboid-dependent Egfr signalling. *Development* **127**, 4729–4741.

Camargo, N., Smit, A. B. and Verheijen, M. H. G. (2009). SREBPs: SREBP function in glia-neuron interactions. *FEBS J.* **276**, 628–636.

Cannon, J. T., Vellutini, B. C., Smith, J., Ronquist, F., Jondelius, U. and Hejnol, A. (2016). Xenacoelomorpha is the sister group to Nephrozoa. *Nature* **530**, 89–93.

Casarosa, S., Fode, C. and Guillemot, F. (1999). Mash1 regulates neurogenesis in the ventral telencephalon. *Development* **126**, 525–534.

Chiodin, M., Børve, A., Berezikov, E., Ladurner, P., Martinez, P. and Hejnol, A. (2013). Mesodermal Gene Expression in the Acoel Isodiametra pulchra Indicates a Low Number of Mesodermal Cell Types and the Endomesodermal Origin of the Gonads. *PLoS One* **8**,.

Cowles, M. W., Brown, D. D. R., Nisperos, S. V, Stanley, B. N., Pearson, B. J. and Zayas, R. M. (2013). Genome-wide analysis of the bHLH gene family in planarians identifies factors required for adult neurogenesis and neuronal regeneration. *Development* **140**, 4691–702.

Dambly-Chaudiere, C. and Ghysen, A. (1987). Independent subpatterns of sense organs require independent genes of the achaete-scute complex in Drosophila larvae. *Genes Dev.* **1**, 297–306.

Dang, C. V, Dolde, C., Gillison, M. L. and Kato, G. J. (1992). Discrimination between related DNA sites by a single amino acid residue of Myc-related basic-helix-loop-helix proteins. *Proc. Natl. Acad. Sci. U. S. A.* **89**, 599–602.

Denison, M. S., Soshilov, A. A., He, G., Degroot, D. E. and Zhao, B. (2011). Exactly the same but different: Promiscuity and diversity in the molecular mechanisms of action of the aryl hydrocarbon (dioxin) receptor. *Toxicol. Sci.* **124**, 1–22.

Eilers, M. and Eisenman, R. N. (2008). Myc's broad reach. *Genes Dev.* **22**, 2755–2766.

Fischer, B., Azim, K., Hurtado-Chong, A., Ramelli, S., Fernández, M. and Raineteau, O. (2014). E-proteins orchestrate the progression of neural stem cell differentiation in the postnatal forebrain. *Neural Dev.* **9**, 23.

Fortunato, S. A. V, Vervoort, M., Adamski, M. and Adamska, M. (2016). Conservation and divergence of bHLH genes in the calcisponge Sycon ciliatum. *Evodevo* 1–12.

García-Bellido, A. and de Celis, J. F. (2009). The complex tale of the achaete-scute complex: a paradigmatic case in the analysis of gene organization and function during development. *Genetics* **182**, 631–9.

bioRxiv preprint doi: https://doi.org/10.1101/237388; this version posted December 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Gavilán, B., Perea-Atienza, E. and Martínez, P. (2015). Xenacoelomorpha: a case of independent nervous system centralization? *Philos. Trans. R. Soc. B-Biological Sci.* **371**,.

Gobel, V., Lipkowitz, S., Kozak, C. A. and Kirsch, I. R. (1992). Nscl-2 - a Basic Domain Helix-Loop-Helix Gene Expressed in Early Neurogenesis. *Cell Growth Differ.* **3**, 143–148.

Guillemot, F. (2007). Spatial and temporal specification of neural fates by transcription factor codes. *Development* **134**, 3771–3780.

Guillemot, F., Lo, L.-C., Johnson, J. E., Auerbach, A., Anderson, D. J. and Joyner, A. L. (1993). Mammalian achaete-scute homolog 1 is required for the early development of olfactory and autonomic neurons. *Cell* **75**, 463–476.

Gyoja, F. (2014). A genome-wide survey of bHLH transcription factors in the Placozoan Trichoplax adhaerens reveals the ancient repertoire of this gene family in metazoan. *Gene* **542**, 29–37.

Gyoja, F. (2017). Basic helix-loop-helix transcription factors in evolution: Roles in development of mesoderm and neural tissues. *genesis* **55**, e23051.

Gyoja, F., Kawashima, T. and Satoh, N. (2012). A genomewide survey of bHLH transcription factors in the coral Acropora digitifera identifies three novel orthologous families, pearl, amber, and peridot. *Dev. Genes Evol.* **222**, 63–76.

Hartl, M., Mitterstiller, A.-M., Valovka, T., Breuker, K., Hobmayer, B. and Bister, K. (2010). Stem cell-specific activation of an ancestral myc protooncogene with conserved basic functions in the early metazoan Hydra. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 4051–4056.

Haszprunar, G. (2016). Review of data for a morphological look on Xenacoelomorpha (Bilateria incertae sedis). *Org. Divers. Evol.* **16**, 363–389.

Hejnol, A. and Lowe, C. J. (2015). Embracing the comparative approach: how robust phylogenies and broader developmental sampling impacts the understanding of nervous system evolution. *Phil. Trans. R. Soc. B* **370**, 20150045.

Hejnol, A. and Pang, K. (2016). Xenacoelomorpha's significance for understanding bilaterian evolution. *Curr. Opin. Genet. Dev.* **39**, 48–54.

Hejnol, A., Obst, M., Stamatakis, A., Ott, M., Rouse, G. W., Edgecombe, G. D., Martinez, P., Baguñà, J., Bailly, X., Jondelius, U., et al. (2009). Assessing the root of bilaterian animals with scalable phylogenomic methods. *Proc. R. Soc. B* **276**, 4261–4270.

Heng, J. I. T. and Tan, S. S. (2003). The role of class I HLH genes in neural development - Have they been overlooked? *BioEssays* **25**, 709–716.

Henry, J. Q., Martindale, M. Q. and Boyer, B. C. (2000). The unique developmental program of the acoel flatworm, Neochildia fusca. *Dev. Biol.* **220**, 285–95.

Horton, S., Meredith, A., Richardson, J. a and Johnson, J. E. (1999). Correct coordination of neuronal differentiation events in ventral forebrain requires the bHLH factor MASH1. *Mol. Cell. Neurosci.* **14**, 355–69.

Huang, X., Powell-Coffman, J. A. and Jin, Y. (2004). The AHR-1 aryl hydrocarbon receptor and its co-factor the AHA-1 aryl hydrocarbon receptor nuclear translocator specify GABAergic neuron cell fate in C. elegans. *Development* **131**, 819–28.

Ishibashi, M; Ang, S; Shiota, K; Nakanishi, S; Kageyama, R; Guillemot, F. (1995). Targeted

bioRxiv preprint doi: https://doi.org/10.1101/237388; this version posted December 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

disruption of mammalian hairy and Enhancer of split homolog-1 (HES-1) leads to up regulation of neural hehx-loop-.hellx factors, premature neurogenesls, and severe neural tube defects. *Genes Dev.* **1**, 3136–3148.

Jarman, A. P., Grau, Y., Jan, L. Y. and Jan, Y. N. (1993). atonal is a proneural gene that directs chordotonal organ formation in the Drosophila peripheral nervous system. *Cell* **73**, 1307–1321.

Jones, S. (2004). An overview of the basic helix-loop-helix proteins. *Genome Biol.* **5**, 226.

Kageyama, R., Ohtsuka, T. and Kobayashi, T. (2008). Roles of Hes genes in neural development. *Dev. Growth Differ.* **50**,.

Kim, W. Y. (2012). NeuroD1 is an upstream regulator of NSCL1. *Biochem. Biophys. Res. Commun.* **419**, 27–31.

Koide, S; Yoshida, I; Tsuji , A; Matsuda, Y. (2003). The Expression of Proprotein Convertase PACE4 Is Highly Regulated by Hash-2 in Placenta: Possible Role of Placenta-Specific Basic Helix-Loop-Helix Transcription Factor, Human Achaete-Scute Homologue-2. 134 (3): 433-440.

Krüger, M., Ruschke, K. and Braun, T. (2004). NSCL-1 and NSCL-2 synergistically determine the fate of GnRH-1 neurons and control necdin gene expression. *EMBO J.* **23**, 4353–4364.

Kury, P., Greiner-Petter, R., Cornely, C., Jurgens, T. and Muller, H. W. (2002). Mammalian achaete scute homolog 2 is expressed in the adult sciatic nerve and regulates the expression of Krox24, Mob-1, CXCR4, and p57kip2 in Schwann cells. *J Neurosci* **22**, 7586–7595.

Lassar, A. B., Davis, R. L., Wright, W. E., Kadesch, T., Murre, C., Voronova, A., Baltimore, D. and Weintraub, H. (1991). Functional activity of myogenic HLH proteins requires hetero-oligomerization with E12/E47-like proteins in vivo. *Cell* **66**, 305–315.

Layden, M. J. and Martindale, M. Q. (2014). Non-canonical Notch signaling represents an ancestral mechanism to regulate neural differentiation. *Evodevo* **5**, 30.

Layden, M. J., Boekhout, M. and Martindale, M. Q. (2012). Nematostella vectensis achaete-scute homolog NvashA regulates embryonic ectodermal neurogenesis and represents an ancient component of the metazoan neural specification pathway. *Development* **139**, 1013–1022.

Ledent, V. and Vervoort, M. (2001). The Basic Helix-Loop-Helix Protein Family: Comparative Genomics and Phylogenetic Analysis. *Genome Res.* **11**, 754–770.

Li, K.-L., Lu, T.-M. and Yu, J.-K. (2014). Genome-wide survey and expression analysis of the bHLH-PAS genes in the amphioxus Branchiostoma floridae reveal both conserved and diverged expression patterns between cephalochordates and vertebrates. *Evodevo* **5**, 20.

Liu, X., Chen, X., Zhong, B., Wang, A., Wang, X., Chu, F., Nurieva, R. I., Yan, X., Chen, P., van der Flier, L. G., et al. (2014). Transcription factor achaete-scute homologue 2 initiates follicular T-helper-cell development. *Nature* **507**, 513–8.

Lu, Q. R., Sun, T., Zhu, Z., Ma, N., Garcia, M., Stiles, C. D. and Rowitch, D. H. (2002). Common developmental requirement for Olig function indicates a motor neuron/oligodendrocyte connection. *Cell* **109**, 75–86.

Martín-Durán, J. M., Pang, K., Børve, A., Lê, H. S., Furu, A., Cannon, J. T., Jondelius, U. and

bioRxiv preprint doi: https://doi.org/10.1101/237388; this version posted December 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Hejnol, A. (2017). Convergent evolution of bilaterian nerve cords. *Nature*.

Martínez, P., Hartenstein, V. and Sprecher, S. (2017). Xenacoelomorpha Nervous Systems.

Massari, M. E. and Murre, C. (2000). Minireview Helix-Loop-Helix Proteins : Regulators of Transcription in Eucaryotic Organisms. **20**, 429–440.

McIntosh, B. E., Hogenesch, J. B. and Bradfield, C. a (2010). Mammalian Per-Arnt-Sim proteins in environmental adaptation. *Annu. Rev. Physiol.* **72**, 625–645.

Mizuguchi, R., Sugimori, M., Takebayashi, H., Kosako, H., Nagao, M., Yoshida, S., Nabeshima, Y. I., Shimamura, K. and Nakafuku, M. (2001). Combinatorial roles of Olig2 and Neurogenin2 in the coordinated induction of pan-neuronal and subtype-specific properties of motoneurons. *Neuron* **31**, 757–771.

Monjo, F. and Romero, R. (2015). Embryonic development of the nervous system in the planarian Schmidtea polychroa. *Dev. Biol.* **397**, 305–319.

Müller, P., Seipel, K., Yanze, N., Reber-Müller, S., Streitwolf-Engel, R., Stierwald, M., Spring, J. and Schmid, V. (2003). Evolutionary aspects of developmentally regulated helix-loop-helix transcription factors in striated muscle of jellyfish. *Dev. Biol.* **255**, 216–229.

Murre, C., McCaw, P. S., Vaessin, H., Caudy, M., Jan, L. Y., Jan, Y. N., Cabrera, C. V., Buskin, J. N., Hauschka, S. D., Lassar, A. B., et al. (1989). Interactions between heterologous helix-loop-helix proteins generate complexes that bind specifically to a common DNA sequence. *Cell* **58**, 537–544.

Nakano, H. (2015). What is Xenoturbella? *Zool. Lett.* **1**, 22.

Northcutt, R. G. (2012). Evolution of centralized nervous systems: Two schools of evolutionary thought. *Proc. Natl. Acad. Sci.* **109**, 10626–10633.

Novitch, B. G., Chen, A. I. and Jessell, T. M. (2001). Coordinate regulation of motor neuron subtype identity and pan-neuronal properties by the bHLH repressor Olig2. *Neuron* **31**, 773–789.

Parker, G. H. (1919). The Elementary Nervous System.

Perea-Atienza, E., Gavilán, B., Chiodin, M., Abril, J. F., Hoff, K. J., Poustka, A. J. and Martínez, P. (2015). The nervous system of Xenacoelomorpha: A genomic perspective. *J. Exp. Biol.* **218**,.

Perea-Atienza, E., Gavilán, B. and Martínez, P. (2018). Inmunostaining and *insitu* hybridization of the developing Acoel nervous system. *Methods in Brain Development. Methods in Molecular Biology.* Springer Nature (in press).

Philippe, H., Brinkmann, H., Copley, R. R., Moroz, L. L., Nakano, H., Poustka, A. J., Wallberg, A., Peterson, K. J. and Telford, M. J. (2011). Acoelomorph flatworms are deuterostomes related to Xenoturbella. *Nature* **470**, 255–258.

Qin, H. and Powell-Coffman, J. A. (2004). The Caenorhabditis elegans aryl hydrocarbon receptor, AHR-1, regulates neuronal development. *Dev. Biol.* **270**, 64–75.

Ramachandra, N. B., Gates, R. D., Ladurner, P., Jacobs, D. K. and Hartenstein, V. (2002). Embryonic development in the primitive bilaterian Neochildia fusca: Normal morphogenesis and isolation of POU genes Brn-1 and Brn-3. *Dev. Genes Evol.* **212**, 55–69.

bioRxiv preprint doi: https://doi.org/10.1101/237388; this version posted December 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Reitzel, A. M., Tarrant, A. M. and Levy, O. (2013). Circadian clocks in the cnidaria: Environmental entrainment, molecular regulation, and organismal outputs. *Integr. Comp. Biol.* **53**, 118–130.

Richards, G. S. and Rentzsch, F. (2014). Transgenic analysis of a SoxB gene reveals neural progenitor cells in the cnidarian Nematostella vectensis. *Development* **141**, 4681–9.

Richards, G. S. and Rentzsch, F. (2015). Regulation of Nematostella neural progenitors by SoxB, Notch and bHLH genes. *Development* **142**, 3332–3342.

Roberts, V. J., Steenbergen, R. and Murre, C. (1993). Localization of E2A mRNA expression in developing and adult rat tissues. *Proc Natl Acad Sci U S A* **90**, 7583–7587.

Ruiz-Trillo, I. and Paps, J. (2016). Acoelomorpha: earliest branching bilaterians or deuterostomes? *Org. Divers. Evol.* **16**, 391–399.

Schmid, T., Kruger, M. and Braun, T. (2007). NSCL-1 and -2 control the formation of precerebellar nuclei by orchestrating the migration of neuronal precursor cells. *J. Neurochem.* **102**, 2061–2072.

Sebé-Pedrós, A., De Mendoza, A., Lang, B. F., Degnan, B. M. and Ruiz-Trillo, I. (2011). Unexpected repertoire of metazoan transcription factors in the unicellular holozoan capsaspora owczarzaki. *Mol. Biol. Evol.* **28**, 1241–1254.

Semmler, H., Chiodin, M., Bailly, X., Martinez, P. and Wanninger, A. (2010). Steps towards a centralized nervous system in basal bilaterians: Insights from neurogenesis of the acoel Symsagittifera roscoffensis. *Dev. Growth Differ.* **52**, 701–713.

Simionato, E., Ledent, V., Richards, G., Thomas-Chollier, M., Kerner, P., Coornaert, D., Degnan, B. M. and Vervoort, M. (2007). Origin and diversification of the basic helix-loop-helix gene family in metazoans: insights from comparative genomics. *BMC Evol. Biol.* **7**, 33.

Srivastava, M., Mazza-Curll, K. L., Van Wolfswinkel, J. C. and Reddien, P. W. (2014). Whole-body acoel regeneration is controlled by Wnt and Bmp-Admp signaling. *Curr. Biol.* **24**, 1107–1113.

Subramanian, S., Denver, D. R., Millar, C. D., Heupink, T., Aschrafi, A., Emslie, S. D., Baroni, C. and Lambert, D. M. (2009). High mitogenomic evolutionary rates and time dependency. *Trends Genet.* **25**, 482–486.

Sugimori, M., Nagao, M., Bertrand, N., Parras, C. M., Guillemot, F. and Nakafuku, M. (2007). Combinatorial actions of patterning and HLH transcription factors in the spatiotemporal control of neurogenesis and gliogenesis in the developing spinal cord. *Development* **134**, 1617–29.

Sun, Y., Nadal-Vicens, M., Misono, S., Lin, M. Z., Zubiaga, A., Hua, X., Fan, G. and Greenberg, M. E. (2001). Neurogenin promotes neurogenesis and inhibits glial differentiation by independent mechanisms. *Cell* **104**, 365–376.

van der Flier, L. G., van Gijn, M. E., Hatzis, P., Kujala, P., Haegebarth, A., Stange, D. E., Begthel, H., van den Born, M., Guryev, V., Oving, I., et al. (2009). Transcription Factor Achaete Scute-Like 2 Controls Intestinal Stem Cell Fate. *Cell* **136**, 903–912.

Van Doren, M., Ellis, H. M. and Posakony, J. W. (1991). The Drosophila extramacrochaetae protein antagonizes sequence-specific DNA binding by daughterless/achaete-scute protein complexes. *Development* **113**, 245–255.

bioRxiv preprint doi: https://doi.org/10.1101/237388; this version posted December 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Wheeler, S. R., Carrico, M. L., Wilson, B. a, Brown, S. J. and Skeath, J. B. (2003). The expression and function of the achaete-scute genes in Tribolium castaneum reveals conservation and variation in neural pattern formation and cell fate specification. *Development* **130**, 4373–4381.

Zhou, Q. and Anderson, D. J. (2002). The bHLH transcription factors OLIG2 and OLIG1 couple neuronal and glial subtype specification. *Cell* **109**, 61–73.

Figure legends:

Figure 1. Expression of bHLH orthologs in embryos and juveniles of the acoel *S. roscoffensis*. First and third columns with the expression patterns from 12 to 24 hours post-fertilization embryos are indicated. Second and fourth column with juveniles from 12 to 24 hours post-hatch juveniles indicated on the top. Name of the corresponded ortholog genes are indicated in each panel. Asterisk indicated probably the anterior pole of the embryo (Some embryos panel lack the asterisk, because we were not able to suggest the position of the anterior pole). Arrowhead indicates the position of the statocyst and therefore the anterior part of the juvenile. Embryos scale bar: 60 um; Juveniles scale bar: 100 um.

Figure 2. Expression of synaptotagmin gene ortholog in embryo and juvenile of the acoel *S. roscoffensis*. A. ISH expression pattern on 12 to 24 hours post-fertilization embryo. Asterisk indicates the anterior pole of the embryo. **B.**ISH expression pattern in juvenile from 12 to 24 hours post-hatch Arrowhead indicates the position of the statocyst and the anterior part of the juvenile. **C.** FISH expression pattern (in green) and anti-SrStg antibody (in pink) on a 12 to 24 post-hatch juvenile. Embryo scale bar: 60 um; Juveniles scale bar: 100 um.

Figure 3: Confocal microscopy projections of different whole mount stages of S. roscoffensis with the positive structures revealed by the specific designed antibody anti-SrASC-A (the protein orthologue sequence of MASH1).A. Embryo between 12-24h post-fertilization showing two positive cells (arrowheads) that are at the first embryonic layers. **B.** Juveniles between about

bioRxiv preprint doi: https://doi.org/10.1101/237388; this version posted December 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

24h post-hatch. Arrows indicate some positive cells in outer layers of the organism. Some developed sensory structures appear connecting outside of the epidermis. **C.** In later stages or old juveniles, this putative sensory cells surround the acoel epidermis, more concentrated in the anterior part of the animal. It should be pointed out that in the hatchlings, these putative sensory cells are not completely develop and shows small dots of signal in cells at the periphery in outer layers.**D.** Anterior part of the body, showing the revealed structures by the antibody (pink) and nuclei (green). Notice how the peripheral sensory cells arrive to the external layer (arrowhead). Arrows show small signal in the cells inside the body. In B, C, D. Anterior part of the body is located up. Scale bars: A: 25 um; B and C: 60 um.

Figure 4. Expression of SrAscA (pink) and SrAscB (green), achaete-scute gene family orthologs in aprox. 24h juvenile of the acoel*S. roscoffensis*. A. Double-FISH expression pattern of a medio-ventral longitudinal section of the anterior part of the animal. Arrowheads indicate the cell clusters where the expression is higher in the CNS (also visible in the SrAscB juvenile pattern of the figure 1). Arrows point tothe peripheral nerve tracks projecting to the animal's surface. Arectangular boxsurrounds the expression domain of both genes in the area of the statocyst. **B. Detailed**image of the**double F**ISH-detecteddomainof expression in the frontal organ'sassociated cell populations (Arrowheads). **C.**Dorso-medial longitudinal section of the double FISHdetecting expression in the anterior part of the animal. Arrowheads point the two lobes of the brain. The circle surrounds the position of the statocyst. C. Dorsal view of a whole mount double FISH-stained animal where can be appreciated the SrAscA expression in the peripheral nerve net. A circle surrounds the position of the statocyst. The union of the cords can be detected anteriorly and posteriorly (arrow). D. Detail in higher magnification of the posterior part of the organism showed in D, arrows point the posterior end of the cords. Scale bars: A=20um;B=5um;C=20um;D=30um.

bioRxiv preprint doi: https://doi.org/10.1101/237388; this version posted December 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 5. Expression domain of SrAscB (pink) and SrHes/Hey (green), in aprox. 24h juveniles of the acoel*S. roscoffensis*. A. Ventral detail of the anterior part of the organism where itis detected the strongest expression of AscB (pink) (arrows). Notice that SrHes/Hey shows only a low expression levelin the ventral nerve cords inserted in the brain (green). B. Composition with three different medial sections of the anterior part of the organism's body showing the expression of SrHes/Hey in a great part of the brain (green). Notice here the shared expression domain with SrAscA (pink) in the area where junctions between the cords and the commissures of the brain occurr. C. Dorsal section detail with a higher magnification in the zone of the brain showing the expression of SrHes/Hey concentrated in the area of the statocyst (circle). D. Medial section, with higher magnification of the brain area, at the level of the statocyst, showing the expression of SrHes/Hey in the dorsal and the medioventral cords. Asterisk signals the position of the statocyst and the arrow marks the commissure located immediately posterior. Scale bars: A=10um;B=40um;C=20um;D=20um.

Figure 6. Expression of SrOligo (pink) and SrE12/E47 (green), in aprox. 24h juvenile of the acoel*S. roscoffensis*.A. Dorsal longitudinal section of a whole mount single FISH animal where can be appreciate the SrOligo expression domain, in a small concentrated domain in a frontal area of the brain (arrowhead) and in the area near the mouth (circle). B. Medio-ventral section of a whole mount single FISH-stained animal where can be appreciate the SrE12/E47 expression in the central part of the body (rectangle) and at low levels in the frontal part of the brain (arrowhead). C. Dorsal section of a whole mount double FISH animal showing the expression pattern of SrOligo and SrE12/E47. Regarding to the nervous system, the first one shows an anterior concentration (arrow) while the second shows expression in the brain area. D. Ventral section of a whole mount double FISH animal showing the expression pattern of SrOligo and SrE12/E47. The expression of SrE12/E47 is located in the central part of the animal and it is almost undetectable in the nervous

bioRxiv preprint doi: https://doi.org/10.1101/237388; this version posted December 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

system. Arrowhead points to the expression domain of SrOligo. E. Ventral section of a whole mount double FISH-stained animal showing the expression pattern of SrOligo and SrE12/E47 at higher magnification; arrowheads point to the expression of SrOligo in the brain. F. Dorsal section of a whole mount double FISH-stained animal showing the expression pattern of SrOligo and SrE12/E47.Scale bars: 40um.

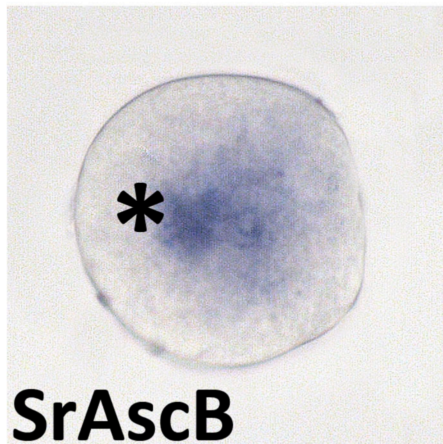
Figure 7. Schematic model of collected mRNA expression patterns obtained in our study for the bHLH genes (with putative expression in the nervous system). Abbreviations**: ns:** nervous system; **m:** mouth; **(*):** anterior statocyst. The model is based, essentially, on the FISH *in situ* data, since they provide us with higher, more accurate, relative patterns.

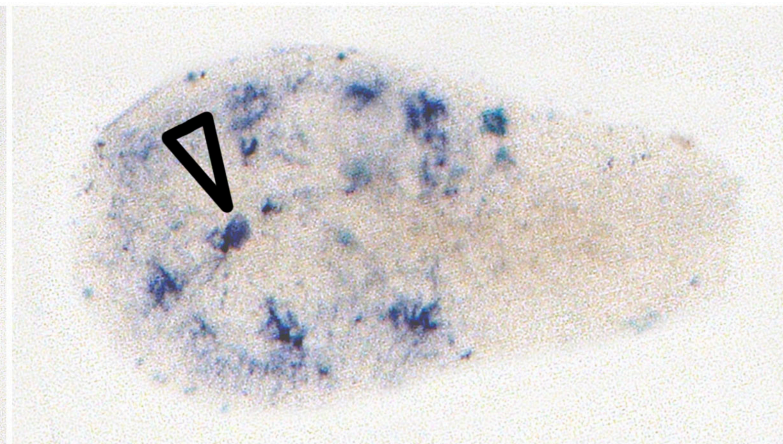
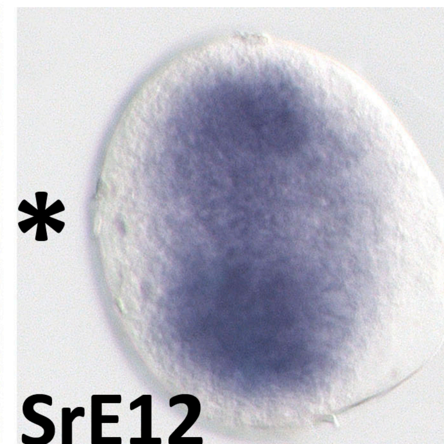
Figure 8. **Families from bHLH group A (which belong most of the neurogenic bHLH genes) represented in different species from several groups.** Colored boxes indicate the presence of the family while empty boxes indicate absence. The families from Xenacoelomorpha species showed with green cubes. Question marks are in boxes the presence of a family need further confirmation. The two last columns represent the number of orphans and the total number of bHLH in each species (not only from group A). The image allows us to see the losses in the different clades produced over evolutionary time. The families represented in our species would agree with a third bHLH expansion between the cnidarians annephrozoan divergences (mainly reflected in the group A), as suggestedby Simionato et al. 2007. Many families have bilaterian, but not cnidarian members;several of them found in *X. bocki's*genome (data unpublished).The data for *X. bocki* and *S. roscoffensis* are derived from our previous analysis (Perea-Atienza et al., 2015). Reference species used were:*H. sapiens, N. vectensis, Daphnia pulex, Caenorhabditis elegans, Tribolium castaneum, Lottia gigantean, Branchiostoma floridae, Amphimedon queenslandica, Strongylocentrotus purpuratus, Capitella sp. I., D. melanogaster and H. magnipapillata*, all derived from the study of Simionato et al., 2007; the data of *Schmidtea mediterranea* was from Cowles et al., 2013; the data

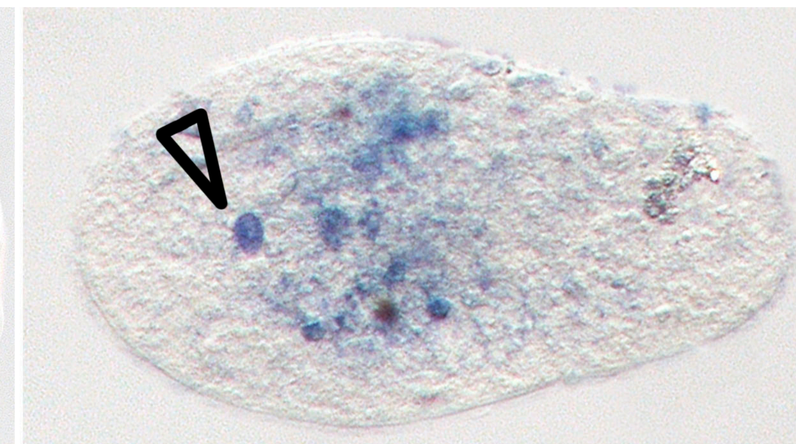
bioRxiv preprint doi: https://doi.org/10.1101/237388; this version posted December 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

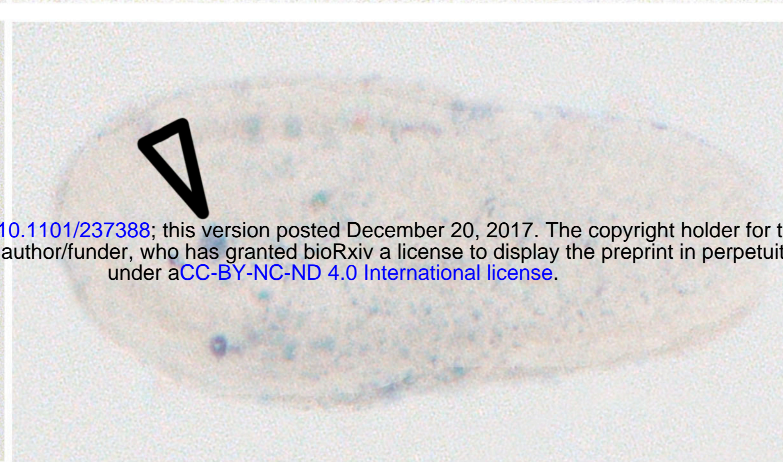
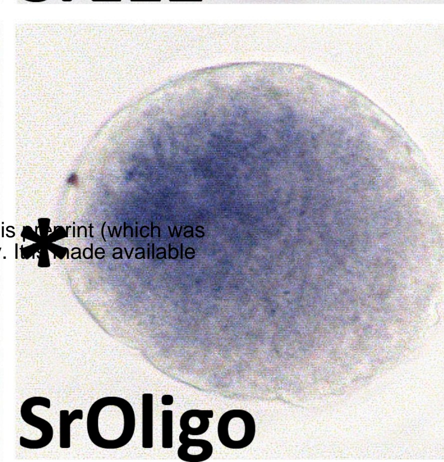
from *A. digitifera* (plus the latest identifications in *N. vectensis*) and *Trichoplax adhaerens* were obtained from Gyoja et al. (Gyoja et al., 2012; Gyoja et al., 2014); the *Sycon ciliatum* data was from Fortunato et al., 2016.

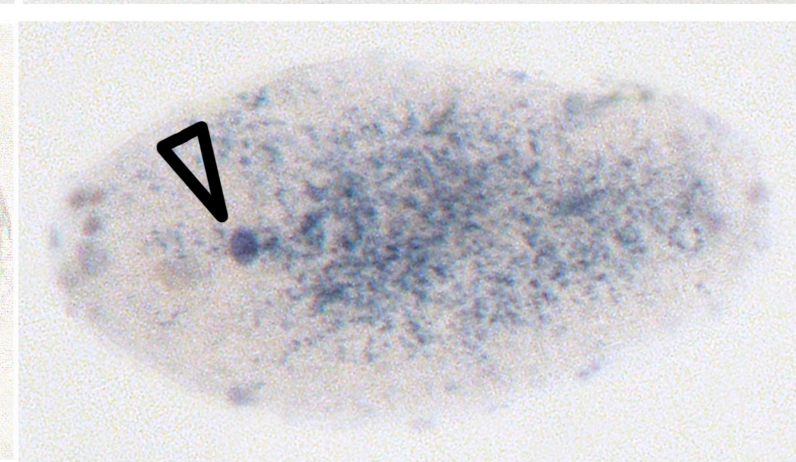
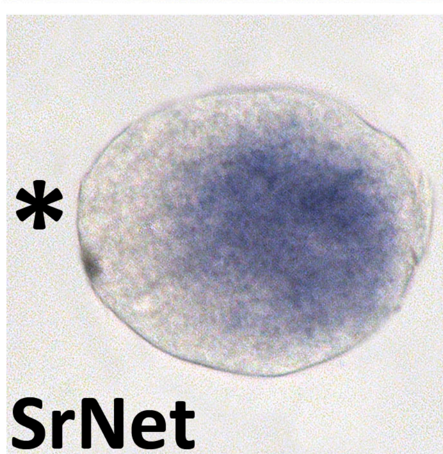
Embryos 12-24 hpf | Juveniles 12-24 hph | Embryos 12-24 hpf | Juveniles 12-24 hph

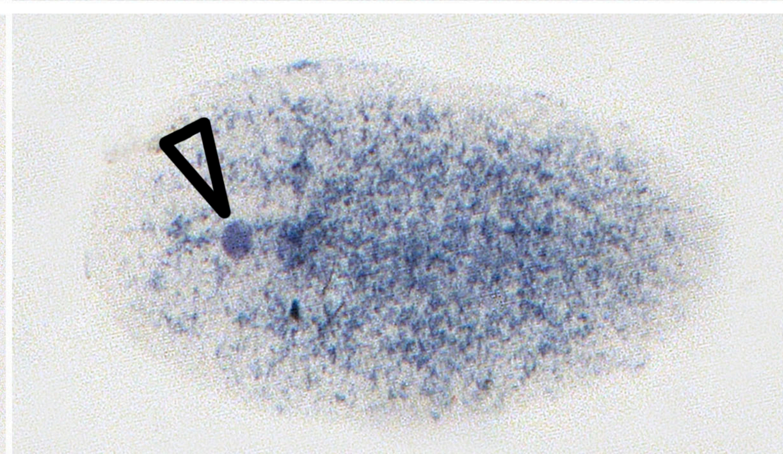
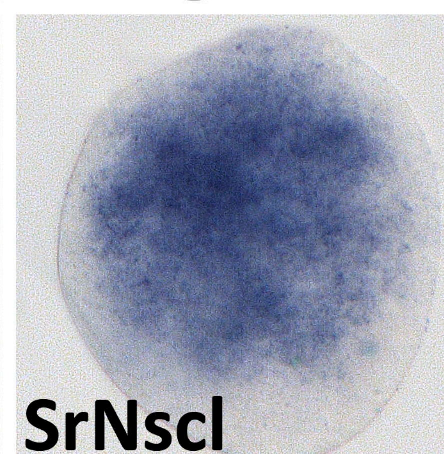


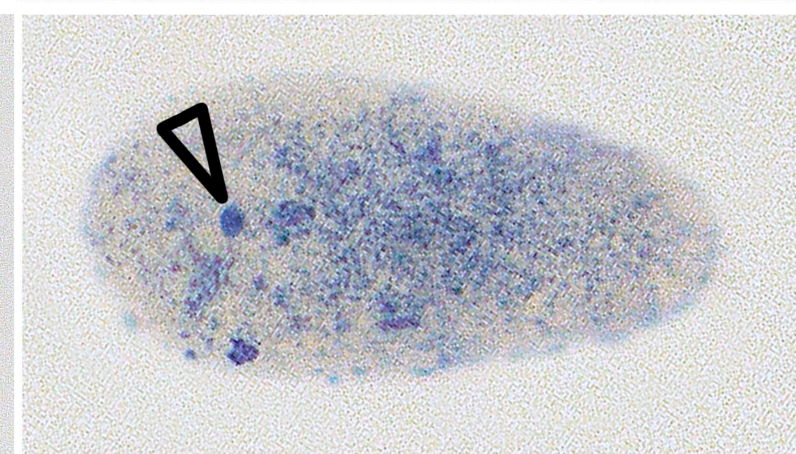
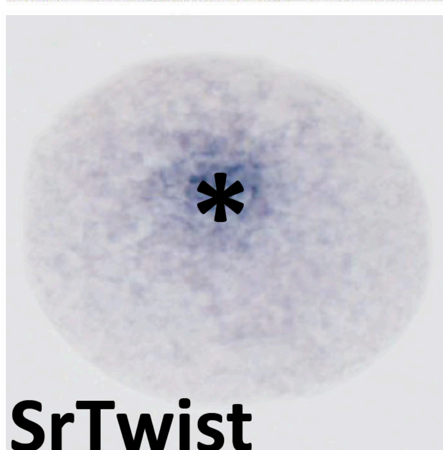



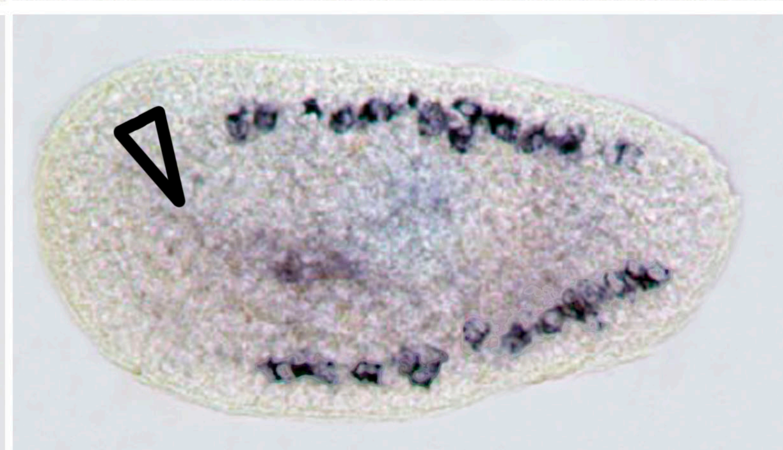
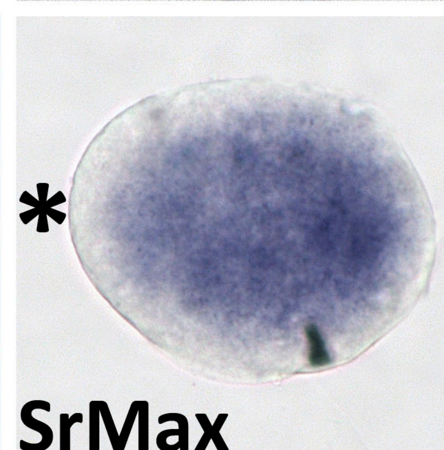


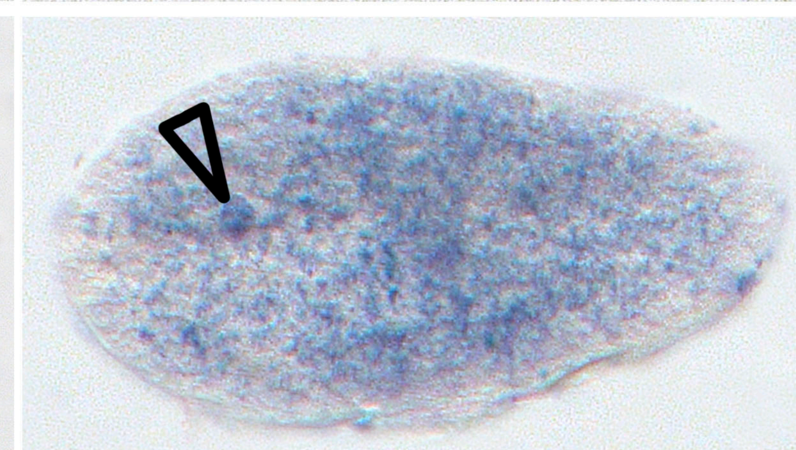
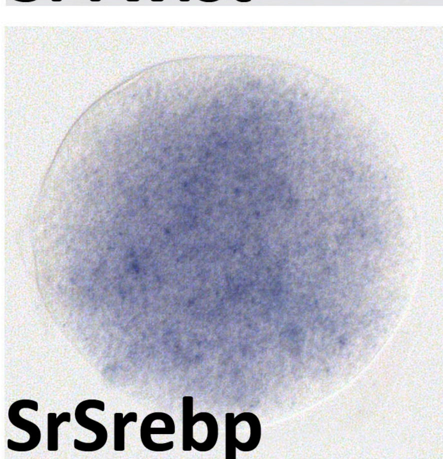

bioRxiv preprint doi: https://doi.org/10.1101/237388; this version posted December 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.


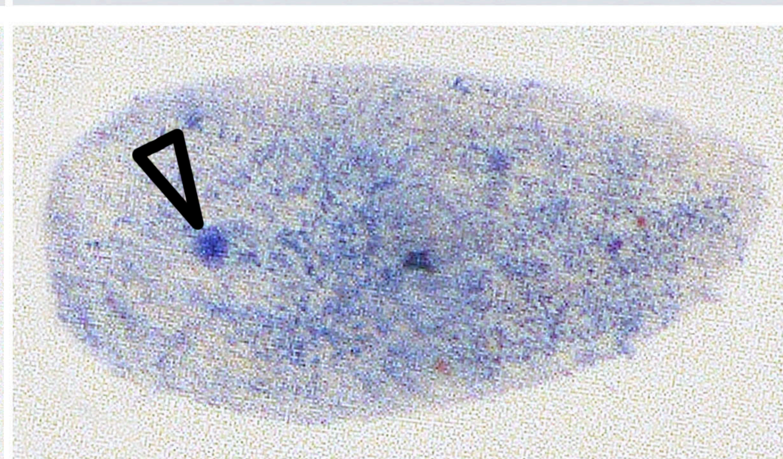
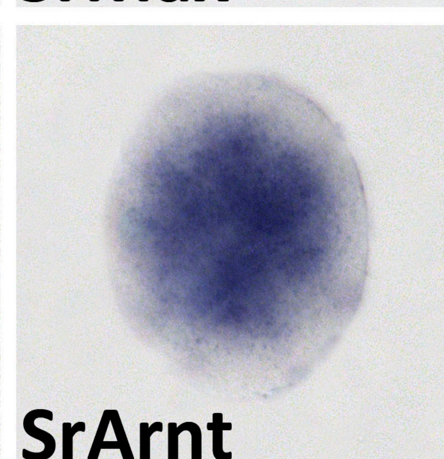



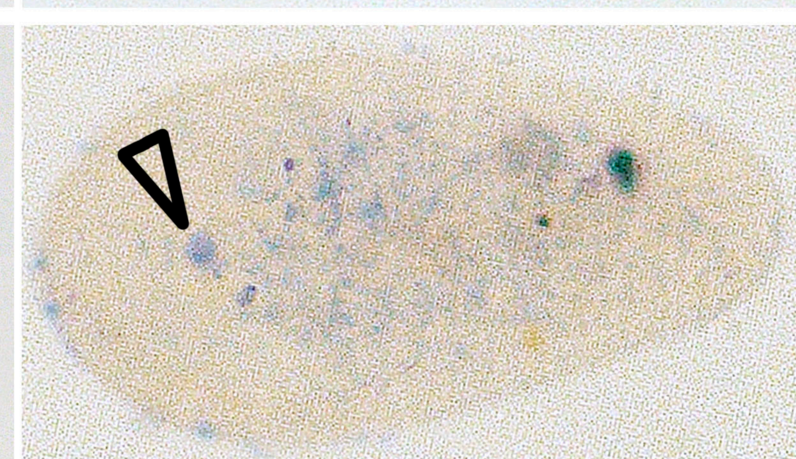
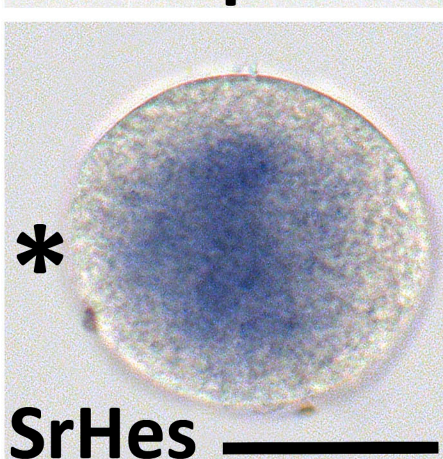



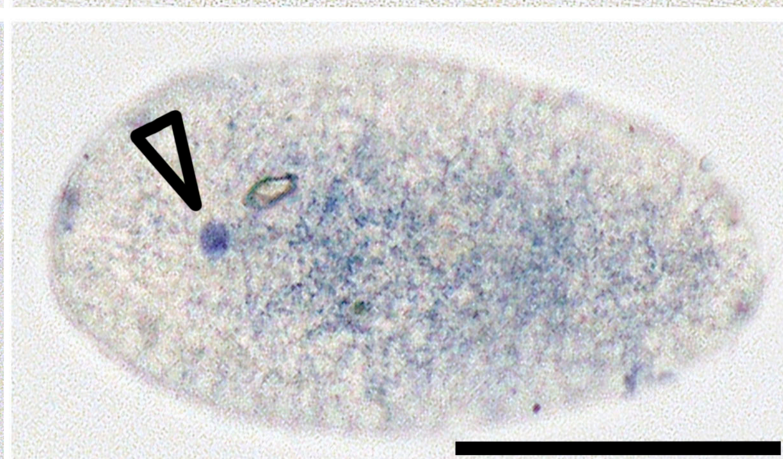
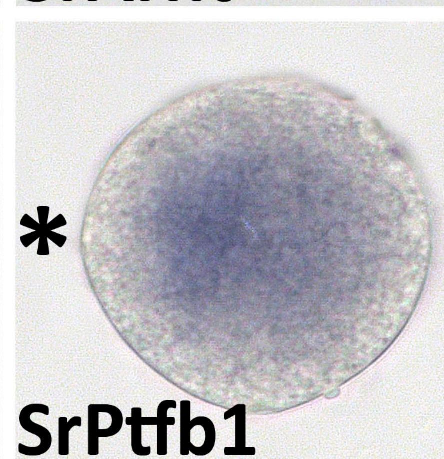



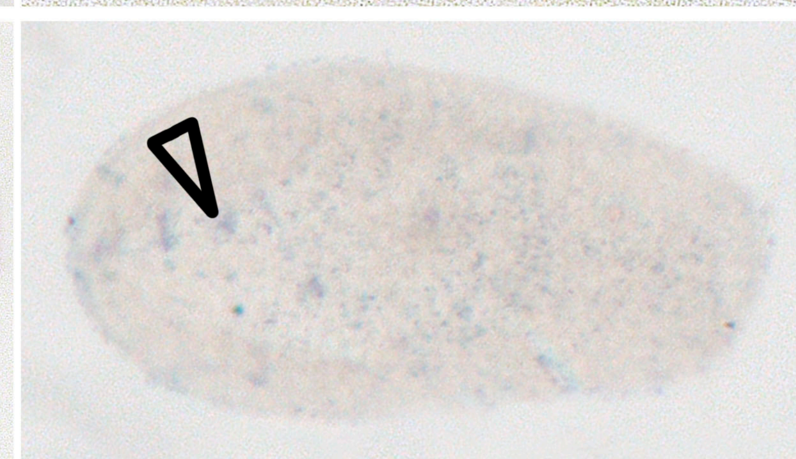



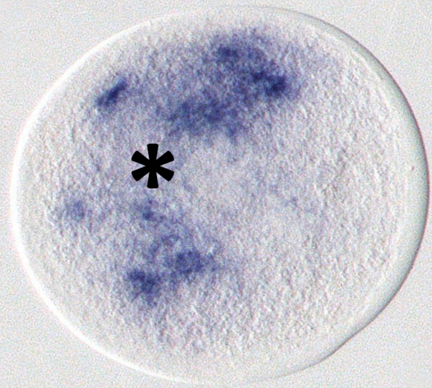



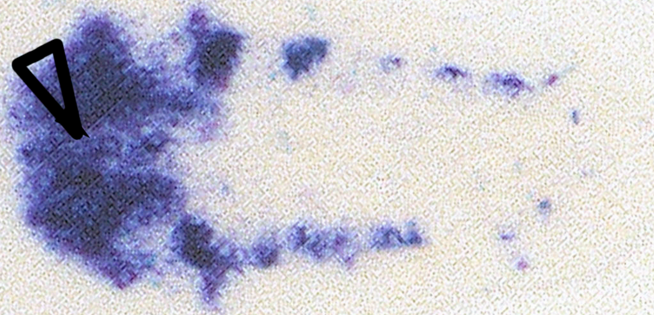



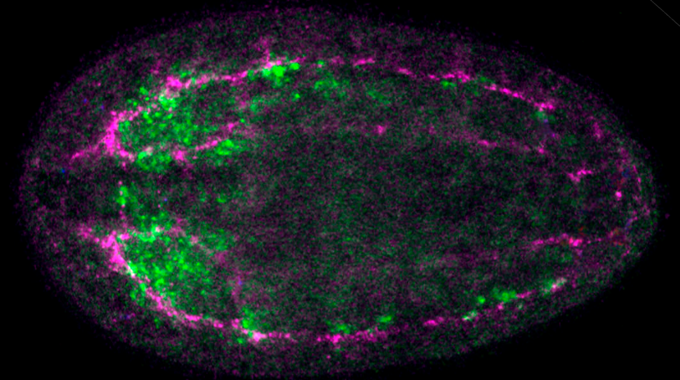


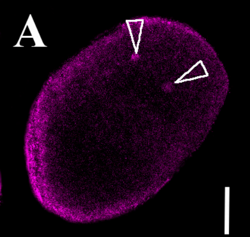
Embryo 12 - 24 hpf

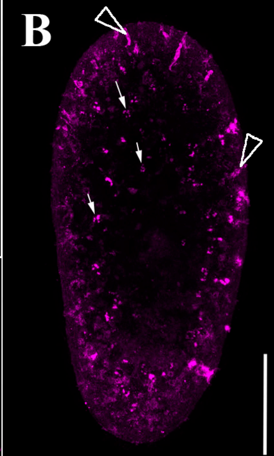
Juveniles 12 - 24 hph



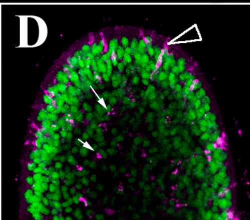


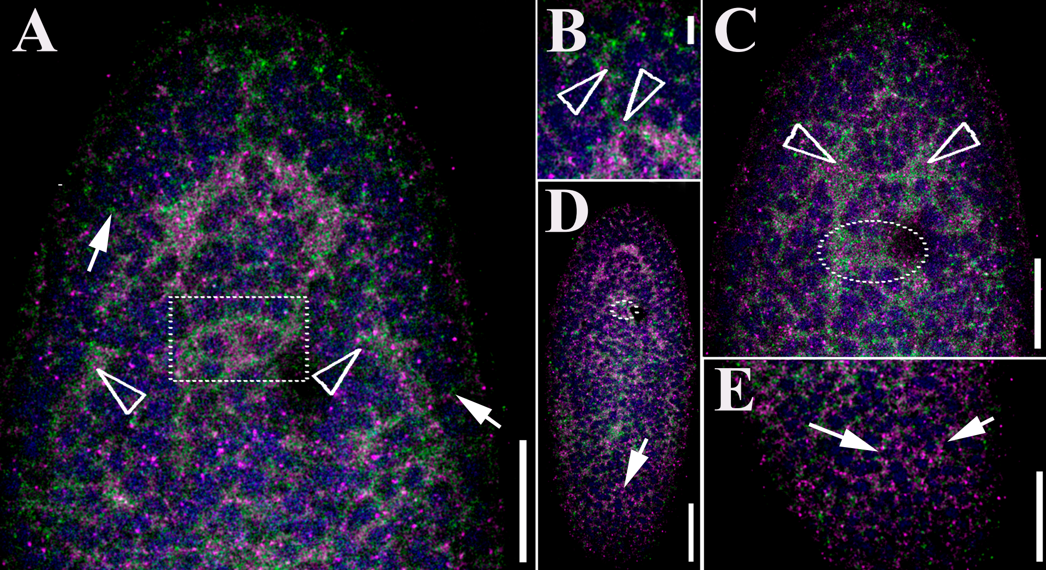


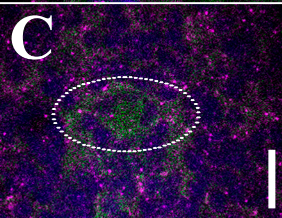


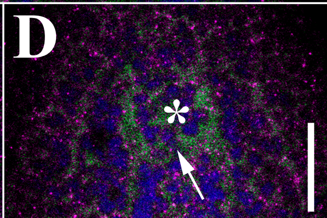


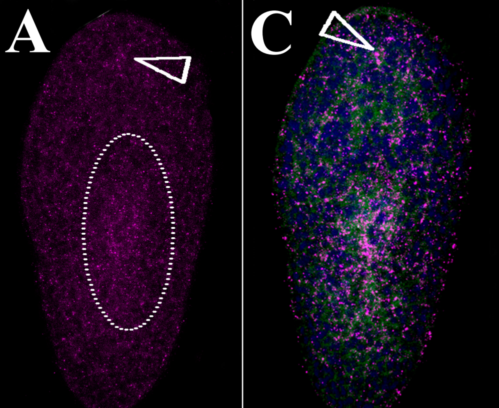
C

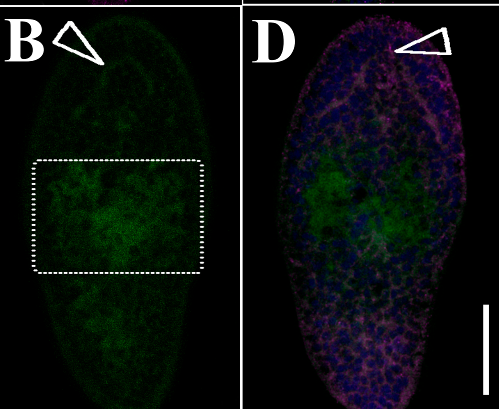


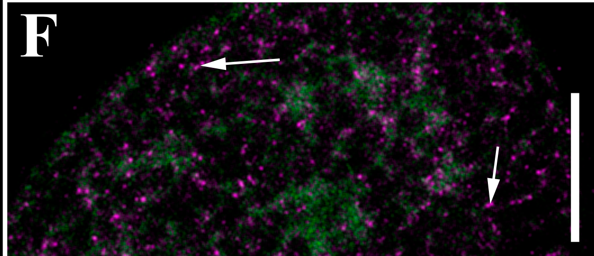

A
B
C
D
E


C


D


A
C


B
D


E
F

bioRxiv preprint doi: https://doi.org/10.1101/237388; this version posted December 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

bioRxiv preprint doi: https://doi.org/10.1101/237388; this version posted December 20, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

ns

m