Asymmetric Distribution of pl10 and bruno2, New Members of a Conserved Core of Early Germline Determinants in Cephalochordates

Simon C. Dailey1 †, Roser Febrero Planas2 †, Ariadna Rossell Espier2, Jordi Garcia-Fernàndez1 and Ildikó M. L. Somorjai1, 2*

1 Gatty Marine Laboratory, Scottish Oceans Institute, University of St Andrews, St Andrews, UK, 2 Department of Genetics, University of Barcelona, Barcelona, Spain

Molecular fingerprinting of conserved germline and somatic “stemness” markers in different taxa have been key in defining the mechanism of germline specification (“preformation” or “epigenesis”), as well as expression domains of somatic progenitors. The distribution of molecular markers for primordial germ cells (PGCs), including vasa, nanos, and piwil1, as well as Vasa antibody staining, support a determinative mechanism of germline specification in the cephalochordate Branchiostoma lanceolatum, similarly to other amphioxus species. pl10 and bruno2, but not bruno4/6, are also expressed in a pattern consistent with these other germline genes, adding to our repertoire of PGC markers in lancelets. Expression of nanos, vasa, and the remaining markers (musashi, pufA, pufB, pumilio, and piwil2) may define populations of putative somatic progenitors in the tailbud, the amphioxus posterior growth zone, or zones of proliferative activity. Finally, we also identify a novel expression domain for musashi, a classic neural stem cell marker, during notochord development in amphioxus. These results are discussed in the context of germline determination in other taxa, stem cell regulation, and regenerative capacity in adult amphioxus.

Keywords: amphioxus, posterior stem cells, evolution, regeneration, PGCs, preformation

INTRODUCTION

One of the key innovations coupled to the evolution of multicellularity was the ability to segregate the germline and the soma, with transcriptional repression of a somatic programme being key to maintaining the germ cell fate (Hallmann, 2011). Historically, two main mechanisms of germline specification have been defined in animals: preformation and epigenesis (Extavour and Akam, 2003), or determinative and inductive modes, respectively. In the determinative mode, cytoplasmic determinants associated with the germ plasm in the egg are inherited by a limited number of daughter cells during cleavage, which are thus specified as presumptive germ cells (PGCs), and go on to form the mature adult gametes. In contrast, during induction, inductive cues cause somatic cells to become specified as germline. Studies in mouse, axolotl, and cricket suggest that BMP signaling may be an ancient mechanism for PGC induction from mesoderm in animals (Chatfield et al., 2014; Donoughe et al., 2014). The phylogenetic distribution of these two mechanisms of germ cell specification suggests that the inductive mode may represent the ancestral state, and that germ
plasm has evolved independently multiple times (Blackstone and Jasker, 2003; Extavour and Akam, 2003; Johnson et al., 2003a, b; Crother et al., 2007; Extavour, 2007; Ewen-Campen et al., 2010).

Comparative studies in multiple taxa have revealed that the molecular signature of germ cells may often be shared across species that use both determinative and inductive modes of PGC specification, leading to the proposition of a conserved germline multipotency programme (Extavour, 2007; Juliano et al., 2010). Interestingly, some basal metazoans appear to use a combination of mechanisms to specify germ cells, and many of the classic germ line markers are in fact also expressed in adult somatic stem cells in these organisms (Alié et al., 2011; Leclère et al., 2012). Germline-associated genes are most often RNA-binding proteins, but there is considerable species-specific variation in the suite employed (see Gazave et al., 2013 for a compilation of much of the recent literature). However, a key core of proteins including Vasa/PL10, Tudor and a PIWI domain containing protein may represent an ancestral "pluripotency module" (Ewen-Campen et al., 2010). Tweaking upstream regulators or downstream targets, combined with the addition of new germ cell genes, such as nanos or bruno, would have generated the diversity in germ line specification mechanisms in early metazoans (Ewen-Campen et al., 2010).

Until recently, little was known about germ line specification in cephalochordates (lancelets or amphioxus), the sister group to the vertebrates and tunicates, and the best living proxy for the ancestral chordate (Bertrand and Escrivà, 2011). Classic studies suggested that lancelets might employ an inductive mode of PGC specification (reviewed in Extavour and Akam, 2003). Electron microscopy data however showed that in Branchiostoma floridae, the pole plasm localizes to the vegetal cortex soon after fertilization and segregates into a single blastomere during cleavage, putting into question this hypothesis (Holland and Holland, 1992). Although functional data are still lacking, blastomere separation experiments combined with expression data for molecular markers traditionally associated with the germ line, including piwi-like1, nanos, vasa, and Vasa protein, strongly support a determinative mode of PGC specification in cephalochordates (Wu et al., 2011; Zhang et al., 2013 and Figure 1). Zygotic expression domains from gastrulation onwards of these genes, as well as piwi-like2 and tudor7, also suggest a function in somatic progenitors/stem cells of the posterior growth zone. Together, these data provide a general framework for understanding how markers for PGCs and posterior progenitors may be expressed in cephalochordates during development (Figure 1).

Currently, the most convincing evidence for the existence of somatic stem cells in cephalochordates comes from studies of tail regeneration in the European amphioxus, Branchiostoma lanceolatum, whose adult regenerative ability is comparable to that seen in many ambulacrarians (echinoderms and hemichordates; Somorjai et al., 2012a, b). Unfortunately, no germ cell markers have so far been characterized during development in this species, and few putative somatic stem cell markers exist in cephalochordates. The purpose of this study was therefore threefold: First, to characterize the early expression of candidate amphioxus germ line markers nanos, piwill1, vasa, and Vasa protein in B. lanceolatum for comparative purposes with other cephalochordates; second, to determine whether germ line markers in other taxa, including pumilio, pufA, pufB, musashi, pl10, bruno2, and bruno4/6, are associated with PGCs in amphioxus; and third, to analyse the late developmental expression of some of these candidates as a prelude to future regeneration studies. Given the considerable conservation in developmental gene expression patterning in cephalochordates demonstrated thus far (e.g., Somorjai et al., 2008; Wu et al., 2011; Zhang et al., 2013), we hypothesize that markers for PGCs and posterior somatic domains will show comparable gene expression profiles in B. lanceolatum to B. belcheri, B. japonicum, and B. floridana.

Here, we present the first analysis of putative germ line and somatic stem cell markers in the European amphioxus, B. lanceolatum. We identify a core set of conserved PGC-associated markers in cephalochordates, including piwill1, nanos, and vasa, characterize Vasa protein distribution, and identify two new candidate germ cell markers in cephalochordates, pl10, and bruno2. We also characterize the amphioxus musashi ortholog, whose expression in the notochord represents a novelty in chordates. The highly conserved molecular expression data in the Branchiostomatiidae support the view that cephalochordate evolution is strongly constrained, and show that data are broadly transposable across species, even in the context of germ line formation. This study also provides the foundation for future studies of regeneration in the amphioxus B. lanceolatum.

MATERIALS AND METHODS

Embryos

Ripe adults were collected in Argelès-sur Mer (France) and spawned as previously described (Fuentes et al., 2007). Embryos were fixed at the relevant time points in 4% PFA in MOPS salts (0.1 M MOPS, 2 mM MgSO4, 1 mM EGTA, and 0.5 M NaCl), and stored in 70% ETOH at −20°C. For phalloidin staining, embryos were stored in PBS at 4°C. Embryos were staged according to Hirakow and Kajita (1991, 1994), with modifications as per Zhang et al. (2013).

Phylogenetic Analysis

If not previously published, putative orthologous sequences were identified using a BLASTp search; reciprocal BLAST was used to confirm identity (Camacho et al., 2009). Protein sequences were aligned in Jalview version 2.8.2 (Waterhouse et al., 2009) using MAFFT on default settings, and checked manually. All positions with less than 95% site coverage were eliminated directly in MEGA5 prior to analysis. Evolutionary models considered to best describe the substitution pattern were identified as those with the lowest BIC (Bayesian Information Criterion) scores using MEGA5 (Tamura et al., 2011). Both neighbor joining (NJ) and maximum likelihood (ML) analyses were performed with 500 and 1000 bootstraps, respectively. The Nearest Neighbor Interchange method was used to infer trees in ML. Unless otherwise noted, and since concordant with results from the NJ method, only ML trees are shown. Model details for each analysis are included in the figure legends for ease of reference. All
sequences used for phylogenetic analyses, including associated accession numbers, are included in Supplementary File 1.

Cloning and Probe Synthesis

RNA was extracted from embryos and adult tissues using Trizol and phenol chloroform extraction; cDNA was generated using Tetro cDNA Synthesis kit (Bioline). Gene fragments for probe generation were amplified by PCR using gene-specific primers designed against the genome of *B. floridana* (Supplementary File 2), ligated into PGEMT–Easy (Promega) and transformed into XL10-Gold (Stratagene) or DH5α (Invitrogen) strains of *E. coli* by heat shock using standard protocols. Selected clones were mini-prepped using peqGOLD or Promega plasmid miniprep kits, and sequence verified. Template was generated by PCR on plasmids using Universal M13F (5′ GTAAACGACGGCCAGT 3′) and M13R (5′ AACAGCTATGACCATG 3′) primers. The band was gel-purified using either the QIAQuick (Qiagen), GFX (Amersham), or Isolate II (Bioline) gel extraction kits following manufacturers guidelines. DIG-labeled (Roche) antisense probes were *in vitro* transcribed using T7, T3, and SP6 enzymes as appropriate following standard protocols. Probes were checked by agarose electrophoresis and purified using miniQuick Spin columns (Roche) or via precipitation with sodium acetate (3 M, pH 5.2) and ethanol.

Whole Mount *In situ* Hybridization (WMISH)

WMISH was performed as previously described (Somorjai et al., 2008). Briefly, fixed embryos were washed in PBT (0.1% Tween), and permeabilized using proteinase K (7.5 mg/ml) for empirically-tested periods based upon embryo stage and enzyme batch. Embryos were postfixed for 40 min in PFA, deacetylated in acetic acid in triethanolamine (0.1 M, pH 8), and pre-hybridized at least 2 h in hybridization solution. Embryos were incubated overnight with shaking at 60–65°C depending on probe. The first post-hybridization washes were performed at the hybridization temperature, with subsequent washes at room temperature in decreasing concentrations of SSC. An RNase step was included (37°C). Embryos were incubated overnight in primary antibody (anti-DIG AP, Roche), pre-adsorbed at 1:3000, with rocking at 4°C. Copious washing in PBT was performed between each step. For the chromogenic reaction we used either BM Purple (Roche) or NBT/BCIP (Roche); embryos were postfixed in PFA for 20 min when the signal/background was deemed appropriate. At least three WMISH were performed for each gene, on 10–50 embryos per stage in total. Embryos were mounted in 80% glycerol/20% PBS, and photographed under a Leitz DMRB microscope (Leica Microsystems) with Normarski optics. Photographs were taken with the Retiga 2000R camera and the QCapture software suite (QImaging), and processed in Adobe Photoshop CS3.

Immunohistochemistry

Immunohistochemistry and Alexa-fluor 568-labeled phalloidin stainings for F-Actin (Invitrogen, 1:400) were carried out as per Somorjai et al. (2012a). Briefly, after fixation, embryos were washed in PBT (phosphate buffered saline plus 0.1% Tween, pH 7.6), and permeabilized in PBS with 0.2% Triton-X for 40 min. After copious washing in PBT, embryos were incubated overnight at 4°C in primary antibody. Embryos were again washed in PBT and incubated in secondary antibody or phalloidin for 2 h at room temperature, or overnight at 4°C. A specific *B. floridana* anti-Vasa antibody, generously donated by Dr Jr-Kai Yu, was used at 1:20,000 (Wu et al., 2011). Secondary antibodies were Alexa fluor 488 and 568 diluted at 1:400 (Molecular Probes). Embryos were mounted in Vectashield (VectorLabs) containing Hoescht 33342 dye to stain nuclei (1:2000 of 10 mg/ml). Confocal images were taken on a Leica TCS SP8 confocal microscope, and processed using NIH ImageJ 1.48 d and Adobe Photoshop CS3.
RESULTS

Identification of Candidate Germline and Somatic Stem Cell Markers in B. lanceolatum

We selected DEAD-box (Vasa, Pl10), Pumilio domain (Pumilio, PuF, PuB), PIWI domain (Piwi1, Piwi2) RRM (Musashi), CELF (Bruno2 and Bruno4/6), and Nanos families as candidate germline and somatic stem cell markers for analysis in B. lanceolatum. When B. floridæ orthologs had not been previously characterized in the extensive phylogenetic analyses of Kerner et al. (2011), we identified putative stem cell markers using BLASTp searches against the genomes of B. floridæ and B. belcheri, and confirmed the identity of our B. lanceolatum proteins by comparison with published sequences in other cephalochordates (Supplementary File 4), including transcriptomic data from B. lanceolatum in the NCBI TSA database (Oulion et al., 2012). We generated phylogenies that include, where possible, sequences from more than one amphioxus species to support the identity of these proteins (Supplementary File 5; and see below). We then cloned partial sequences of orthologs in B. lanceolatum using primers designed in its sister species B. floridæ (Supplementary File 2). Using this approach, we successfully cloned 12 genes (including two piwil1; not shown) with known function in the germline or somatic cell lineage (Table 1). While previous phylogenies show that the distinction among Piwi clades is unequivocal (Kerner et al., 2011), the evolutionary history of piwi genes in cephalochordates is more complex. We identified a single piwil2 (piwila in Kerner et al., 2011) and three piwil1 (piwilB in Kerner et al., 2011) genes in the genomes of B. belcheri and B. floridæ. The latter belong to an apparent tandem duplication cluster (not shown and Yue et al., 2015) that appears to be present in all Branchiostoma, as we successfully cloned two of the three paralogs of piwil1 in B. lanceolatum. We also identified an ortholog of piwiX (“piwilike” in Zhang et al., 2013), but have been unable to clone the gene in B. lanceolatum. While EST data collected in NCBI and B. floridæ EST databases (Yu et al., 2008) support the expression of piwil1 and piwil2 (Supplementary File 3), we have not identified any expression data for piwiX in any database, including our own tail regenerate transcriptome dataset (Dailey and Somorjai, unpublished).

We also cloned partial pi10, vasa, nanos, Bruno2 (brunoB or CELF2 in Kerner et al., 2011), Bruno4/6 (brunoA or CELF4/5/6 in Kerner et al., 2011), PufA, PufB, and Pumilio sequences. The phylogenetic analyses broadly confirm previous studies (Kerner et al., 2011), though we could only confirm the existence of single A-type and B-type Bruno sequences. In most cases we could identify B. belcheri orthologs for the B. floridæ proteins, in addition to several B. lanceolatum sequences (Supplementary Files 4, 5). EST data in B. floridæ also supported the expression of these putative germline and somatic stem cell markers (Supplementary File 3).

As we were interested specifically in stem cell-related Musashi, and relationships among Musashi-related protein families are complex (Gasparini et al., 2011) we generated phylogenies utilizing the available full length B. floridæ and B. belcheri sequences, and included putative Saccoglossus kowalevski orthologs. We clearly identified sequences belonging to the TARDBP43 and hnRNP A3/hnRNP D clades (Figure 2). The close relationship between Musashi-like and DAZAP proteins is also strongly supported by this analysis, although the branching order is unclear particularly within DAZAP sequences and in basal metazoans. Notably, we were unable to find an amphioxus sequence with convincing affinity to DAZAP/hnRNP 27 genes in either species. We did however identify a Musashi-like sequence in both B. belcheri and B. floridæ (Figure 2, Supplementary File 4). In spite of the relatively low support for the Musashi clade, most likely due to the inclusion of

FIGURE 2 | Phylogenetic analysis of the RRM domain containing protein family in animals, including Musashi-like, DAZAP, hnRNP, and TARDBP43 clades. Maximum likelihood analysis was performed in MEGAS with 1000 bootstrap replicates, indicated as a percentage at each node. The model used was rtREV + G with five rate categories on 173 sites. Branches are colored according to the level of node support; amphioxus species names are highlighted in blue and red, with the clade representing musashi genes boxed in blue. Protein names are taken directly from the literature where available. See text for details. Species name abbreviations are as follows: Aqu, Amphimedon queenslandica; Bfl, Branchiostoma floridæ; Bbe, Branchiostoma belcheri; Bsc, Botryllus schlosseri; Cgi, Crassostrea gigas; Cte, Capitella teleta; Dme, Drosophila melanogaster; Efl, Ephydatia fluviatilis; Hsa, Homo sapiens; Nve, Nematostella vectensis; Pdu, Platynereis dumerinii; Sko, Saccoglossus kowalevski; Tad, Trichoplax adherens; Tca, Tribolium castaneum; Tk1, Thelohanellus kiauiae; Xtr, Xenopus tropicalis.
Insect “Musashi” proteins, the amphioxus sequences groups with vertebrate and hemichordate sequences with strong support (85), in addition to the recently identified “real” Drosophila Musashi-related protein Rbp6 (Siddall et al., 2012). Insect “musashi” and Rbp6 may therefore represent clade-specific duplications in this group from a musashi-like ancestor. We therefore propose that the Rbp6/Msi sequences be referred to as Musashi-like (blue boxed region in Figure 2), and all others outside the clade as DAZAP. Based on this nomenclature and the firm position of amphioxus musashi among deuterostome sequences, we are therefore confident that we identified a musashi gene orthologous to vertebrate musashi1 and musashi2.

Candidate Marker Expression in Putative PGCs

Recently, expression patterns for putative germline markers have been described in three other species of amphioxus: B. floridæ, B. belcheri, and B. japonicum (Wu et al., 2011; Zhang et al., 2013). We therefore performed WMISH for piwil1, piwil2, vasa, and nanos orthologs in early developmental stages of B. lanceolatum. Figures 3A–D show the characteristic expression in single “points” from the two cell stage to the gastrula stage in all four genes with the exception of piwil2. In some cases, the morulae or gastrulae contained up to three points (not shown). By the early neurula stages, the punctate distribution may be masked by the zygotic tailbud expression (discussed below).

Another dead-box containing gene, pl10, has been implicated in germ cell specification, and in some cases regeneration, from sponges to annelids (Alié et al., 2011; Rebscher et al., 2012; Leininger et al., 2014; Kozin and Kostyuchenko, 2015). PL10 is closely related phylogenetically to the Vasa protein (Kerner et al., 2011), but expression of pl10 has so far not been described in any cephalochordate. We therefore cloned a clear pl10 ortholog in B. lanceolatum (Supplementary Files 4, 5) and determined its expression using WMISH. Like vasa, pl10 is expressed in a punctate pattern from fertilization until gastrula stages, consistent with a role in PGC specification or maintenance (Figure 3E).

We also determined early expression of markers of three other classes of RNA-binding proteins that we might expect to have a stem cell association based on reports in other species: the Pumilio domain containing genes pumilio, pufA, and pufB; the CELF/Bruno genes bruno2 and bruno4/6, and musashi (Gazave et al., 2013 and references therein). Up to gastrulation, pumilio, pufA, and pufB show no clear localization in the presumptive germline (Supplementary Files 6A–C). Interestingly, pufA ESTs are found in blastula-stage embryos, and we observe several independent but convincing instances in which pufA appeared to be expressed in a punctate distribution reminiscent of our other PGC-associated patterns in some cleavage stage embryos (Supplementary File 6A). Similarly to PUM domain containing genes, bruno4/6 was absent in B. floridæ EST databases, and showed no convincing expression until gastrulation (Figure 3G). In contrast, bruno2 showed clear and strong localization to nuage or PGCs (Figure 3F). No other marker analyzed had specific expression in the presumptive PGCs (Supplementary File 6), including musashi, which had diffuse ubiquitous expression at early stages (Supplementary File 6D; see Supplementary File 7 for sense control).

Vasa Protein Distribution is Consistent with PGCs and Somatic Progenitor Cell Domains in B. lanceolatum

Along with transcript expression, localization of Vasa is a hallmark of primordial germ cells (PGCs) in multiple species. In order to confirm the identity of PGCs in B. lanceolatum, we took advantage of the recent generation of an antibody against B. floridæ Vasa (Wu et al., 2011) to perform immunohistochemistry. Given its clear cross-reaction in several amphioxus species (Zhang et al., 2013), we reasoned that α-BfVasa should also label PGCs in the European amphioxus.

---

**Table 1** | Genes cloned in this study and accession numbers for all identified sequences in *Branchio/stoma* species.

<table>
<thead>
<tr>
<th>Gene name</th>
<th>B. lanceolatum</th>
<th>B. floridæ</th>
<th>Identity Bla/Bfl%</th>
<th>Bla transcriptome</th>
<th>B. belcheri</th>
<th>B. japonicum</th>
</tr>
</thead>
<tbody>
<tr>
<td>bruno2</td>
<td>KT354035.1</td>
<td>XP_002589327.1</td>
<td>64.91</td>
<td>J7864629.1</td>
<td>Unpublished data¹</td>
<td>N.D.</td>
</tr>
<tr>
<td>bruno4/6</td>
<td>KT354036.1</td>
<td>XP_002588860.1</td>
<td>98.91</td>
<td>Not found</td>
<td>N.D.</td>
<td>N.D.</td>
</tr>
<tr>
<td>musashi</td>
<td>KT354038.1</td>
<td>XP_002590753.1</td>
<td>96.42</td>
<td>J7855672.1</td>
<td>N.D.</td>
<td>N.D.</td>
</tr>
<tr>
<td>nanos</td>
<td>KT354040.1</td>
<td>ADM26639.1</td>
<td>96.64</td>
<td>Not found</td>
<td>AGI96000.1</td>
<td>AGI96005.1</td>
</tr>
<tr>
<td>piwi-like 1</td>
<td>KT354041.1</td>
<td>XP_002611937.1</td>
<td>66.31</td>
<td>J7877108.1</td>
<td>AGI96002.1</td>
<td>AGI96006.1</td>
</tr>
<tr>
<td>piwi-like 2</td>
<td>KT354042.1</td>
<td>AGI96997.1XP_002596179.1</td>
<td>85.46</td>
<td>J7858457.1</td>
<td>N.D.</td>
<td>N.D.</td>
</tr>
<tr>
<td>pi10</td>
<td>KT354043.1</td>
<td>XP_002593182.1</td>
<td>94.24</td>
<td>J7857642.1</td>
<td>ADD25830.1</td>
<td>Unpublished data²</td>
</tr>
<tr>
<td>pufA</td>
<td>KT354044.1</td>
<td>XP_002606015.1</td>
<td>95.08</td>
<td>JT902408.1</td>
<td>N.D.</td>
<td>N.D.</td>
</tr>
<tr>
<td>pufB</td>
<td>KT354045.1</td>
<td>XP_002607419.1</td>
<td>80</td>
<td>JT863417.1</td>
<td>N.D.</td>
<td>N.D.</td>
</tr>
<tr>
<td>pumilio</td>
<td>KT354046.1</td>
<td>XP_002601469.1</td>
<td>91.46</td>
<td>Not found</td>
<td>N.D.</td>
<td>N.D.</td>
</tr>
<tr>
<td>vasa</td>
<td>KT354047.1</td>
<td>ADM26640.1</td>
<td>84</td>
<td>JT861702.1</td>
<td>N.D.</td>
<td>AGI96004.1</td>
</tr>
</tbody>
</table>

Protein sequences predicted in genome or transcriptome assemblies are indicated by italicized accession numbers. Percentage identity of each B. lanceolatum clone is given relative to the most-complete available B. floridæ protein (Bla/Bfl). Abbreviations: Bla, B. lanceolatum; Bfl, B. floridæ; *, additional sequences; N.D., not determined. The sequence listed as “Unpublished data¹” is provided in Supplementary File 1 as “Bbe_Bruno2_076200F_001000_in,” and “Unpublished data²” as “Bbe_PL10_173980F_003600_in.”

---

¹ Unpublished data
² Unpublished data
confirming our expression data. The protein distribution resembles that of *vasa* transcripts (Figure 3D), with a pattern reminiscent of germplasm in fertilized eggs and cleavage stages (Figures 4A–D). In the late gastrula/early neurula, the protein is perinuclear in small clusters of cells within the ventral endoderm (Figures 4E,F); although variable in number (or at least detection), we could clearly identify as many as eight cells by the careful analysis of series of confocal image z-sections (Figure 4F). Such clusters could be identified even in some mid-neurula stage embryos, either on one side in the ventral mesoderm (Figure 4G and inset), or in most cases posteriorly congruent with the zygotic tailbud domain (Figure 4H and inset). Vasa expression was however most conspicuous in the posterior neural tube throughout neurulation (Figures 4I,J). Only in premouth stage and later larvae was it possible to again more easily identify posterior clusters of Vasa-expressing cells as distinct from posterior neural and tailbud expression (Figures 4K,L). Vasa also appeared to demarcate the posteriormost somites (not shown), similarly to *vasa* transcripts (Figure 5C, see below).

**Candidate Stem Cell Marker Expression in Developing Somatic Tissues**

We performed WMISH for selected genes from gastrulation onwards, reasoning that they should show expression patterns with possible roles in late developmental processes (Figure 5 and Supplementary File 8). We thus identified two classes:

<table>
<thead>
<tr>
<th>Candidate Stem Cell Marker Expression in Developing Somatic Tissues</th>
</tr>
</thead>
</table>
| **Early expression of putative somatic stem cell and germline markers in *B. lanceolatum*.** (A) *piwi-1*; (B) *piwi-2*; (C) *nanos*; (D) *vasa*; (E) *pl10*; (F) *bruno2*; (G) *bruno4/6*. Purple arrowheads indicate putative PGC-associated expression, as early as the fertilized egg, of *piwi-like1*, *nanos*, *vasa*, *pl10*, and *bruno2*. Stages are indicated along the top of the figure in panel (A), from egg to mid-gastrula stages (G3). In morula (ca 64-cell) and blastula stages, macromeres are oriented downwards. G3-bla, blastopore view; G3-lat, lateral view. Scale bars = 50 microns.
FIGURE 4 | Vasa protein expression during *B. lanceolatum* embryogenesis. Vasa is expressed in a single point in fertilized eggs (A), two cell (B), eight-cell (C), and monula (D) stages (arrows). (E,F) Perinuclear punctae are evident in late gastrula/early neurula (N0–N1) stages in up to eight cells (two different embryos shown, E and F). (E) shows a cluster of cells of which two clearly express Vasa (arrow). In (F), a series of z-stacks shows a cluster of eight cells expressing Vasa. Mid-neurula (N2) stage embryos (G,H) are variable, with some individuals showing a cluster of Vasa-positive cells in the ventral endoderm on one side (G, arrow and boxed inset); in others the putative germ cells are just ventral to and contiguous with the posterior Vasa expression domain (H, arrow and boxed inset). (I,J) germline expression is difficult to distinguish in late neurula (N3) stage embryos; zygotic expression is most apparent in the posterior neural tube and posteriormost ventral endoderm. (K,L) Clusters of cells are again evident in some pre-mouth stage and late larvae (L1, L2 arrows), and neural expression is much reduced. The dotted line in (K) indicates where two confocal images have been combined. Nuclei stained with Hoechst are false-colored magenta, and Vasa protein is shown in cyan. Acetylated tubulin marks axons and cilia in yellow in (E), (I–K). Phalloidin stains F-actin yellow in the membranes in (F) and (L). With the exception of scale bars in (F) and in boxed insets (G,H), which are 10 microns, all scale bars are 50 microns. Orientations are as follows: (E–L) anterior, left; (E,F,H) ventral views; (G,I–L) lateral, dorsal is to the top.

“tailbud-enriched” and, broadly speaking, “anterior endoderm-associated.” We found that *piwil1*, *nanos*, and *vasa* have strong tailbud expression throughout development (Figures 5A–C). *piwil1* and *nanos* also show clear posterior neural tube expression in N4 neurulae and L1 stage premouth larvae, as well as expression outlining the posterior somites (black arrowheads, Figure 5A; Supplementary File 8A). Though weaker, *piwil2* and *pl10* both show tailbud expression at later stages, and *piwil2* is clearly expressed in the neural tube (Supplementary File 9).

In contrast, Pumilio domain containing genes appear enriched in anterior endoderm (Figures 5D,E and Supplementary Files 8D,E). During gastrulation, *pumilio* shows weak expression around the blastopore. In early and mid-neurula stages, stronger expression is evident in the neural plate and anterior ventral endoderm, as well as anterior mesoderm. As neurulation proceeds, *pumilio* appears mostly restricted to the anterior endoderm, with expression much weaker in the last third of the embryo (Figure 5E). Expression continues to be strongest in the future pharyngeal domain until the pre-mouth larval stage. Weaker expression is evident in the rest of the endoderm, with some conspicuous staining in the mesoderm and endoderm of the tailbud region. Expression of...
**DISCUSSION**

**Germline-Associated Gene Expression Conservation in Cephalochordates**

Recent work in *B. floridæ, B. japonicum*, and *B. belcheri* has suggested that germline specification occurs by the asymmetric segregation of cytoplasmic determinants during cleavage, with expression of key conserved germline markers such as *vasa* and *nanos*, as well as *piwi1* and *tudor-related7*, in the germ plasm and PGCs (Wu et al., 2011; Zhang et al., 2013). We set out here to characterize the expression of germline-associated markers in the European amphioxus, *B. lanceolatum*, for which there were until now no data. Similarly to other species, our results also argue against an inductive mechanism for PGC specification: we demonstrate here that *B. lanceolatum* expresses *nanos*, *piwi1*, and *vasa* in the putative PGCs, as well as Vasa protein, suggesting the presence of a conserved core of germline-associated transcripts in cephalochordates. Stasis in developmental gene expression over millions of years of evolution is considered typical of *Branchiostoma* (Somorjai et al., 2008), paralleling the genus’ relative genomic and morphological conservativeness. The apparent conservation in germline-associated gene expression in amphioxus species is in stark contrast to hypotheses derived in vertebrates that suggest the evolution of germ plasm is coupled to increased speciation in this lineage (Johnson et al., 2011; Evans et al., 2014). Data in *Asymmetron*, the earliest diverging and most slowly evolving of the three extant amphioxus lineages (Kon et al., 2007; Yue et al., 2014), will be invaluable in evaluating the degree of conservation of germline specification mechanisms in cephalochordates.
Our research also identifies pl10, a DEAD-box gene related to vasa, and bruno2 as putative PGC markers in amphioxus. Accumulating evidence suggests that pl10 often plays a role in the germline in metazoans. In addition to Drosophila, pl10 orthologs are expressed in the germinal cells or their derivatives in the annelid Platynereis dumerilii (Rebscher et al., 2007; Gazave et al., 2013), the platyhelminth Dugesia japonicum (Shibata et al., 1999; reported as vasa-related genes) several hydrozoan cnidarian species (Leclère et al., 2012; Siebert et al., 2015), the ctenophore Mnemopsis leydiid (Alié et al., 2011), the sponge Sycon ciliatum (Leininger et al., 2014), and in the colonial urochordate B. schlosseri (Rosner et al., 2009). vasa is coexpressed with pl10 in the latter, similarly to our results in B. lanceolatum. In contrast, data are sparse for the second gene identified, bruno2. Homologs of bruno are expressed in PGCs and/or germline derivatives in ctenophores (Alié et al., 2011), but not in Platynereis (Gazave et al., 2013). Interestingly, using RNAi, the Bruno-like gene brul1 was shown to be required for maintenance of a subset of neoblasts in the asexual planarian Schmidtea mediterranea (Guo et al., 2006), but this gene is not homologous to canonical bruno genes. Confirmation of expression of pl10 and bruno2 in other species, and Tudor related tdrd7 in B. lanceolatum, will further expand this repertoire.

We also identified several species with weak ubiquitous expression during early development. For instance, musashi, piwil2, and genes of the Pumilio domain family do not appear to be associated specifically with PGCs in B. lanceolatum or B. floridae (this study; Yue et al., 2015). A possible exception is pufA, which we found to be concentrated in a PGC-like domain in some cleavage stage embryos (two-cell to morula) in several independent experiments. Given the variability of the expression observed, we hesitate to classify this as bone fide expression in PGCs. However, pufA is expressed in germ cells in other species, including zebrafish (Kuo et al., 2009) and P. dumerilii (Gazave et al., 2013). Interestingly, a global search of germline and reproduction-associated genes using the transcriptome and genome of Asymmetron lucayanum and B. floridae, respectively, identified a pumilio/puf gene with expression in oocytes (Yue et al., 2015). Similar studies in maturing gonads in B. lanceolatum may also reveal functions for some of our candidates during germline maturation.

### Evolution of Musashi Related RRM-Containing Proteins and Novel Expression of Amphioxus musashi

The Musashi related proteins belong to a larger superfamily of RRM containing proteins, including Musashi, DAZAP, hnrp, and TARDBP clades. Although the evolutionary history of RRM domain containing proteins is complex, orthologs of musashi-related genes have been identified from sponge to human (Gasparini et al., 2011; Okamoto et al., 2012), including lancelets (Gasparini et al., 2011, this study). One of the principal findings of this study is that cephalochordates appear to have lost the ortholog of DAZAP, as we were unable to identify the gene in either the genomes of B. floridae or B. belcheri. Considerable confusion exists in the nomenclature in the literature due to the difficulty in distinguishing between musashi and DAZAP related genes. This is particularly evident in basal metazoans, where phylogenetic signal is weak (Okamoto et al., 2012, this study). Gasparini et al. (2011) first suggested that previously identified musashi-like genes in Halocynthia roretzi and Ciona intestinalis (Kawashima et al., 2000) are in fact DAZAP. This gene is expressed in the brain and nerve cord, as might be expected from musashi-like genes (Kawashima et al., 2000). However, the bona fide DAZAP1 in Botryllus schlosseri is expressed both during asexual (spermatogenesis) and sexual (embryonic) development in many proliferating cell types, including the new growing vessels of the colonial circulatory system and the embryonic nerve cord, and is not restricted to neural stem cells as in other systems. Likewise, in the planarian D. japonicum DAZAP/musashi-like gene Djdmlg is expressed in differentiated tissues as well as X-ray sensitive neoblasts (Higuchi et al., 2008). In this context, it would be particularly interesting to determine whether the cephalochordate musashi is taking on any of the DAZAP functions, or whether a different functional homolog might be involved.

Our observation that neural cells within the developing CNS of amphioxus express musashi is broadly consistent with data in bilaterians. For instance, in the flatworm Dugesia japonica, three musashi-like genes have been identified with expression in the brain primordia (Higuchi et al., 2008). Similarly, in zebrafish, musashi1 is expressed in neural tissues during early development, and knockdown by morpholino results in aberrant CNS formation (Shibata et al., 2012). Surprisingly however, amphioxus does not express musashi in a pattern consistent with a role in PGC specification or maintenance, in contrast to many other taxa. In Drosophila, musashi is required to maintain stem cell identity in GSCs (Siddall et al., 2006), and Rbp6, which is more closely related to vertebrate musashi1/2 (Siddall et al., 2012; this study), also may play some function in the germ line (Siddall et al., 2012). In mice, the msi1 and msi2 orthologs appear to have sub-functionalized such that msi1 is required to maintain stem cell identity during early spermatogenesis, whereas msi2 plays a role in differentiation (Siddall et al., 2006). The generation of specific antibodies will be critical to gaining an understanding of the distribution of Musashi protein during amphioxus development and stem cell regulation.

Given its known neural and germline functions, the finding that musashi is predominantly expressed in the developing notochord in amphioxus was unexpected. We are not aware of any data demonstrating a specific function for musashi in the notochord in any chordate. However, the ancestral function of these RRM containing proteins may simply be in the switch between undifferentiated/stem cell and differentiated cell types and in the regulation of proliferation (Potten et al., 2003; MacNicol et al., 2011; Hochgreb-Hägele et al., 2014). Supporting this, the anterior endoderm encompassing the zones that will form the mouth and gill slits in amphioxus larvae, which has conspicuous musashi expression, is a zone of extensive proliferation and remodeling (Holland and Holland, 2006). The expression in developing notochord described here, which is unique to amphioxus, might also reflect a role in differentiation...
of this structure. Functional studies will help elucidate the role of Musashi in this and other structures.

**Posterior Stem Cells and Implications for Amphioxus Regeneration**

The zygotic expression of several markers in the tailbud, including *nans*, *vasa*, *piwil1*, and *piwil2* among others, combined with circumstantial evidence that PGCs may migrate at the neurula stage toward the posterior (this study; Wu et al., 2011), suggest that the tailbud may be a source of progenitors or stem cells in larval amphioxus. Posterior elongation in amphioxus involves budding of somites directly from the tailbud, a source of Wnt ligand (Holland et al., 2000; Schubert et al., 2000, 2001; Somorjai et al., 2008). Although architecturally different, the tailbuds of vertebrates like mouse and chick are also sources of multipotent stem cells for embryonic elongation whose fate is Wnt signaling-dependent (Wilson et al., 2009; Garriock et al., 2015). The posterior growth zone may also act as a niche for progenitor cells even into adulthood, particularly in animals that add segments throughout their lives, such as arthropods and most annelids (Bely and Wray, 2001; de Rosa et al., 2005; Seaver et al., 2005). The observation that the Vasa-positive PGCs lie within a stem cell marker-expressing posterior growth zone in amphioxus larvae (Wu et al., 2011; this study), representing a “mosaic” of PGCs and somatic stem cells, is however not unique to amphioxus. Gazave et al. (2013) have proposed the existence of an RNA binding protein signature for a new type of animal stem cell, termed “posterior stem cells,” in *P. dumerilii*. Lineage analysis and EdU labeling have also revealed that the 4 presumptive PGCs, which appear during gastrulation, are derived from a mesoderm posterior growth zone (MPGZ; Rebscher et al., 2007, 2012). While the mechanisms employed by these anelids and cephalochordates to specify the germine are somewhat different, the use of such techniques in amphioxus will be instrumental in elucidating the origin and fate of different cell types during posterior elongation.

The existence of a posterior stem cell in the tailbud, or any other resident stem cell population that could be activated following tail amputation, has clear implications for regeneration in amphioxus. Although it has recently been demonstrated that the European amphioxus has considerable regenerative ability, most notably of the tail (Somorjai et al., 2012a,b), we still know next to nothing about the molecular signature or function of the somatic stem cells/progenitor pools involved in the process. This study represents the first step toward identifying a putative posterior stem cell pool in *B. lanceolatum*. Our prediction is that somatic stem cell markers that are normally expressed during tailbud development, such as *vasa*, *nans*, *piwil1*, or *piwil2*, will also be expressed during the adult tail regeneration process. We are currently analyzing blastema transcriptomes and proteomes to test this hypothesis (Dailey and Somorjai, unpublished). We might also expect to find genes traditionally associated with the germline to be expressed during tail regeneration, if common expression of “stemness” markers in PGCs and somatic stem cells reflect broader roles in developmental regulation, as has recently been demonstrated for Vasa in the sea urchin (Yajima and Wessel, 2015). Although functional experiments are lacking, comparative expression data in annelids are beginning to provide compelling evidence for this. In *P. dumerilii*, a number of RNA binding protein genes are expressed in PGCs as well as in putative posterior mesodermal and ectodermal stem cells during caudal regeneration, including *vasa*, *pi10*, *piwi*, *pufA*, *pufB*, *nans*, and several *tudor* related genes (Rebscher et al., 2007; Gazave et al., 2013). Of these, several markers are also differentially expressed both in the germline and terminal growth zone during normal development and regeneration in the polychaetes *Alitta virens* and *Capitella sp I* (Dill and Seaver, 2008; Giani et al., 2011; Kozin and Kostyuchenko, 2015). However, the most striking example of a germline-independent redeployment of classic PGC markers in somatic tissues has been shown in the freshwater annelid *Pristina leidyi*, which reproduces exclusively asexually in the laboratory via paratomic fission. As might be expected, *nans*, *piwil1*, and *vasa* are expressed in the posterior growth zone and developing (but unused) gonads. Notably, transcripts are also detected following amputation in the *anterior* blastema as well as the fission zone (Bely and Sikes, 2010; Ózpolat and Bely, 2015), highlighting a more general role in tissues undergoing proliferation and remodeling. This phenomenon is not restricted to invertebrates or basal metazoans, as *piwil1* and *piwil2* are expressed in a complex spatiotemporal sequence during axolotl limb regeneration, with knockdown of either gene resulting in retardation of the regenerate outgrowth (Zhu et al., 2012). Future work in amphioxus will assess the tissue-specific expression pattern of some of the candidates identified here during adult tail regeneration. Development of knockdown tools and lineage analysis will be indispensable to elucidate their functional role during the regeneration process. Moreover, these methodologies will permit the comparative analyses of cellular and molecular processes necessary to understand the evolution of regeneration mechanisms in deuterostomes. More broadly, these types of studies should add to the growing body of literature aimed at understanding the link between soma and germine evolution.

**AUTHOR CONTRIBUTIONS**

SD, RF, and AR performed experiments. JG discussed experiments and contributed reagents. IS conceived the study, performed experiments, contributed reagents, analyzed the data and wrote the manuscript.

**ACKNOWLEDGMENTS**

This work was carried out sporadically over the course of several years in various countries. We would like to thank Jir-Kai Sky Yu (Academia Sinica, Taiwan) for contributing the Vasa antibody, and Irene Garcia for help in the laboratory (Barcelona). SD is funded through a MASTS PhD studentship (St Andrews, UK). IS gratefully acknowledges previous funding from Marie Curie IEF postdoctoral fellowship, FP7 People Programme (Barcelona, Spain); as well as MASTS (Marine Alliance for Science and Technology Scotland) laboratory start-up funds (St Andrews, UK). Embryo collection was made possible in part through the ASSEMBLE access programme (grant agreement
no. 227799). We thank the Laboratoire Océanologique de Banyuls-sur-Mer, and most especially Dr. Hector Escrivà and Dr. Stéphanie Bertrand for hosting us. The University of St Andrews Library fund for open access supported the article publishing fee.

**REFERENCES**


**SUPPLEMENTARY MATERIAL**

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fevo.2015.00156