HAL
open science

Amphioxus illuminates the origin of the vertebrates' head

Daniel Aldea

To cite this version:

Daniel Aldea. Amphioxus illuminates the origin of the vertebrates' head. Development Biology. Université Pierre et Marie Curie - Paris VI, 2016. English. NNT : 2016PA066704 . tel-01879096

HAL Id: tel-01879096

https://theses.hal.science/tel-01879096v1

Submitted on 22 Sep 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

University Pierre and Marie Curie

Doctoral School 515

Team “Evolution and development of Chordates”
BIOM UMR7232

Amphioxus illuminates the origin of the vertebrates’ head

Thesis submitted for the degree of
Doctor of Philosophy (Ph.D) in Developmental Biology and Evolution

by Daniel Aldea

Thesis directors: Mr. Héctor Escrivà and Ms. Stephanie Bertrand

Thesis defended in public on September 20th, 2016

PhD Committee:

Mr. Vincent Laudet	President of the Jury
Mr. Robert Kelly	External examiner
Mr. Ricard Albalat	External examiner
Ms. Estelle Hirsinger	Internal examiner
Ms. Mathilde Paris	Internal examiner
Mr. Héctor Escrivà	Thesis director
Ms. Stephanie Bertrand	Thesis co-director

Acknowledgments

(Purposely written in Spanish)

Primero que todo quiero dar las gracias a Héctor y Stephanie, por recibirme en su laboratorio durante estos 4 años. Gracias por guiarme en este camino llamado tesis, por darme las herramientas para resolver preguntas a través de experimentos, por enseñarme el cómo explicar mejor las cosas y a la vez ordenar mis ideas en mi cabeza. Luego y por sobre todo lo demás, gracias por la buena atmosfera que existe en el laboratorio, definitivamente las noches y noches de trabajo no hubiesen sido igual sin ello. A Sylvain gracias por enseñarme biología molecular, sin lugar a dudas mi paso por el LADE me lleno de excelentes herramientas para mi futuro profesional. A mis compañeros de laboratorio, Lucie, gracias por toda la ayuda en experimentos; Anthony, por tu humor con tus videos y Yann por tu humor bretón. Por el laboratorio pasaron muchas personas y de cada uno de ellas he aprendido algo, seguro nombres se me quedaran en el tintero; Helene, Marc, Mathieu, Hugo L., Hugo B., Margot, Marine, Denisse, Mariana, Mathias, Laetitia, Stephan, Nataly, Sandrine, Elsa, Sandra, Nacho, Fanny, Eleonora, Perrine, Brieuc, Sheree, Medhi, Amandine, Valeria, Tatiana, Remy, y Claire gracias a ustedes y los muchos asados, fiestas, cervezas y salidas que hicieron de mi paso por Banyuls, unos años muchos más que agradables. A Phillipe, Christophe, Jean-Loius, Christian, Emilie y Vanda por su excelente humor en la cocina y alrededores día a día.

A la tropa de chilenos que pasaron por Banyuls, desde mis primeros días y hasta el último haciendo que me sintiera como en casa, Carlos, Estrella, Valentina, Marilo, Claudia, Benjamín, Tomas y en especial a María Jesús y Belén. A Banyuls-sur-mer mismo por regalarme los atardeceres más bellos que he visto y entregarme maravillosos parajes para hacer bicicleta, y aquí gracias a Lucie Jaugeon por acompañarme en cada ruta durante estos últimos meses.

Te mereces un párrafo especial, porque si no fuera por ti este último año no hubiese sido lo mismo, Sarah, muchas gracias por escucharme, reírte de mis bromas, y acompañarme en esta última parte de la tesis, tendremos muy buenos momentos grabados para siempre. No importa que nos depara el futuro, juntos o separados, siempre estaré deseando lo mejor para ti.

Por último y los más importantes, a mi familia por ser siempre la piedra base de mi vida, muchísimas gracias por apoyarme en cada decisión que he tomado. Gracias por hacerme sentir querido y acompañado, incluso en un comienzo cuando las cosas no se daban bien. Sé que estarán por mí en cada paso y en el siguiente. A mis Papas, sin duda yo no estaría aquí si no fuera por ustedes. Esta tesis es por y para ustedes.

A todos ustedes y a los que seguro he olvidado.

GRACIAS TOTALES!!!

Index

Acknowledgments 2

Preface 7

1. Introduction 8

1.1 HYPOTHESES FOR THE ORIGIN OF THE VERTEBRATES' HEAD 8

1.1.1 Segmentalist Hypothesis 8

1.1.2 Non Segmentalist Hypothesis 10

1.1.3 The "New Head" Hypothesis 12

1.1.4 Final remarks 16

1.2 AMPHIOXUS AS A MODEL 17

1.2.1 Identification and Description of Amphioxus 17

1.2.2 Phylogenetic Position and Amphioxus Species 18

1.2.3 Environment 20

1.2.4 Reproduction and Life Cycle 20

1.2.4 Embryonic Development 21

1.2.5 Amphioxus as an Animal Model in EVO-DEVO 25

1.2.6 Final remarks 28

1.3 SOMITOGENESIS 29

1.3.1 Somitogenesis in vertebrates and amphioxus: a morphological description 29

1.3.2 Molecular Control of Somitogenesis 31

1.4 MYOGENESIS 37

1.4.1 Genetic control of truncal myogenesis in vertebrates and amphioxus 37

1.4.2 Formation of Facial Muscles in Vertebrates 41

1.4.3 Orthologues of vertebrate head muscles formation in amphioxus 44

1.4.4 Final remarks 45

1.5 FGF SIGNALING PATHWAY AND ANTERIOR SOMITOGENESIS IN AMPHIOXUS 46

1.5.1 FGF signaling pathway 46

1.5.2 Expression of FGFs in amphioxus 47

1.5.3 FGF signaling pathway controls anterior somitogenesis in amphioxus 48

1.5.4 Final remarks 50

2. Introduction of Article 1 51

2.1 Evolution of the role of RA and FGF signals in the control of somitogenesis in Chordates 53

3. Introduction Article 2 67

3.1 Anterior somitogenesis in amphioxus sheds lights on the origin of the vertebrates' head 68
4. General discussion and additional data 111
4.1 Hox genes, FGF signal and the anterior somitogenesis in amphioxus 111
4.2 RNA-seq analysis and the role of the FGF signal in anterior somitogenesis 112
4.3 Possible role of upregulated genes in anterior somitogenesis 116
4.4 Six1/2 and Pax3/7 control somitogenesis in amphioxus..... 117
4.5 FGF signal and vertebrates head mesoderm..... 118
4.6 Our results under the context of the evolution of the vertebrates' head 119
5. Annex Articles..... 121
5.1 Expression of Fox genes in the cephalochordate *Branchiostoma lanceolatum*..... 122
5.2 A single three-dimensional chromatin compartment in amphioxus indicates a stepwise evolution of vertebrate Hox bimodal regulation 135
6. Bibliography..... 145

Preface

"Nothing in biology makes sense except in the light of evolution"

Theodosius Dobzhansky (Dobzhansky, 1973)

After obtaining its own division in the Society for Integrative and Comparative Biology (SICB), the Evolutionary Developmental Biology (Evo-Devo) emerged officially as a discipline in 1999. However, the combination between developmental biology and evolutionary biology took place at least one decade before, when biologists began to look at and compare the expression patterns of developmental genes in different organisms. Unexpectedly, when they analyzed the expression of *Hox* genes at early stages in embryos of invertebrates (fruit fly) and vertebrates (mice), they found a conserved expression pattern in the antero-posterior axis among these animals. This was the beginning of fruitful decades for this "new discipline". In general terms, the Evo-Devo tries to unravel the evolutionary scenarios in which, from a unique ancestor, the appearance of all the morphological/anatomical characteristics and shapes observable today occurred.

The body of vertebrates is characterized by a highly specialized anterior structure called "the head". First attempts to understand the evolutionary origin of the vertebrates' head date from the beginning of the 19th century. At that time, the discussion was taken under the framework of comparative embryology and the first hypotheses were postulated. Unravelling how novelties arise during evolution is one of the major tasks in Evo-Devo, thus in the last decades efforts to understand the origin of the vertebrates' head have brought new hypotheses to the scenario but many questions still remain to be fully clarified.

Since the appearance of the "new genomic era" in the early 2000s, many high-throughput tools have been developed and other classical tools have been improved. Thus, taking advantage of these new technologies it seems more than ever necessary to unravel the origin of the vertebrates' head (nothing in evolution makes sense except in the light of these new technologies). Under this framework, in the following pages, I will present my work using a cephalochordate (amphioxus) as an animal model to shed light on the origin of the vertebrates' head. In other words, in the context of the Evo-Devo discipline, through a comparative approach between amphioxus and vertebrates, and using both classical developmental biology and recent high-throughput techniques, I tried here to make one more step in the understanding of the evolutionary changes that precluded the evolution of the vertebrate's head.

1. Introduction

Unlike the other chordates (i.e. urochordates and cephalochordates) the vertebrates possess an extremely specialized structure in the anterior part of their body called "the head". This structure is composed by skeletal structures, muscles, primary sensory organs (vision, taste, smell, hearing and balance), and a complex organ (the brain) that process the information coming from outside. The arising of this novelty (the head) over more than 500 million years ago paced the transition between a filter-feeding to a predator life-style, and supposed the appearance of the first vertebrates. Even if no demonstration exists, it is extensively accepted that the last common ancestor of all chordates possessed its body completely segmented from the most anterior to the most posterior part of the body. However, the head of the vertebrates is unsegmented, even if some nerves appear as segments. At contrary to the trunk of the vertebrates, where the mesoderm shows a clear segmentation brought by the somites (structures derived from the paraxial mesoderm). In this first part of the introduction I will present you the three major postulated hypotheses about the origin of the vertebrates' head during the last centuries, as well as supporting data or controversial points for each of these hypotheses.

These three hypotheses are:

- Segmentalist hypothesis (it claims that the body -the head and the trunk- of vertebrates is formed by the same segmental process)
- Non segmentalist hypothesis (it claims that the head is formed by a process different from the trunk)
- The "New Head" hypothesis (it claims that the head is a completely new structure originated thank to the appearance of neural crest cells and placodes)

1.1 HYPOTHESES FOR THE ORIGIN OF THE VERTEBRATES' HEAD

1.1.1 Segmentalist Hypothesis

We can attribute the first ideas about the origin of the vertebrates' head to Goethe, who inspired by the thoughts of the German philosopher Immanuel Kant, proposed a model where the body of all living animals consists of equivalent segments (vertebrae), and the skull represents a modified part of these segments (vertebral theory) (**Figure 1**) (Goethe, 1790). The idea of an "archetype" proposed by Goethe and from which any type of living animals could be derived by simple modifications was, later on, modified by the British zoologist Owen, who proposed an extreme "archetype" for vertebrates, where the final drawing showed an animal that possesses different characteristics coming from derived and ancestral traits. In this way Owen proposed that the formation of the anterior portion of the body (head) is similar to the posterior one (trunk), meaning that the whole body is formed in a metameric fashion (repeated segments) (**Figure 1**) (Owen, 1854).

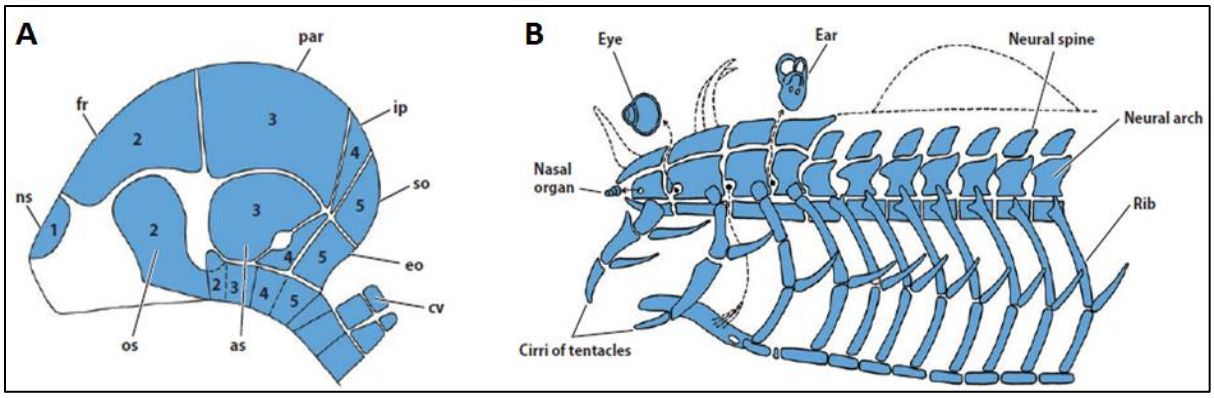


Figure 1. Early hypotheses of head segmentation. (A) The vertebral theory of Goethe. Goethe proposed that similarly to the vertebrae of the body, the skeleton of the mammalian skull is segmented. Then the different bones of the head would represent five head segments (1 to 5), (modified from (Jollie, 1977)). **(B)** The vertebrate archetype of Owen, who proposed that the anterior portion of the vertebrate body is similar to the posterior part of the vertebrate body (trunk), (modified from (Owen, 1854)). Abbreviations: ns, nasal; fr, frontal; par, parietal; ip, interval parietal; so, supraoccipital; eo, exoccipital; cv, cervical; as, aliphenoid; os, orbitosphenoid.

Decades later, at the end of the 19th century, Balfour described for the first time "head cavities" in shark embryos. In fact, he found three pairs of cavities (premandibular, mandibular and hyoid), that he compared with the somitic coeloms (cavities) of the trunk (Balfour, 1874; Balfour, 1876). At the beginning of the 20th century, Koltzoff and Damas described head somites during the development of lamprey embryos (the earliest divergent group of vertebrates), supporting the idea of a segmented origin of the head (Damas, 1944; Koltzoff, 1902). At the same time, Goodrich, who was one of the major proponents of a segmental structure of the head, proposed the presence of primary mesodermal segments in the head comparable to the somitic segments found in the trunk. He also claimed that the head was segmented into eight units, which represent a primitive condition of jawed vertebrates, and that the ancestor of all vertebrates was an amphioxus-like creature (Goodrich, 1918; Goodrich, 1930) (**Figure 2**). Notably, the body of extant amphioxus is completely segmented from the most anterior to the most posterior part, and possesses a similar morphology to vertebrates but much simpler. These reasons pushed zoologists at that time to think that amphioxus represented the primitive state of all vertebrates. In sum, unlike Goethe and Owen, who considered only skeletal elements in their hypotheses, Goodrich's model integrated nerves, muscles, and pharyngeal elements into one single metamere.

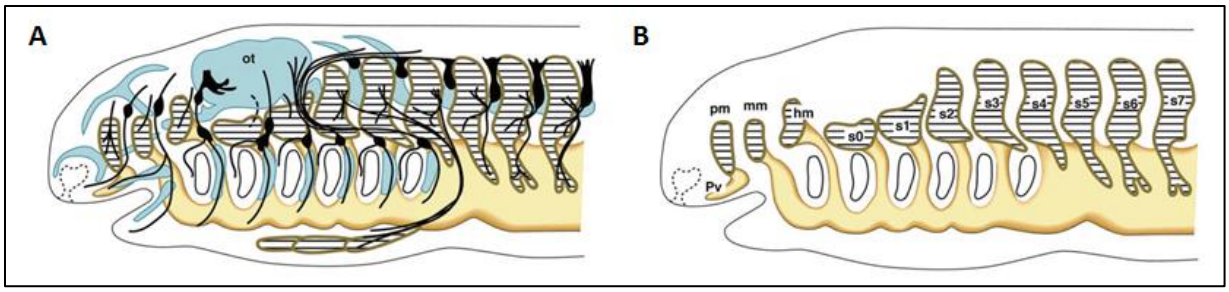


Figure 2. The head metamerism theory of Goodrich. (A) In the head region, each compartment contains a nerve, a head somite, and a branchial arch. **(B)** When the skeletal and peripheral nervous systems are removed from the scheme, it is clear that Goodrich based his model on the segmentation of the paraxial mesoderm. Abbreviations: pm, premandibular cavity; mm, mandibular cavity; hm, hyoid cavity; ot, otic vesicle; pv, Platt's vesicle; s0-7, somites. Both figures were modified from (Goodrich, 1918; Kuratani et al., 1999).

1.1.2 Non Segmentalist Hypothesis

In contrast with the segmental view of the head, the "non segmentalist" movement, which arose late in the 19th century, proposed that the head of vertebrates was formed by a different segmental process of that of the trunk. Thus, Froriep (Froriep, 1892, 1894) put in manifest his doubts about the homology between the preotic and postotic cavities and the most posterior trunk somites observed in lamprey embryos. Later, Kingsbury and Adelmann proposed that the components of the vertebrate head such as neuromeres (segments of the central nervous system), somitomeres (segments of paraxial mesoderm), and branchiomeres (segments of branchial arches), should not be integrated into single series of units (a single metamere), but as separated developmental processes (Kingsbury, 1920; Kingsbury, 1926; Kingsbury and Adelmann, 1926). In 1972, Romer proposed a dual segmental theory, claiming that the segmental process forming the body somites must be considered as an independent process of that of the gills (pharyngeal arches) (**Figure 3**). In the same work, Romer also proposed that during the evolution the "somatic or active" part of the embryo, (i.e. most part of the muscles, bones and central nervous system) have been trying to gain the control of the "visceral or passive" part of the embryo, (i.e. the digestive tract and its appendages) (Romer, 1972). In this way, Romer also claimed that the chordates evolved from a primitive sessile arm-feeder (passive animal) to a tunicate. Then the free-swimming larva of the tunicate underwent secondary evolutionary events giving rise to a primitive filter-feeding vertebrate (active animal) (Romer, 1972). Nowadays, tunicates are not considered in the discussion of the origin of the vertebrates' head since they have apparently lost muscular somites and their phylogenetic position has been revised (see below).

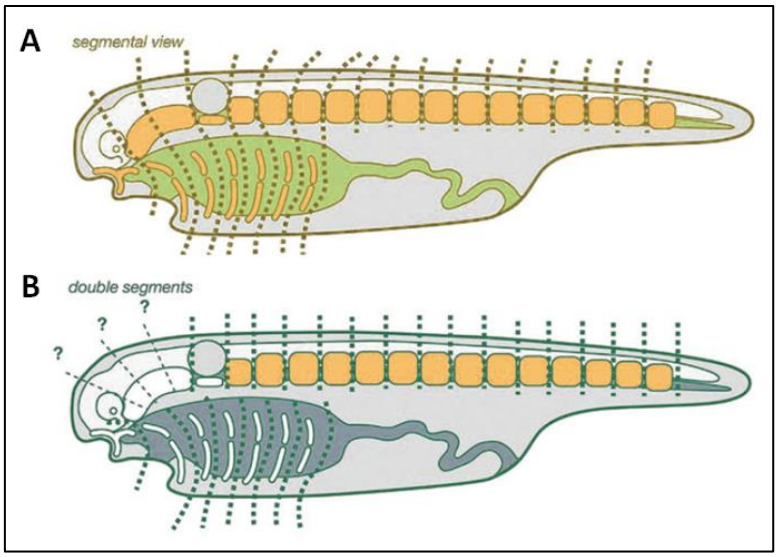


Figure 3. Segmental/Non segmental theories. (A) In the segmental hypothesis of Goodrich the vertebrate's head is assumed to contain only one type of segmentation that involves metamerism of paraxial mesodermal segments and pharyngeal arches. **(B)** Non segmentalist theory assumes independent patterns of metamerism for somites and pharyngeal arches questioning the presence of segments in the mesoderm of the head. Modified from (Kuratani, 2003).

One of the most controversial points between both hypotheses (segmentalist/non segmentalist) is the existence or not of the "cephalic somitomeres" (i.e. segments in the cephalic mesoderm). The "cephalic somitomeres" were described for the first time in 1980s as bulges in the cephalic mesoderm of chick embryos (Anderson and Meier, 1981; Jacobson, 1988; Jacobson and Meier, 1984; Meier, 1979; Meier and Packard, 1984; Meier and Tam, 1982). Nevertheless, since then, there has been no clear evidence, at a molecular or cell lineage levels, about the existence of these "cephalic somitomeres" (Freund et al., 1996; Jouve et al., 2002). Thus, from a morphological level, any clear metameric pattern is observed in the head mesoderm that suggests the presence of the "cephalic somitomeres". Conversely, a segmental pattern is observed in the trunk due to the metameric formation of the somites. Similarly, a segmental pattern is observed in the cranial nerves due to rhombomeres and the pharyngeal pouches. (Begbie et al., 1999; Begbie and Graham, 2001; Kuratani and Eichele, 1993). Altogether, this lead to Kuratani et al. to propose that the "cephalic somitomeres" rather than a real metameric process, could be a regionalization of the mesoderm into several domains induced by some other embryonic structures in the vicinities (Horigome et al., 1999; Kuratani et al., 1999). Nevertheless, new molecular data shows the expression of the segmental gene *c-hairy* in chicken embryos indicating an oscillatory pattern with two pulses entering into the cephalic mesoderm. The premandibular mesoderm (prechordal mesoderm) representing one wave and the rest of the head mesoderm representing the second wave (Jouve et al., 2002). Taken together, this suggests that the entire head mesoderm of vertebrates might be formed by two segments, in opposition to the hypothetical number of head mesoderm segments (seven or eight) assumed in the hypothetical vertebrate ancestor (Holland, 2000) (**Figure 4**).

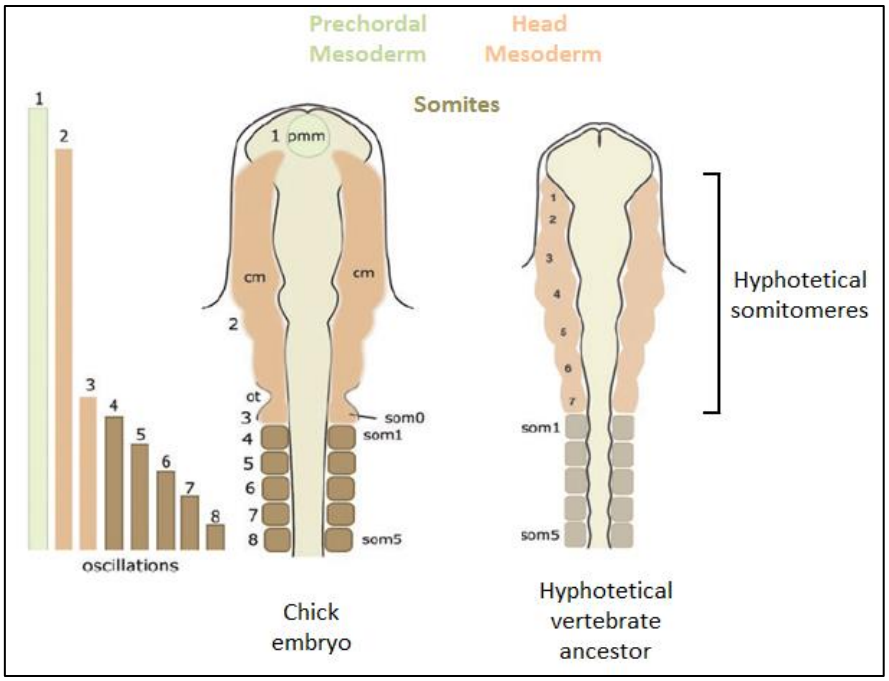


Figure 4. *c-hairy* expression in chick embryos and somitomeres. The morphological pattern of somitomeres in the chick embryo is shown on the right, as a simplified illustration. Hypothetical somitomeres are numbered. On the left is shown the oscillating expression of the gene *c-hairy 1*, in the early chick embryo, based on (Jouve et al., 2002). Each oscillation is numbered together with the mesodermal part generated after that oscillation. Note that there are only two oscillations in the head mesoderm, one for the premandibular mesoderm, and the other for the rest of the cephalic mesoderm. Abbreviation: (pmm) premandibular mesoderm or prechordal mesoderm, (cm) cephalic mesoderm, (som) somite. Modified from (Kuratani, 2005).

Another controversial point between these two hypotheses is the presence of head cavities described in lampreys (Damas, 1944; Koltzoff, 1902). Their observation suggested that head mesoderm was segmented in the ancestral vertebrate. However, thanks to the development of scanning electron microscopy techniques in the last decades, it was possible to understand the head mesoderm morphology of lampreys in a much better way. Thus, Kuratani's laboratory was able to show in lamprey embryos that (i) there are no overt head cavities in the premandibular, mandibular, or hyoid mesoderm and (ii) the developmental sequence of the head mesoderm in lamprey embryos is completely different from that of trunk somites (Kuratani et al., 1999). Cyclostomes, to which belong lampreys and hagfishes, are the most early divergent group of vertebrates. Therefore, the results of Kuratani's laboratory suggest that the head cavities (premandibular, mandibular and hyoid cavities) observed in sharks are probably derived and not ancestral features.

1.1.3 The "New Head" Hypothesis

In 1983, a new hypothesis for the origin of the vertebrates' head was postulated by Northcutt and Gans (Gans and Northcutt, 1983). Thus, thanks to the discovery and description of new cell populations as the neural crest and neurogenic placodes, they

realized that these cell populations together with an unsegmented head mesoderm (derived from the lateral plate mesoderm) are exclusive to vertebrates and that during the evolution these structures played a crucial role to switch from filter feeding to active predation. Thus, they proposed that the rostral head of vertebrates is a completely new structure and that the neural crest and neurogenic placodes evolved from the epidermal nerve plexus of ancestral deuterostomes (Gans and Northcutt, 1983; Northcutt and Gans, 1983). Recently, Northcutt has rejected their last claim, proposing that neural crest and neurogenic placodes evolved due to the rearrangement of germ layers in the blastulae of the ancestral deuterostome that gave rise to the chordates (Glenn Northcutt, 2005). To better understand the "new head" hypothesis, the neural crest cells and neurogenic placodes derivatives are examined in more details below.

1.1.3.1 Neural Crest Cells

Firstly described in 1868 by the Swiss Embryologist Wilhelm His, the neural crest cells (NCCs) are a migratory group of cells derived from the ectodermal tissue and de-epithelialized during embryogenesis. The NCCs were initially associated with the origins of neurons and ganglia, until the 1890s when Julia Platt demonstrated their role in the formation of the visceral cartilage of the head and in the teeth of the mud puppy *Necturus* (Hall, 2008; Platt, 1897; Trainor et al., 2003). However, this breakthrough was controversial and not completely accepted until 50 years later (Horstadius, 1950).

Nowadays, we have a more detailed view about NCCs and their derivatives. Molecularly, the NCCs are characterized by the expression of a set of transcription factor genes as *AP2*, *Snail1/2*, *FoxD3* and *SoxE* (Green et al., 2015). In addition, and thanks to the initial work leaded by Le Douarin using chicken/quail chimeras (Le Douarin, 1982), it was possible to follow the derivatives of the NCCs. Thus, it is known that NCCs generate the peripheral nervous system (PNS), establishing the connection between the central nervous system (CNS) and the periphery. The melanocytes of the body also derive from NCCs, as the mesenchymal cells that are able to differentiate into connective tissue, adipose tissue, bone, cartilage and into cells forming the wall of blood vessels (Dupin et al., 2006). Regarding the contribution of the NCCs to the formation of the cranium and facial skeleton structures, it was established that most of these craniofacial structures in vertebrates are derived from cephalic NCCs (Couly et al., 1993) and that only the posterior part of the neurocranium (the portion of the skull that covers the brain) has a mesodermal origin (**Figure 5**) (Couly et al., 1992, 1993). Additionally, while head muscles are derived mostly from cranial mesoderm (it will be treated later) their connective cells and attached tendons derive from cephalic NCCs (Grenier et al., 2009).

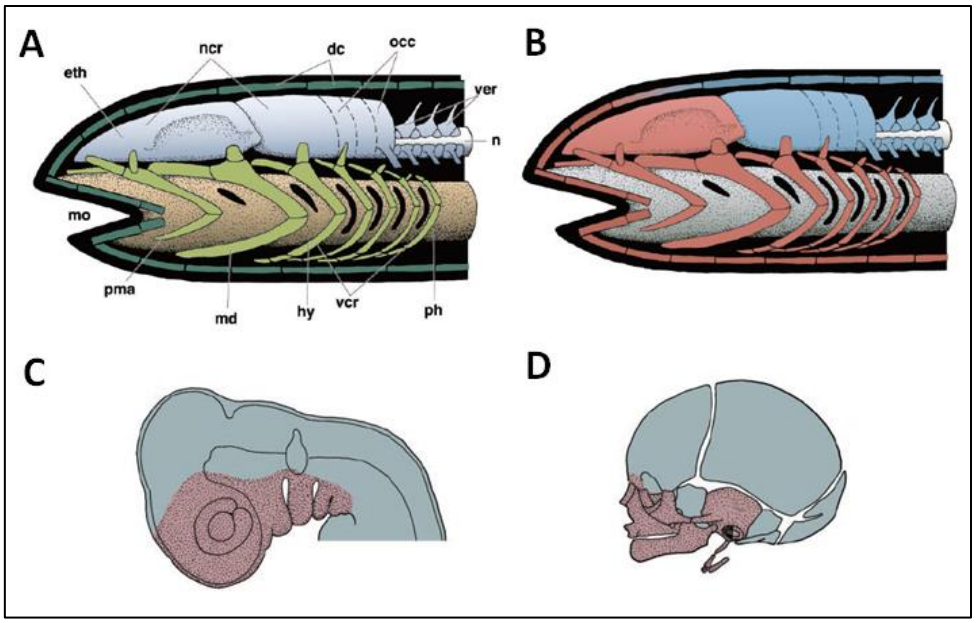


Figure 5. Neural crest cells and their skeletal derivatives in the vertebrate cranium. (A) According to the original scheme (Portman, 1969), the vertebrate (gnathostome) skull is assumed to be composed of the cartilaginous neurocranium (light blue), the viscerocranium (light green), and the dermatocranium (brown). **(B)** Neural-crest-derived elements have been colored in red, and the mesodermal elements in blue, based on several cell-labelling and molecular genetic experiments reported by (Couly et al., 1993; Le Lievre, 1978; Le Lievre and Le Douarin, 1975; Morriss-Kay, 2001; Noden, 1984). **(C)** Distribution of the cephalic mesoderm (blue) and crest-derived ectomesenchyme (red) in the chicken pharyngula by (Noden, 1988). **(D)** Results from chicken experiments were extrapolated to the human perinatal skull. Abbreviations: dc, dermatocranium; eth, ethmoidal region of the neurocranium; hy, hyoid arch; md, mandibular arch; mo, mouth; n, notochord; ncr, neurocranium; occ, occipital; ph, pharynx; pma, premandibular arch; vcr, viscerocranium; ver, vertebrae or vertebral column.

1.1.3.2 Neurogenic Placodes

In vertebrates the neurogenic placodes give rise to part of the cranial sensory apparatus (vision, taste, smell, balance and hearing). It has been established that all placodes arise from a common precursor territory, the preplacodal ectoderm (PPE) located around the anterior neural plate and neural crest, a region characterized by the expression of the transcription factor genes of the *Six1/2* and *Six4/5* families and of their coactivators from the *Eya* family (Schlosser, 2006). The different derivatives of placodes are as follows: (i) the adenohypophyseal placode, which gives rise to the anterior pituitary, (ii) the olfactory placodes that generate the chemosensory neurons of the olfactory epithelium and the vomeronasal organ, (iii) the lens placodes that will form the lens of the eyes, (iv) the profundal and trigeminal placodes, which generate somatosensory neurons sensing temperature, touch and pain in the head, (v) the lateral line placodes that generate mechanosensory hair cells to detect movement in the water (these placodes have been lost in amniotes), (vi) the otic placodes that generate mechanosensory hair cells to detect auditory stimuli and, (vii) the epibranchial placodes, which form viscerosensory neurons (**Figure 6**) (Schlosser, 2015).

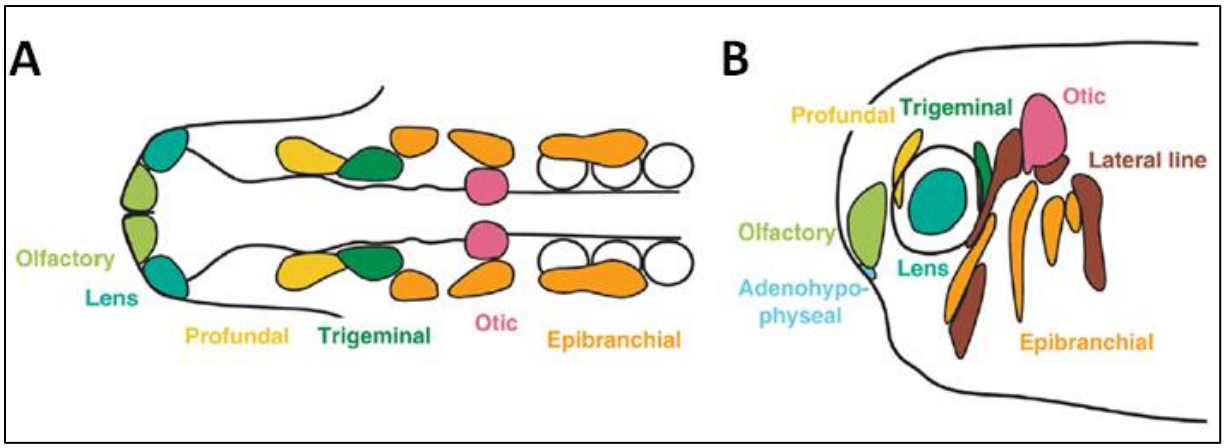


Figure 6. The cranial placodes of vertebrates. (A) Cranial placodes in a 10-13 somite stage chick embryo. In amniotes, profundal and trigeminal placodes are commonly referred to as ophthalamic and maxillomandibular placodes of the trigeminal nerve, respectively. Modified from (Streit, 2004). **(B)** Cranial placodes in a tailbud stage *Xenopus* embryo. Modified from (Schlosser and Northcutt, 2000).

1.1.3.3 The "New Head" hypothesis and the Brain

As Gans and Northcutt observed 30 years ago, the contribution of the NCCs and neurogenic placodes to the formation of the head is tremendous, indicating that these vertebrate novelties have clearly played a crucial role for the appearance of the head. Interestingly, some facial structures do not develop from these tissues. These structures are the head muscles (their origin will be treated later). Additionally, in their work, Gans and Northcutt argued the difficulty to incorporate into the classical segmentalist views a NCCs origin for the skull, since segmentalists claimed a metameric process for the entire vertebrate body and the NCCs contribute only to the anterior (head) and not to the posterior (trunk) skeletal parts of the vertebrate body. Thus they discussed that it was unlikely that the vertebrate's head originated through modifications of the same processes that pattern segments in the trunk, i.e. the metameric idea of Goodrich.

Another point that they discussed is the fact that there is no homologue of the brain in either the cephalochordates or the urochordates. Therefore, in vertebrates, the sister group of both, the forebrain and midbrain would represent a neomorphic structure (Gans and Northcutt, 1983; Glenn Northcutt, 2005). Nevertheless, more recent molecular and transmission electron microscopy (TEM) studies have revealed that indeed the cephalochordates present an homologue of the vertebrate midbrain (Schubert et al., 2006) and an homologue of the vertebrate diencephalon (posterior part of the forebrain) (Lacalli and Kelly, 2000; Lacalli, 2008). Thus, the only neomorphic structure in vertebrates is the telencephalon (the anterior part of the forebrain) (Holland, 2015). Moreover, an homologous genetic program for patterning the brain in vertebrates is present in the hemichordate *Saccoglossus kowalevskii* (that belongs to the ambulacraria, the sister group of chordates) and partly in the invertebrate chordates (i.e. cephalochordates and urochordates), suggesting that these animals have not completely retained the basal genetic program patterning the brain, and that indeed the genetic program to build the entire brain is more ancient that previously thought (Holland et al., 2013; Pani et al., 2012).

1.1.4 Final remarks

The vertebrates' head is one of the most complex structures of all animals. During the last centuries, different works have contributed to increase our knowledge about its development in different vertebrate species (mice, frog, lamprey, and chicken). Nevertheless, its evolutionary origin still remains unclear. Firstly, hypotheses on the origin of the vertebrates' head claimed that the entire body of vertebrates, including the head, is formed by the same segmental process. Contrary to the segmental vision of the head, the non segmentalist movement claimed that the formation of the head and the trunk are independent processes. In the early 1980s, Gans and Northcutt highlighted that neural crest cells, neurogenic placodes and unsegmented head mesoderm are unique innovations of vertebrates and are linked to the origin of the vertebrates' head, and established the "new head" hypothesis. Thanks to the development of better microscopy techniques, in the 1990s, Kuratani showed that the cyclostomes, the most basally divergent group of vertebrates, do not possess head cavities, indicating that i) the origin of head mesoderm was probably unsegmented and ii) head cavities in sharks are probably a derived feature in Chondrichthyes. Therefore, even if the question of the origin of the vertebrates' head seems to be blurry, the data presented so far allow us to conclude that (i) there is a widely acceptance that the ancestor of all chordates was completely segmented all along the body (ii) the vertebrates' head is formed by a completely different process from the truncal segmentation (iii) neural crest cells and neurogenic placodes form most of the facial structures of the vertebrates' head and (iv) that the brain do not represent a unique novelty of the vertebrates since the genetic program for building a brain was already present in hemichordates.

Importantly the anterior mesoderm, which give rise some head structures in vertebrates (mostly muscles, it will be treated later), is an unsegmented tissue. Thus the evolutionary question arising from this observation is how an unsegmented anterior mesoderm arose from a hypothetical ancestor of all chordates possessing all its body segmented? Three different scenarios emerge to answer this question:

(i) The addition of a completely new structure in the most anterior part of the body, a "new head".
(ii) The loss of anterior mesoderm segmentation
(iii) The loss of anterior mesoderm with a secondary acquisition of head mesoderm

The first scenario, the Gans and Northcutt scenario, supposes the appearance of a completely new structure, notably a complex structure including both the crane and the brain. This hypothesis seems unlikely because the cephalochordates (the most basally divergent group among chordates) possess homologous structures to the vertebrates' brain. Therefore, to distinguish between the last two scenarios, the use of an invertebrate chordate for comparative studies shedding light about the origin of the vertebrates' head seems an excellent choice. Below, I introduce the cephalochordate amphioxus as an animal model for these comparative studies and I explain why amphioxus is the best model to try to answer our question about the origin of the vertebrate's head.

1.2 AMPHIOXUS AS A MODEL

1.2.1 Identification and Description of Amphioxus

In 1774, the German zoologist Peter Simon Pallas described and classified a new mollusk that he called "*Limax lanceolaris*" (Limax = slug ; lanceolaris = lancet); it was the first description of amphioxus (**Figure 7**) (Pallas, 1774). Decades later, in 1834, Gabriel Costa, an Italian zoologist, recognized this "mollusk" as being closer to the vertebrates and renamed it as "*Branchiostoma lubricus*", because of its "mouth gills" (branchio = gills; stoma = mouth); this mouth gills were in fact the oral cirri of amphioxus. Thus, the name *Branchiostoma* remained as the Linnaean name of the genus (Costa, 1834). In 1836, William Yarell described in *Branchiostoma lubricus* a structure characteristic of all chordates, the notochord. This structure in amphioxus extends all along the body (from the most anterior to the most posterior part), this is why amphioxus was named as cephalochordate (kephalé = head; khordé = chord). At the same time the name amphioxus was first used to designate cephalochordates (from the Greek; amphioxus = pointed on both sides) (Yarrell, 1836).

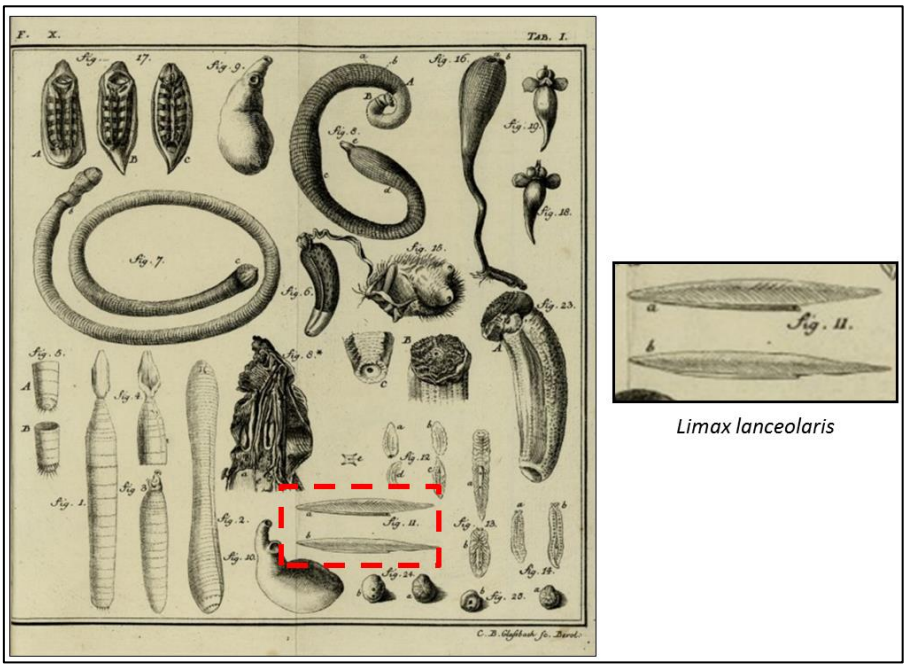


Figure 7. *Limax lanceolaris.* The first description of amphioxus was made by Peter Simon Pallas in 1774 classifying amphioxus as a "mollusk" (Pallas, 1774). A zoom of the drawing (dotted red rectangle) shows an enlargement of the amphioxus drawing.

The phylum Cephalochordata is composed by three genus; *Branchiostoma*, *Asymmetron* and *Epigonichthys* (Poss and Boschung, 1996). All of them possess characteristics shared with all chordates (urochordates and vertebrates). Thus the cephalochordates possess a dorsal hollowed neural tube, a dorsal notochord, pharyngeal gill slits, segmented muscles, ventral gut, mouth, and anus (**Figure 8**). Amphioxus also possesses some structures shared with vertebrates as the endostyle and preoral pit that are homologous to the thyroid gland and to the adenohypophysis, respectively. On the other hand, amphioxus lacks some essential vertebrate structures, as an internal skeleton, neural crest cells (NCCs), placodes and paired sense organs (Bertrand and Escriva, 2011).

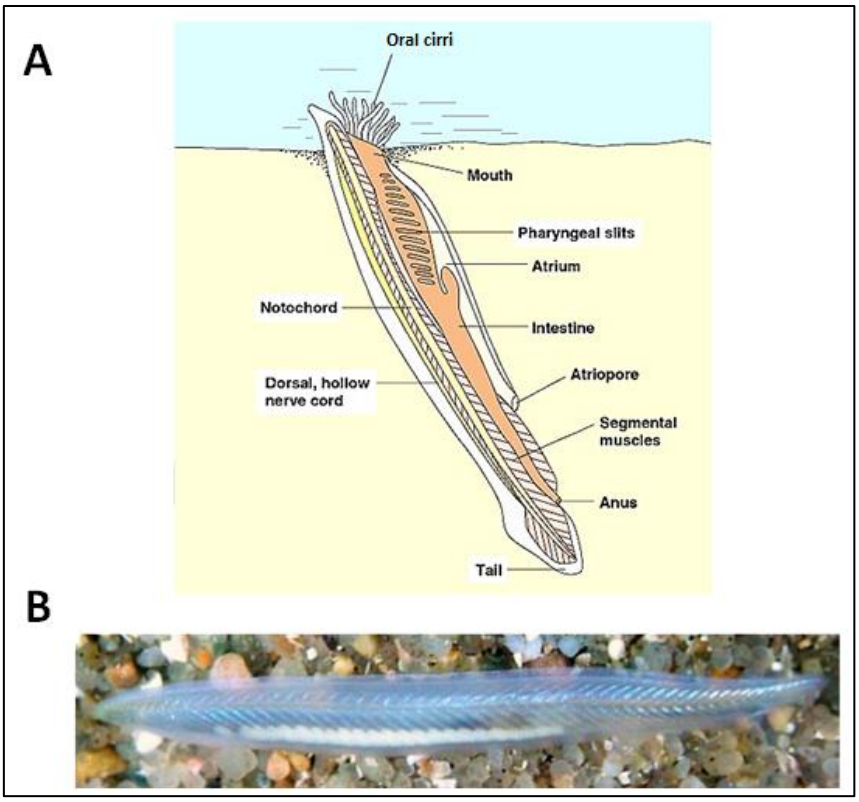


Figure 8. Amphioxus basic anatomy. (A) Schematic view of amphioxus basic anatomy. The most important morphological characteristics are depicted. These include the dorsal hollow neural tube, the dorsal notochord, the intestine and segmented muscles. The schema depicts an amphioxus burrowed in the sand. Anterior is to the top **(B)** *Branchiostoma lanceolatum* adult individual. It is possible to observe the segmented muscles and gonads (white). Anterior is to the left.

1.2.2 Phylogenetic Position and Amphioxus Species

Based on morphological characteristics and phylogenetic studies using the ribosomal ribonucleic acid coding gene sequences (rRNA) (Winchell et al., 2002), cephalochordates were considered as the sister group of vertebrates for a long-time. However, recent genome-scale studies have shown that urochordates and not cephalochordates are the closest phylum to vertebrates (Delsuc et al., 2006; Delsuc et al., 2008; Putnam et al., 2008). Additionally, new studies have identified in urochordates a type of migratory neural crest-like cells similar to the vertebrate neural crest cells, a typical characteristic of the vertebrates and absent in amphioxus (Jeffery et al., 2004). Thus, the new phylogenetic tree of life places the phylum Cephalochordata as the most basally divergent group among chordates (**Figure 9**).

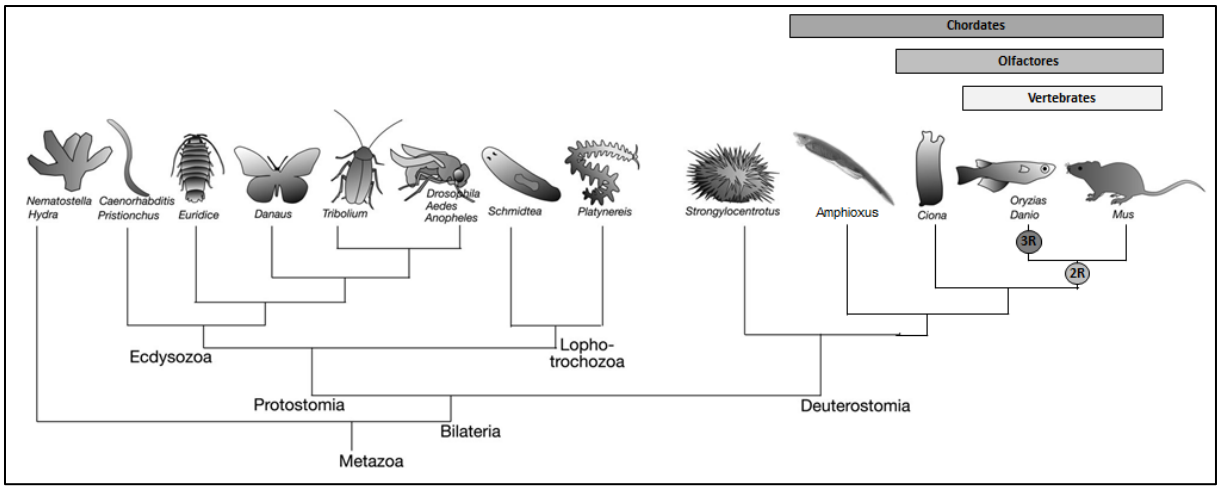


Figure 9. Selected animals from Bilateria showing its three main branches Ecdysozoa, Lophotrochozoa and Deuterostomia (Hydra as outgroup). New genomic studies allowed to place cephalochordates (amphioxus) at the base of the chordates whereas the urochordates (i.e. *Ciona*) get placed as the sister group of vertebrates (Delsuc et al., 2006). Two whole genome duplications occurred after the divergence of urochordates (2R). Another whole genome duplication took place during the evolution of the teleosts (3R). Times of phylogenetic divergence are not to scale, and the tree branches are intended only to depict general relationships. Modified from (Aguinaldo et al., 1997; de Rosa et al., 1999)

In their study based on morphological characteristics, Poss and Boschung identified and characterized at least 23 different species from the genus *Branchiostoma*, and 7 from the genus *Epigonichthys*, showing that amphioxus has colonized all the seas except for the Arctic and Antarctic oceans (there is no study showing the contrary). They described a worldwide repartition of amphioxus living in shallow waters in the Mediterranean or Caribbean sea, and Pacific, Atlantic or Indian oceans (**Figure 10**) (Poss and Boschung, 1996). Recent molecular analyses suggest that they might in fact be more species than described previously at least in the genus *Branchiostoma and Asymmetron*. Thereby, *Branchiostoma belcheri* has been subdivided into three different species, *Branchiostoma belcheri*, *Branchiostoma japonicum* and *Branchiostoma tsingtauense* (Zhang et al., 2006), and *Asymmetron lucayanum* also seems to be a multispecies clade (Kon et al 2006). Thus, new molecular analyses accompanied by new field collection seem to be needed to clarify the number of species and their phylogenetic relationships within the phylum Cephalochordata.

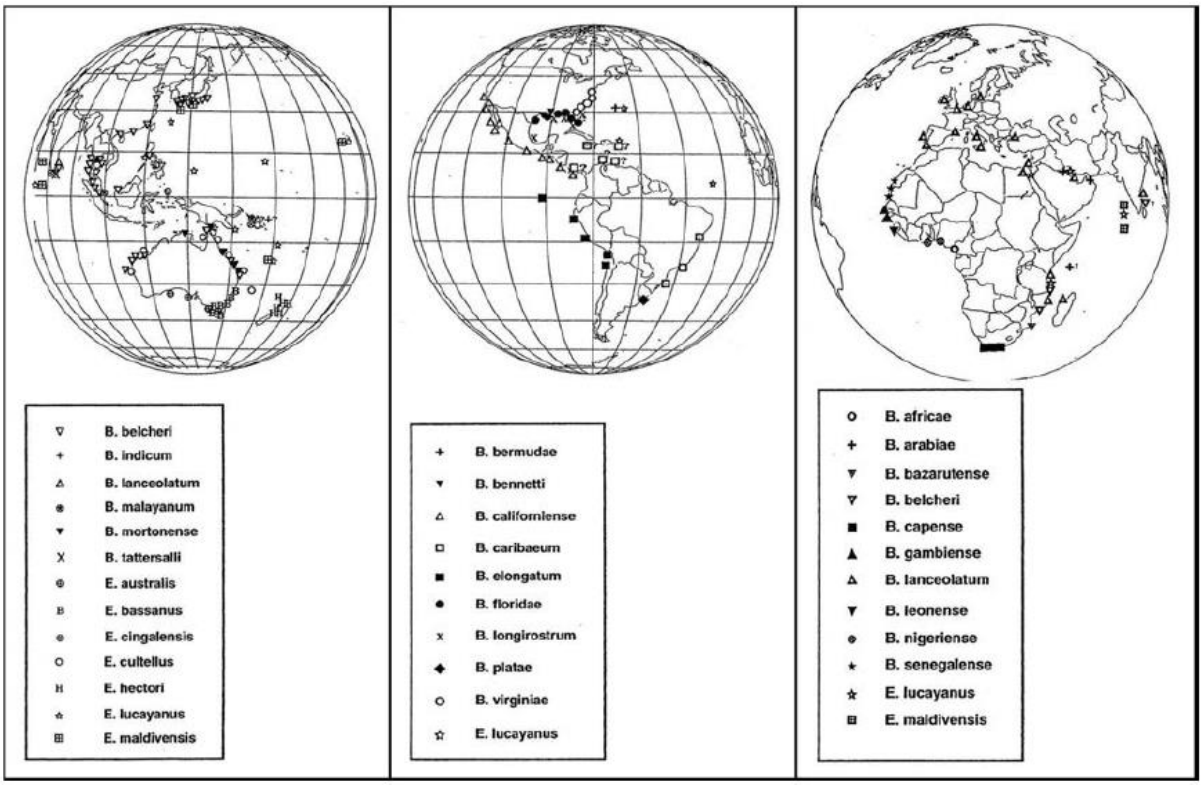


Figure 10. Amphioxus global distribution. Representation of the 23 species from the genus *Branchiostoma* and of the 7 species from the genus *Epigonichthys* described by Poss and Boschung. Modified from (Poss and Boschung, 1996).

1.2.3 Environment

Amphioxus is a filter-feeder animal that usually lives in shallow waters burrowed into the sand leaving only its mouth outside the sediment for filtering sea water. Amphioxus lives in tropical and temperate sea waters with a preference for coarse sand as sediment. However, some species like for example the Caribbean species *Branchiostoma floridae* can be found in thinner sand (Desdevises et al., 2011; Gosselck and Spittler, 1979; Webb and Hill, 1958). So far, only one species living in deep waters (229 meters depth) has been identified and called *Asymmetron inferum* (Kon et al., 2007).

1.2.4 Reproduction and Life Cycle

All the species in the phylum Cephalochordata are gonochoric and reproduce sexually by external fertilization. The spawning season of different amphioxus species usually corresponds to the spring-summer season and it spans during three to six months per year depending on the species. For instance, the spawning season for *B. belcheri* and *B. lanceolatum* takes place during two or three months, whereas for *B. floridae* it lasts almost five months. The increase of the temperature of the sea water during the spring-summer triggers the spawning in most of the amphioxus species. *A. lucayanum* is an exception since it is able to spawn during two different periods of the year, one during the summer and the

other one during the autumn (Holland and Holland, 2010). Normally, after the sunset, amphioxus swim up into the water column and release their gametes, coming back afterwards into the sand. Different intervals of spawning have been observed in each species. For instance, the species *B. belcheri* spawn during a short period of days, whereas *B. floridae* is able to spawn in a synchronic way every two weeks. Another case is represented by *A. lucayanum*, that is apparently influenced by the lunar cycle, thus most of the population tends to spawn the day after the full moon (Holland and Holland, 2010). After fertilization, the embryos develop and form a larva with a planktonic life style until they reach metamorphosis. The length of the planktonic period depends on each species. Thus in the case of *B. lanceolatum* this period takes 2-3 months and for *B. floridae* 2-3 weeks. At the end of metamorphosis, the juvenile become benthonic and goes into the sand where it continues growing until it reaches the adult stage (Bertrand and Escriva, 2011). Regarding the lifespan of amphioxus, again it depends on the species. Thus, it has been published that *B. floridae* can live between 2-3 years and *B. lanceolatum* between 5-8 years (Bertrand and Escriva, 2011; Futch and Dwinell, 1977).

1.2.4 Embryonic Development

The embryonic development of the genus *Branchiostoma* has been very well described and studied for more than 150 years (Cerfontaine, 1906; Conklin, 1932; Hatschek, 1893; Kowalevsky, 1867, 1876; Wilson, 1892, 1893). Regarding the genus *Epigonichthys*, there is no study describing its embryonic development. For the genus *Asymmetron*, there is only one study where the authors show that the embryonic development of *A. lucayanum* is similar to what is observed in the genus *Branchiostoma*, finding differences only at the beginning of the larva stage (Holland and Holland, 2010). During my research project, I used the species *B. lanceolatum* as an animal model, therefore the embryonic stages described in this work correspond to the embryonic development of this species at 19 °C (Bertrand and Escriva, 2011; Fuentes et al., 2007; Fuentes et al., 2004).

From fertilization until the gastrula stage, in particular during gastrulation, the development of amphioxus is similar to the one of invertebrate deuterostomes (i.e. a hollow blastula invaginates to form a gastrula in a similar manner to the sea urchin). However, the gastrula stage could be considered as a transition state towards a vertebrate-like development, since at the end of this stage the formation of characteristic structures of chordates as the notochord, the neural tube and the somites starts. The embryonic development of amphioxus is carried out as follows.

1.2.4.1 Fecundation to blastula stage

Once the spermatozoid has fertilized the oocyte, the chorion or the fertilization membrane raises and coats the zygote avoiding polyspermy and protecting the zygote from the exterior. The next phase, called cleavage, corresponds to the segmentation of the zygote. Thus, the first division occurs after 90 minutes, then after seven synchronic divisions the blastula stage is reached, this stage is characterized by the presence of a blastocoel, an internal cavity without communication with the exterior. During the blastula stage, it is

possible to identify the cells that will give rise to the mesendoderm and ectoderm, being the mesendodermal cells bigger than the ectodermal cells. This observation is based on morphological description or cell lineage labelling, but not on molecular identification (Holland and Onai, 2012).

1.2.4.2 Gastrulation

The beginning of gastrulation corresponds to a flattening of the blastula at the vegetal pole where the cells will become mesendodermal cells. Then, a movement of invagination of the vegetal pole into the blastocoel is observed, until the vegetal pole touches the animal pole (**Figure 11**). In amphioxus, during gastrulation, a second movement is observed, that is the involution of some cells at the level of the blastoporal lip. At the end of gastrulation, two germ layers are formed: (i) one internal layer called mesendoderm that will give rise to the endoderm in the ventral part, and to the mesoderm in the dorsal part; (ii) and one external layer, the ectoderm, that will give rise two different tissues, the epidermis in the anterior and ventral part, and the neural plate (neuroectoderm) in the dorsal region of the gastrula (**Figure 11**).

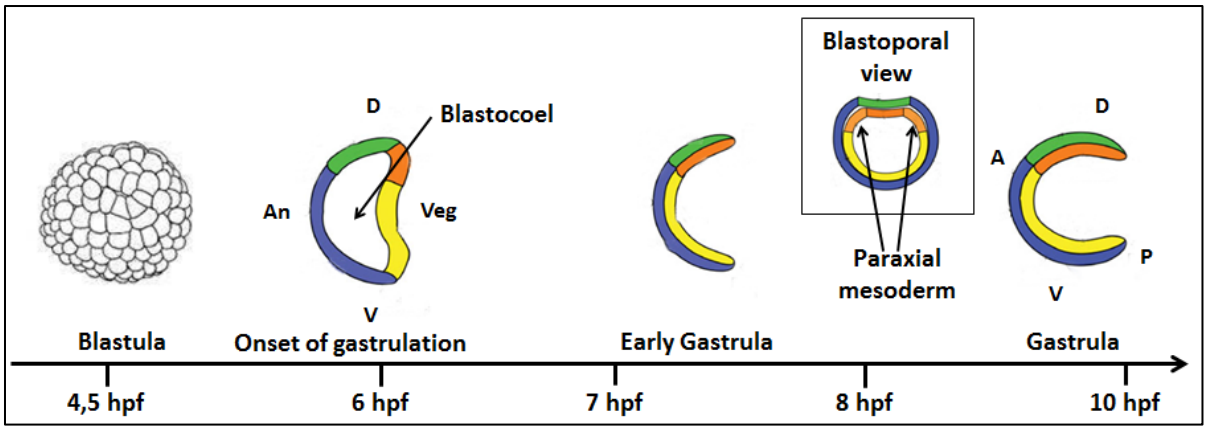


Figure 11. Two germ layers are formed during gastrulation in amphioxus. In this schematic representation the fate map of the different tissues and their movements during gastrulation are depicted. The presumptive ectoderm (blue), endoderm (yellow), neural plate (green), notochord (orange) and the paraxial mesoderm that will give rise to the somites (light orange) are depicted during gastrulation. Lateral views for all except for blastula stage and blastoporal view of gastrula. Abbreviations: (An) animal pole, (Veg) vegetal pole, (D) dorsal part, (V) ventral part, (A) anterior part, (P) posterior part of the embryo. Modified from (Holland and Onai, 2012).

1.2.4.3 Neurulation to metamorphosis

The neurula stage starts by the flattening of the dorsal part of the gastrula forming the neural plate. Then major events are observed. At the level of the ectoderm, the epidermal part completely detaches from the neural plate and cells fuse in the dorsal midline. On the other hand, the edges of the neural plate begin to fold until they fuse to form the neural tube. Regarding the dorsal mesoderm, it is possible to differentiate three regions, the axial

mesoderm (central position), that will give rise to the notochord, and the paraxial mesoderm (both sides of the axial mesoderm), that will give rise to the anterior somites. At the same time, the blastopore begins to close posteriorly. Then, the embryo elongates through the addition of new structures in the posterior part produced by the tailbud, which derives from the blastoporal lips, until the larval stage (**Figure 12**).

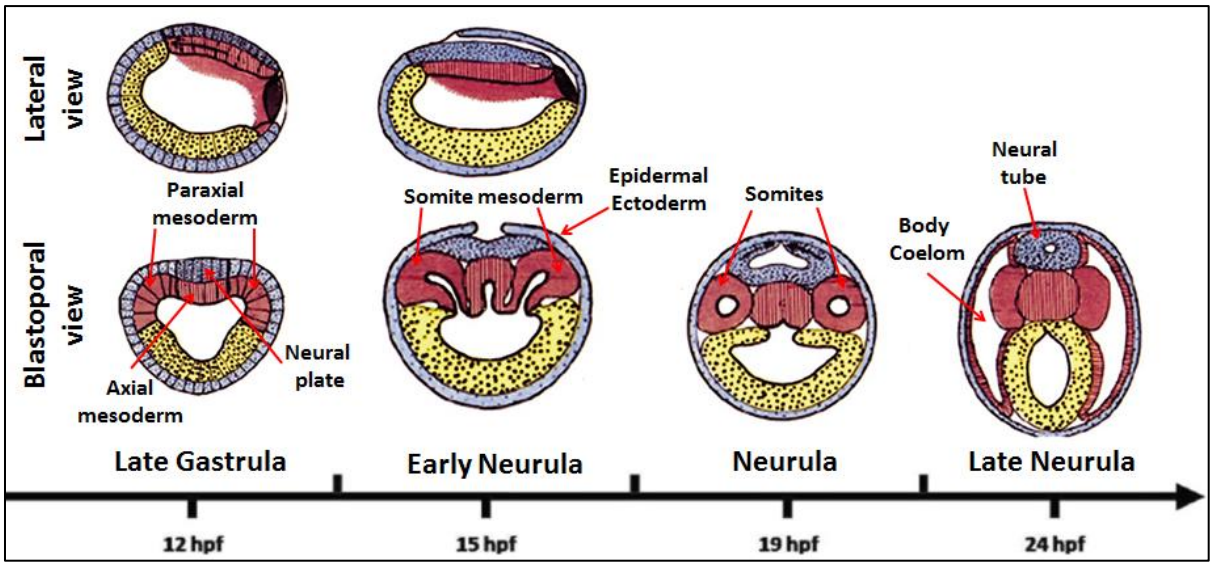


Figure 12. Neurulation in amphioxus. In this schematic representation it is possible to observe the events occurring during neurulation in amphioxus. At late gastrula stage the mesoderm (red) extends along the antero-posterior axis, and the paraxial mesoderm (red) and neural plate (dark blue) are differentiated from the axial mesoderm and non-neural ectoderm, respectively. At early neurula stage the mesoderm separates by constriction laterally to form somites and middorsally to form the notochord. The epidermal ectoderm spreads over the neural plate. At neurula and late neurula stages the somites are completely formed and spreads laterally and ventrally to form the body coelom. The neural tube is formed through the dorsal folding of the lateral edges of the neural plate until they fuse at the midline. Modified from (Langeland et al., 1998).

At the end of the neurula stage (subdivided into; early neurula (N1), mid neurula (N2), and late neurula (N3) stages) the first signs of pharyngeal enlargement are observed (Hirakow and Kajita, 1994). From this moment the embryo enters into the so-called premouth stage, which is characterized by the development of structures such as the pharynx, mouth and digestive tube. Once these structures are developed the larval stage begins. The planktonic larvae remain growing and adding new somites at the posterior part and gill slits in the pharyngeal region. The number of gill slits that are formed before metamorphosis is different for each amphioxus species, but in the case of *B. lanceolatum* metamorphosis occurs after the formation of 13-15 gill slits which takes between 2-3 months. During the larval stage, the gill slits are formed asymmetrycally as well as other body structures. Thus the mouth is placed on the left side of the pharynx and the gill slits on the right ventro-lateral side. Moreover, the left-side somites are positioned more rostrally (half-somite) than the right-side somites. Finally, during the metamorphosis, the mouth gets positioned rostrally. The gill slits are duplicated into two rows in a first time, and then one of these rows migrates to the left side. Moreover, the metapleural folds develop and cover the pharynx forming the atrium. The intestine gets also regionalized and the hepatic caecum is formed (**Figure 13**).

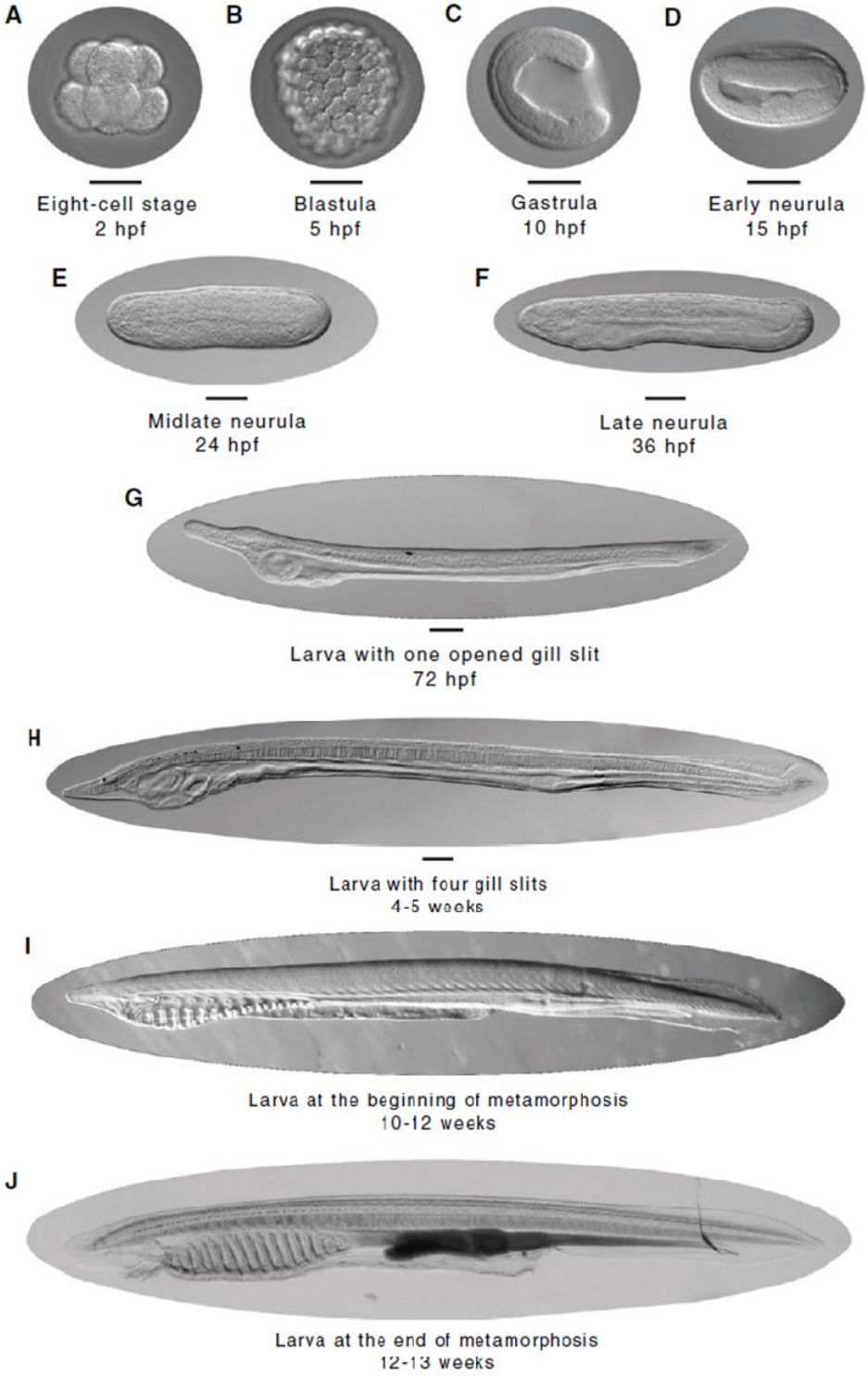


Figure 13. Embryonic development and metamorphosis of amphioxus. Photographs of the most representative embryonic and larval stages of amphioxus are presented (*B. lanceolatum*). The embryos were grown at 19°C. (A-C) Early stages of development including gastrula stage. (D-F) Neurulation stages. (G-J) Larval stages. Scale bar 50 µm. Modified from (Bertrand and Escriva, 2011).

1.2.5 Amphioxus as an Animal Model in EVO-DEVO

Amphioxus has always been considered a fascinating animal model for answering evolutionary questions. Thus, with the recent phylogenetic data, amphioxus is now occupying an interesting phylogenetic position as the most basally divergent group among chordates (Delsuc et al., 2006). Moreover, fossil records dating from 520 million years ago (*Pikaia gracilens* from the middle Cambrian found in Burges Shale, and fossil records from the lower Cambrian, in particular *Yunnanozoon*, *Haikouichthys* and *Myllokunmingia* found in Chengjiang, China) show morphological characteristics (**Figure 14**) that could be considered similar to those harbored by amphioxus (**Figure 8**). Even if the phylogenetic position of some fossil records is still controversial, it is likely that the ancestor of all chordates possessed an amphioxus-like body-plan, making amphioxus the only living animal with a high resemblance to the hypothetical chordate ancestor.

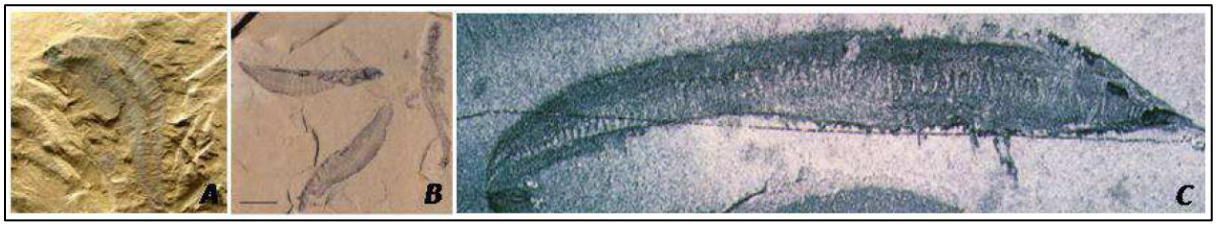


Figure 14. Fossil records from the Cambrian. (A) *Yunnanozoon livium*, found in 1984 in Chengjiang in China. **(B)** *Haikouella lanceolata*, collection from professor Jun Yuan Chen. Scale bar: 1 cm, modified from (Bertrand et al., 2007). **(C)** *Pikaia gracilens* found in Burges Shale in Canada, modified from (Long, 1995).

During the last decades the advances in whole genome analyses showed that amphioxus possess a "simple" genome (Putnam et al., 2008). Thus, as it was proposed by Ohno (Ohno, 1970), and confirmed later by Dehal et al. (Dehal and Boore, 2005) two rounds of whole genome duplications occurred during the evolutionary history of vertebrates (three in teleosts) (**Figure 9**) (Jaillon et al., 2004; Meyer and Schartl, 1999; Taylor et al., 2003)). What still remains controversial is the precise timing of these duplications. In any case, the duplications happened after the divergence of cephalochordates and urochordates (Putnam et al., 2008) and before the divergence of chondrichtyans (Robinson-Rechavi et al., 2004). Unfortunately, the complete genome sequence of a key animal to understand genome duplications in vertebrates, the lamprey, was not able to completely demonstrate the exact timing of the two rounds of genome duplications, even if it was suggested that both occurred before the cyclostomes divergence (Smith et al., 2013). An example of the result of the whole genome duplications is represented by the *Hox* clusters in vertebrates. Thus, there is a unique *Hox* cluster in amphioxus (Garcia-Fernandez and Holland, 1994) compared to four clusters in mammals, and for each specific *Hox* gene in amphioxus there are between one and four *Hox* orthologue genes in most vertebrates (**Figure 15**). Nevertheless, as any other animal, amphioxus has its own evolutionary history and possesses 15 *Hox* genes instead of the 14 assumed to have been present in the chordate ancestor, suggesting that a specific duplication of one *Hox* gene occurred in amphioxus (Holland et al., 2008a). The fact

that amphioxus possess a "simple" genome represents an advantage to understand the evolution of the function of different signalling pathways. For instance in amphioxus there is only one FGF receptor and 8 ligands compared with the four FGF receptors and 22 ligands in vertebrates (Oulion et al., 2012b). The simple inhibition of the FGF receptor in amphioxus can show us its direct role during the embryonic development of amphioxus (Bertrand et al., 2011). On the contrary, in vertebrates, the multiple FGF receptors and ligands makes more complicated the interpretation of the results of their inhibition during development.

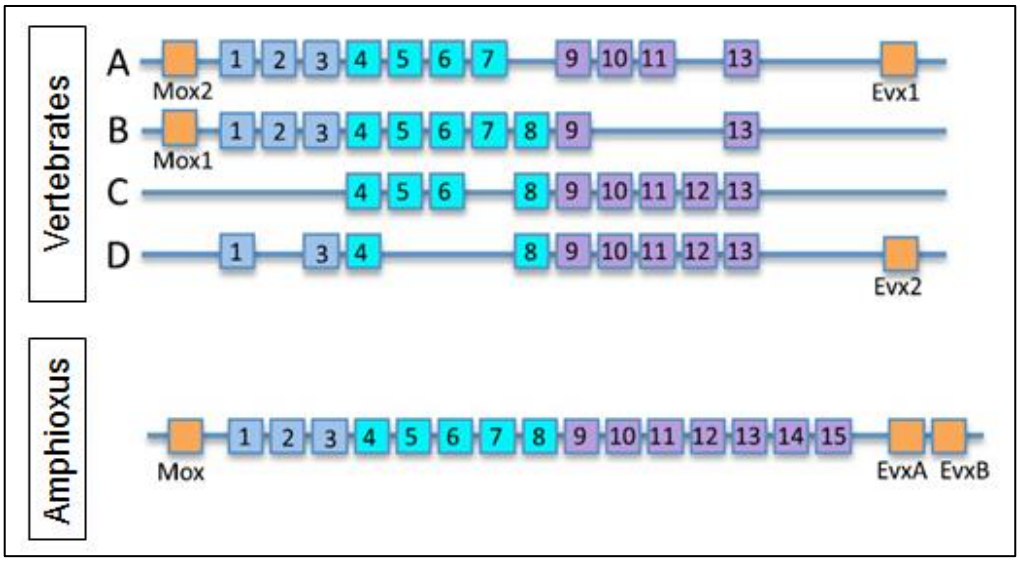


Figure 15. *Hox* clusters in mammals and amphioxus. Due to the two rounds of whole genome duplications proposed by (Ohno, 1970) (2R), there are four *Hox* clusters (HoxA to HoxD) in vertebrates. In amphioxus there is only one *Hox* cluster since it diverged before the 2R. The genes *Mox* and *Evx* flanking the *Hox* cluster in amphioxus and vertebrates confirmed the synteny. Also it is possible to observe the losses of certain *Hox* genes in vertebrates. Modified from (David and Mooi, 2014).

Even if several species form the phylum Cephalochordata, only three or four of them are used for Evo-Devo studies. The most used species are the Mediterranean species *B. lanceolatum*, the Caribbean specie *B. floridae* and the Asian species *B. belcheri.* Importantly, for these species the genome and transcriptome are publicly available (Huang et al., 2014; Mou et al., 2002; Oulion et al., 2012a; Putnam et al., 2008). Additionally, a few studies have also been undertaken after the correct classification of the different Asian species, so today we can find literature for *Branchiostoma japonicum* and *Branchiostoma tsingtauense* and more interestingly in a species from a different genus, *Asymmetron lucayanum* (Holland et al., 2015). Regarding the possibility to use amphioxus as a model for evo-devo studies, a lot has been done during the last 30 years. Thus, during the summer time (from May to August), for the Mediterranean species *B. lanceolatum*, it has been shown that an increase of 3-4 °C of the water temperature during 36 hours can trigger spawning of the animals in captivity, making it possible to obtain embryos every night (Fuentes et al., 2007; Fuentes et al., 2004). On the contrary, for the species *B. belcheri* and *B. floridae*, the induction of spawning has

been less studied, and it is only possible to obtain embryos during the natural field spawning nights.

Besides *in situ* hybridization as the classical experimental technique, amphioxus offers the possibility to interfere with several signaling pathways by using pharmacological treatments directly added to the seawater in which the embryos develop (Bertrand and Escriva, 2011). Performing immunohistochemistry staining is also possible in amphioxus using specific antibodies or heterologous antibodies used in vertebrates designed against conserved epitopes (**Figure 16**). Finally, concerning our ability to modify gene function, although microinjections in eggs of amphioxus were established 10 years ago, a lot of improvements are still needed. Indeed, mRNA injection allows overexpression of a given gene in all the amphioxus species, but knock-down is only efficiently working using morpholino antisense oligonucleotides injection in the Caribbean species *B. floridae* (Holland and Onai, 2011). In addition, recently for the Asian specie *B. belcheri*, a new method to induce direct deletions, mutations or insertions in the genome have been reported, thus the transcription activator-like effector nuclease (TALEN) method seems to be effective in amphioxus as in vertebrates (zebrafish, frog, rat, mouse) (Li et al., 2014). Finally, microinjections of plasmids can also be used in several species to obtain transient mosaic transgenic embryos.

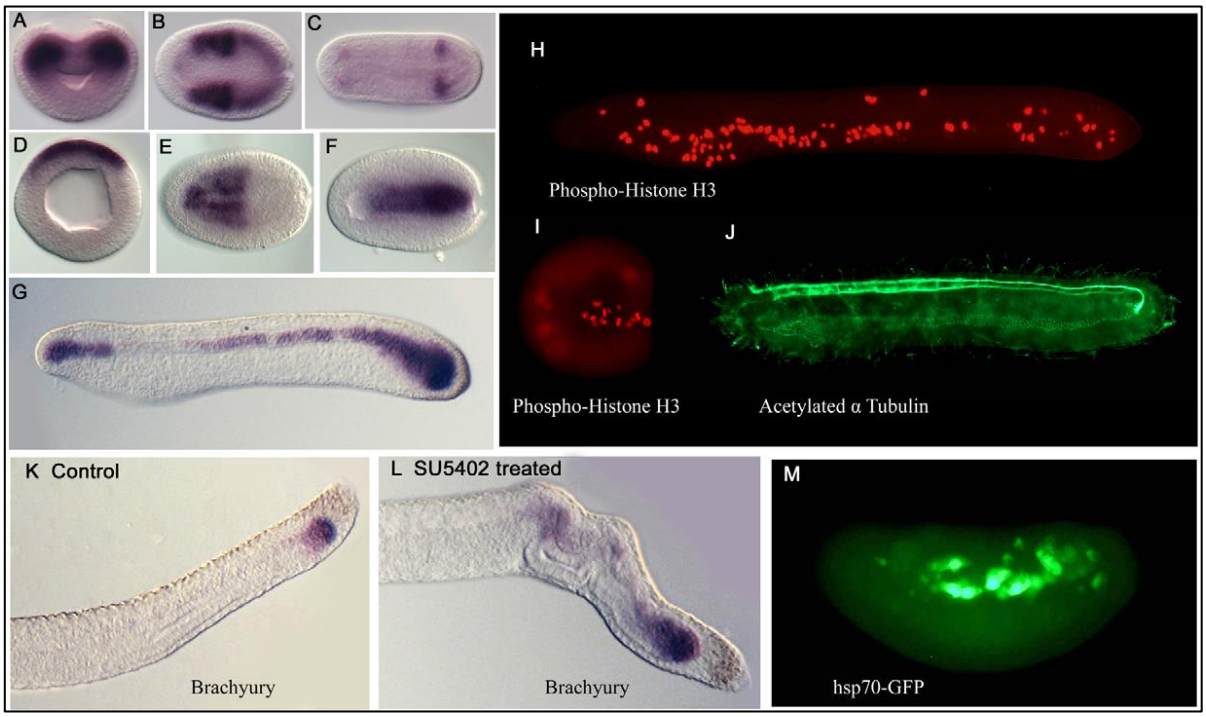


Figure 16. Experimental approaches developed during the last 30 years in amphioxus. (A-G) *In situ* hybridization showing the expression pattern of key developmental genes as *Delta* (A-C), *Neurogenin* (D-E), *Netrin* (F) and *Brachuyry* (G, K, L) at different stages as gastrula (A-D), neurula (B, C, E, F), and late neurula (G). **(H-J)** Immunohistochemistry labelling using antibodies against phosphorylated histone H3 (H, I) and acetylated tubulin (J) in embryos at gastrula (I) and late neurula (H, J) stages. **(K-L)** Pharmacological treatment using the inhibitor of the fibroblast growth factors (FGFs) signaling pathway SU5402. In control larva the expression of *Brachyury* is restricted to the tailbud (K), whereas

in treated larva the expression is observed throughout the entire notochord which elongated during the treatment period (L). **(M)** Transient transgenic amphioxus obtained by microinjection of the reporter plasmid p339_hsp70-GFP. Figure extracted from (Bertrand and Escriva, 2011).

1.2.6 Final remarks

Firstly described more than 200 years ago, amphioxus called the attention of researchers because of its morphologically simple characteristics resembling a vertebrate, but at the same time lacking some essential features as limbs, internal skeleton, neural crest cells or neurogenic placodes among others. In the last decades, and supported by phylogenetic data, amphioxus was positioned as the most basally divergent group among chordates. In addition, even if the interpretations based on fossil records are always controversial, it is extensively accepted that the ancestor of all chordates possessed an amphioxus-like morphology with a body completely segmented. Due to the effort of several laboratories around the world, amphioxus has emerged as a new animal model for the study of the invertebrate-chordate to vertebrate transition. Importantly, in the European species *B. lanceolatum*, the spawning can be controlled by a temperature shock allowing us to get embryos every night during the natural spawning season of this species. In addition, several molecular tools have been implemented (microinjection of unfertilized eggs, pharmacological treatments, *in situ* hybridization or high throughput analyses among others). Undoubtedly, a lot has to be done in the future (CRISPR/Cas9 technology, the possibility to get embryos all year long or to complete the life cycle in captivity). Nonetheless the tools available today allow us to use amphioxus as a good approach to understand how novelties arose in vertebrates.

The body of amphioxus shows a clear segmentation given by the somites (structures derived from the paraxial mesoderm) along the antero-posterior axis. Moreover, the most anterior part of amphioxus lacks all the structures that define a vertebrate head, such as sensory paired organs (eyes, noise or ears), a complex brain, skeletal elements or particularly, an unsegmented mesoderm. Indeed, it is believed that the loss of segmented somites in the anterior part of the vertebrate embryo during evolution allowed to release the developmental constrains imposed by the somites and the formation of new structures. Thus, functional elucidation of the development of anterior somitogenesis in amphioxus may shed light on differences with vertebrates that explain the evolution of the head. Are all the somites of amphioxus formed by the same segmental process? Are there any differences between the anterior and posterior somites? Is the genetic program for the formation of somites similar between vertebrates and cephalochordates? These are the questions that will be treated in the next chapters.

1.3 SOMITOGENESIS

The most conspicuous part of the amphioxus body are the muscles. In amphioxus, muscles extend from the most anterior to the most posterior part of the body and give to amphioxus the ability to generate undulatory movements to escape from predators or to change their position into the sand. The muscles in amphioxus derive exclusively from the somites, therefore in the following pages I will show the morphological, developmental and genetic programs known to be involved in somitogenesis and myogenesis in amphioxus and vertebrates. Importantly, since facial muscles in vertebrates do not develop from somites, I will also treat in a separate chapter the formation of facial muscle structures in vertebrates. Finally, I will treat the differences between anterior and posterior somitogenesis in amphioxus and I will discuss all these data in the context of the origin of the vertebrates' head.

1.3.1 Somitogenesis in vertebrates and amphioxus: a morphological description

In vertebrates, somites develop from the mesenchyme present in the tailbud region called presomitic mesoderm (PSM), in the most posterior region of the embryo. Somites form as spherical epithelial structures, derived from this presomitic mesoderm (PSM) *via* a mesenchymal-to-epithelial transition (Hubaud and Pourquie, 2014; Yabe and Takada, 2016). Once the somites are formed, they are compartmentalized. Thus, the ventral portion of the somite is de-epithelialized to form the mesenchymal sclerotome, the dorsal portion called dermomyotome remains as an epithelial sheet and the myotome arise later between the sclerotome and the dermomyotome by delamination of its edges (**Figure 17a-b**). Each of these compartments will give rise to different tissues. Hence, the sclerotome will give rise to the axial skeleton, the dermomyotome to the dorsal dermis and skeletal muscles and the myotome to the skeletal muscle precursors (Brent and Tabin, 2002).

In amphioxus, somites can be divided morphologically in two classes: the most anterior 8-10 somite pairs, pinch off by enterocoely from dorsolateral grooves of the archenteron at early neurula (N1) stage as a single mesodermal layer. And then, at the beginning of late neurula stage (N3), the new somites are formed by schizocoely one at a time, by budding off from the epithelium that surround the neurenteric canal, allowing the elongation of the body in the posterior part (**Figure 18**) (Beaster-Jones et al., 2008; Mansfield et al., 2015; Schubert et al., 2001b). Thus, the posterior somites of amphioxus unlike those in vertebrates that are generated by budding off from the PSM, are derived directly from an epithelium (i.e. amphioxus lacks a PSM region in its posterior part of the body), and do not undergo a mesenchymal-to-epithelial transition. Then, amphioxus somites are divided into myotome (medial) and non-myotome (lateral) compartments. The myotome will give rise to the myomeres that will constitute the body musculature. The non-myotome compartment is divided in dermomyotome (external cell layer), presumptive lateral plate and sclerotome. The dermomyotome will give rise to the dermis, connective tissues, and the fin box (a segmented structure formed dorsal to the neural tube). The presumptive lateral plate will

give rise to the perivisceral coelom. Thus, while in vertebrates the sclerotome gives rise to structures as cartilages or bones, amphioxus do not possess such structures, instead the sclerotome in amphioxus will give rise to the mesothelium that encloses the sclerocoel separating the myotomes from other structures as notochord and neural tube (**Figure 17c**) (Mansfield et al., 2015; Scaal and Wiegreffe, 2006).

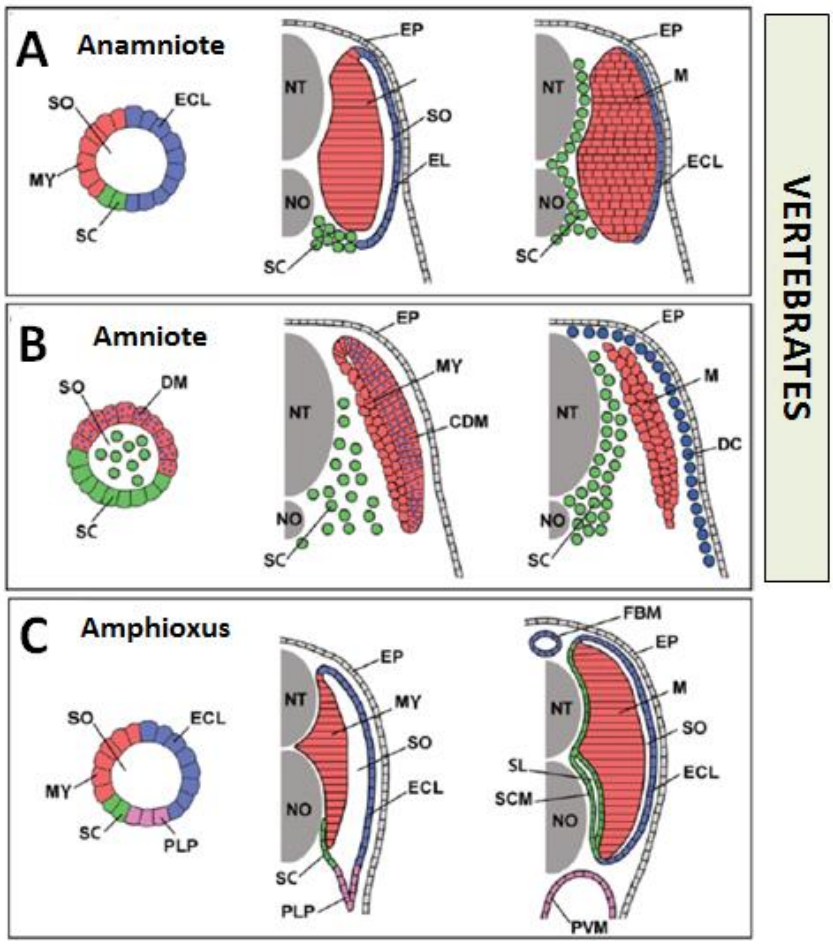


Figure 17. Somites development in amphioxus and vertebrates. A, B and C shows early (left panels), mid (middle panels) and late (right panels) stages for development and organization of the somites in anamniote vertebrates (e.g. fishes, amphibians), amniote vertebrates (mammals) and amphioxus. The schema shows somites unbent from their true chevron- or W-shape and for simplicity the ribs and ventral muscles were omitted (and in anamiotes, the myoseptal cells), which are derived from somites and migrate ventrally into the lateral plate mesoderm. Abbreviations: (CDM) central dermomyotome, (DC) dermal cells, (DM) dermomyotome, (ECL) external cell layer, (EP) epidermis, (FBM) fin box mesothelium, (MY) myotome, (NT) neural tube, (NO) notochord, (PLP) presumptive lateral plate, (PVM) perivisceral mesothelium, (SL) sclerocoel, (SCM) scleromesothelium, (SO) somitocoel, (SC) sclerotome, (M) trunk muscle. Modified from (Mansfield et al., 2015)

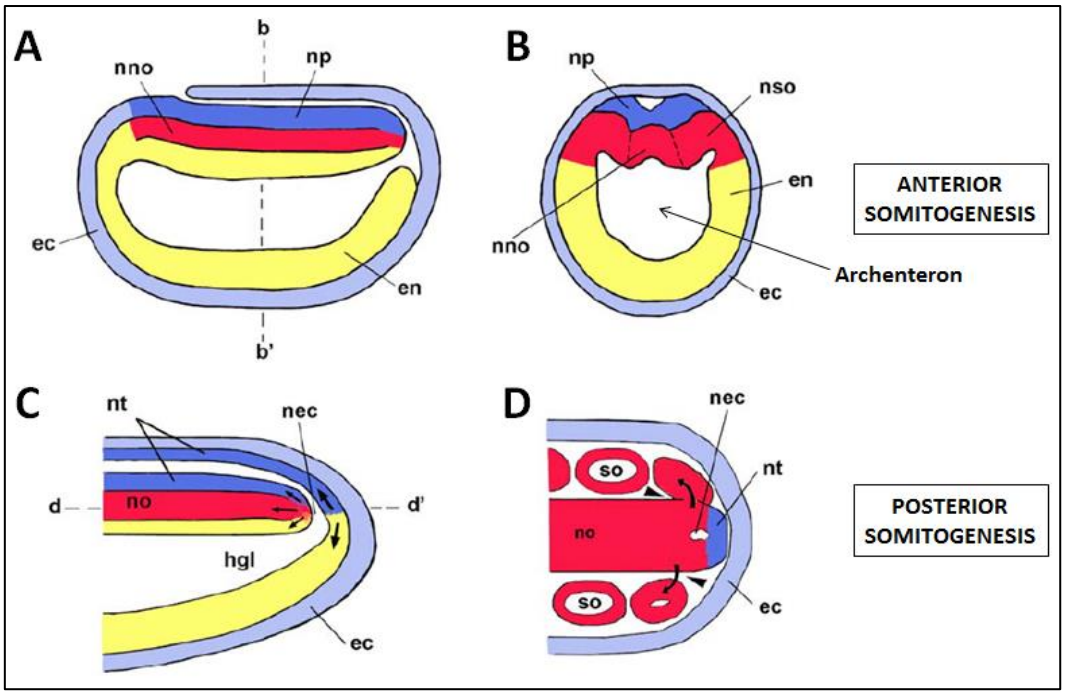


Figure 18. Somitogenesis in amphioxus. Schema of early neurula stage (N1) embryo in lateral view (A) and cross section at the level of b-b' (B). Posterior portion of 10-12 somites stage embryo in lateral view (C) and cross section at the level of d-d' (D), anterior is to the left in (A, C and D). The first somites pinch off from the grooves of the dorsolateral wall of the archenteron (A, B). The posterior somites pinch off from the neurenteric canal one at a time, first from one side and then from the other as indicated by arrows in the scheme. Abbreviations: (np) neural plate, (en) endoderm, (nno) nascent notochord, (nso) nascent somites, (ec) ectoderm, (nec) neurenteric canal, (hgl) lumen of the hindgut, (nt) notochord, (no) neural tube. Modified from (Beaster-Jones et al., 2008).

1.3.2 Molecular Control of Somitogenesis

1.3.2.1 The clock and wavefront model

If the precise anatomical structures of vertebrate's tailbuds vary from one species to the other, the global mechanism of posterior elongation is similar, and the molecular control of this process seems to be conserved. Indeed, it has been shown that FGF, Wnt and retinoic acid (RA) signaling pathways are interacting in all vertebrates to allow a harmonious elongation of the posterior body structures even if the fine interactions between these pathways are still not completely elucidated (Aulehla and Pourquie, 2010; Hubaud and Pourquie, 2014; Krol et al., 2011).

Different studies on the elongation process have shown that the function of FGF signaling is both to maintain a proliferative zone formed by pluripotent cells in the most posterior part of the embryo, thereby allowing growth of the axis, and to control the timing of cell differentiation. Indeed, FGFs ligands have been shown to inhibit cell differentiation in the caudal part of the embryo (Boulet and Capecchi, 2012; Dubrulle and Pourquie, 2004). Thus, the attenuation of the FGF signal is required for mesodermal differentiation as well as

for positioning the segment boundaries of the nascent somites in the mesoderm (Akiyama et al., 2014). But absence of the FGF signal is not sufficient to induce cell differentiation, and another signal is needed. This signal is RA, coming from the anterior already formed somites (Maden et al., 2000). Thus, RA attenuates FGF signaling in the paraxial mesoderm, where it also controls somite boundary position. In fact it has been proposed that FGF and RA form opposing signals that downregulate each other (Diez del Corral et al., 2003; Kumar and Duester, 2014). Many data also demonstrate the implication of Wnt signaling in the posterior elongation of vertebrate embryos. One of the roles of Wnts is to balance the opposing FGF/RA signals (Olivera-Martinez and Storey, 2007). Indeed, some Wnts expressed in the caudal part of the embryo can activate FGF signaling. Wnts also act in concert with FGFs to maintain the cells of the proliferative zone in an undifferentiated state, whereas more anteriorly, where FGF signaling is absent, Wnts can activate RA signaling by increasing Raldh2 expression (an enzyme of RA synthesis) (Aulehla et al., 2003; Aulehla et al., 2008; Diez del Corral et al., 2003; Dunty et al., 2008; Vermot and Pourquie, 2005; Zhao and Duester, 2009)

This model in which FGF, Wnt and RA interact allows both cell proliferation and regionalization of the extending embryonic axis. Indeed, regionalization of the mesoderm is concomitant with a segmentation process giving rise to the somites. In this case, FGF, Wnt and RA signals, as well as Notch, interact through the so-called “clock and wavefront model” (Cooke and Zeeman, 1976; Dequeant et al., 2006; Hubaud and Pourquie, 2014). This interaction permits the synchronized activation of segmentation genes in the presomitic mesoderm (PSM) in response to a periodical signal emitted by a mechanism acting as a segmentation clock. So, sustained oscillation occurs in the FGF, Wnt, and Notch signaling pathways mainly due to negative feedback processes (mediated by Lunatic fringe in the Notch pathway, by Axin2 in the Wnt pathway, and by MKP3/Dusp6 in the FGF pathway) (Aulehla et al., 2003; Dequeant et al., 2006; Ferjentsik et al., 2009). Oscillations start at the most posterior part of the presomitic mesoderm and at its anterior part, the opposition of the FGF and RA pathways define a border, the so-called wavefront, where the oscillations end, and the tissue is subdivided into a prospective somite with rostro-caudal polarity (Diez del Corral et al., 2003; Kumar and Duester, 2014). Thus, the opposition of the FGF and RA pathways defines the position of the future somite boundary during the process of segmentation. The final segmentation step is the establishment of a morphological somite boundary which involves a mesenchymal-to-epithelial transition. Moreover, RA signal, besides its role in the differentiation and patterning of the emerging somites, is also required for the synchronous and symmetrical left-right development of somites (Vermot and Pourquie, 2005).

In addition to FGF, Wnt and RA other genes, mainly controlled by these signals, play important roles in all vertebrates studied so far suggesting their conserved role across vertebrates. For example, the segmentation clock is defined by at least one member of the HES/HER family, (Hubaud and Pourquie, 2014). Thus, the oscillatory expression of the segmentation gene Hairy1, an effector of the Notch pathway, patterns the PSM (Palmeirim et al., 1997). Additionally, when FGF signal is inhibited the posterior expression of the master regulator for the segmentation clock Hes7 disappear (Niwa et al., 2007). Moreover, Axin2 a negative regulator of the WNT pathway shows oscillatory expression in the PSM. Additionally, mesoderm posterior (MESP) genes are master regulators of the segmental

program that forms the future segment boundary. They are activated immediately after the determination front, as a bilateral stripe. Indeed, inactivation of *Mesp1* and *Mesp2* genes in the PSM blocks segment formation in mouse (Oginuma et al., 2008). Interestingly, *Mesp2* is activated by *Tbx6*-mediated Notch signaling and repressed by the FGF signal (Kageyama et al., 2012) (**Figure 19**).

Taken together, all these data show an extraordinary complex mechanism controlling posterior elongation and somite formation in vertebrates. Moreover, even if conserved among vertebrates, slight differences exist between different vertebrate species, what complicates even more the whole picture. But despite the complexity of the system, a striking cross-talk between the FGF, WNT, Notch and RA signaling pathways is fundamental for the process. Then, the question that arises is the evolutionary origin of such mechanisms. In other words, is the molecular core for the control of vertebrates somitogenesis conserved in non-vertebrate chordates?

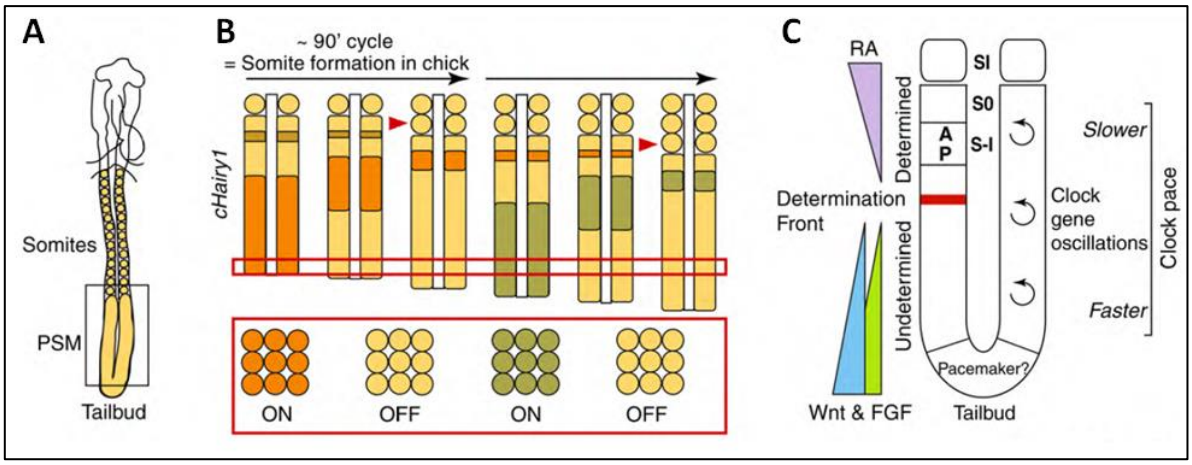


Figure 19. The clock and wavefront model. (A) In vertebrates the somites bud off from the anterior PSM periodically and sequentially. **(B)** In chick embryos one somite is formed approximately every 90 minutes, the oscillatory expression of the segmentation gene *c-Hairy1* is shown in different colors, the expression of individual cells (boxed in red) shows how the cells turn on and off the expression of a gene in synchrony, resulting in apparent waves of gene expression across the PSM. **(C)** The wavefront is characterized by opposite gradients of the signaling pathways, Wnt (light blue) and FGF (green) in the posterior tailbud and RA (purple) more anteriorly. The segmentation clock is represented on the right side of the scheme, showing that the oscillations slow down as they reach the anterior part of the PSM. Wnt activity seems to act as a pacemaker mechanism to regulate the periodicity of cyclic gene oscillations. Prospective somites in the PSM are numbered with S0 being the forming somite and the somite next to form labelled S –I and being already patterned in its rostral (A) and caudal (P) portion. The somite recently formed is numbered as S I. Modified from (Gibb et al., 2010).

1.3.2.2 Genetic control of somitogenesis in amphioxus: does it exist a clock and a wavefront?

In all vertebrates studied so far the "clock and wavefront model" appears as a conserved mechanism for the formation of somites. In amphioxus, gene expression of most of the genes involved in the "clock and wavefront model" have been described suggesting that these genes might be playing a similar role in amphioxus. Thus, it is possible to classify the genes involved in somitogenesis of amphioxus in two groups (i) those putatively involved in keeping a presomitic mesoderm, and (ii) those putatively involved in segmental processes. Belonging to the first group and expressed in the posterior mesoderm at gastrula stage and in the tail bud at neural stage are *Paraxis*, *Lcx*, *Pbx*, *Axin* and *OligoA* (Beaster-Jones et al., 2008). In the second group and expressed as stripes are *Delta* (Rasmussen et al., 2007), *Notch* (Holland et al., 2001), *Tbx15/18/22*, *Hey1*, *NeuroD/atonal-related*, *Pcdhδ2-17/18*, *Uncx4.1* (Beaster-Jones et al., 2008), *HairyB*, *HairyC* and *HairyD* (Minguillon et al., 2003), *IrxB* (Kaltenbach et al., 2009), *Ripply* (Li et al., 2006), and *Six1/2* (Kozmik et al., 2007). Interestingly, the first 8-10 somites which form by enterocoely, express a set of genes that are uniquely expressed in these somites and not in those derived from the tail bud. These genes are: *Engrailed*, a segmentation gene in flies, with detectable expression starting from early neurula stage (N1) (Holland et al., 1997); *HairyB*, a modulator of Delta-Notch signaling during segmentation of vertebrate somites (Gibb et al., 2010), with a detectable expression from gastrula stage (Minguillon et al., 2003); *Pbx*, a homeodomain containing cofactor for *Hox* genes (Laurent et al., 2008); and *OligA*, probably also involved in neurogenesis in amphioxus as it was shown in vertebrates (Bronchain et al., 2007). From these genes, only *Engrailed* and *HairyB* are expressed in the paraxial mesoderm as stripes, and might be playing a specific role during the segmentation of the anterior 8-10 somites (**Figure 20**) (Beaster-Jones et al., 2008).

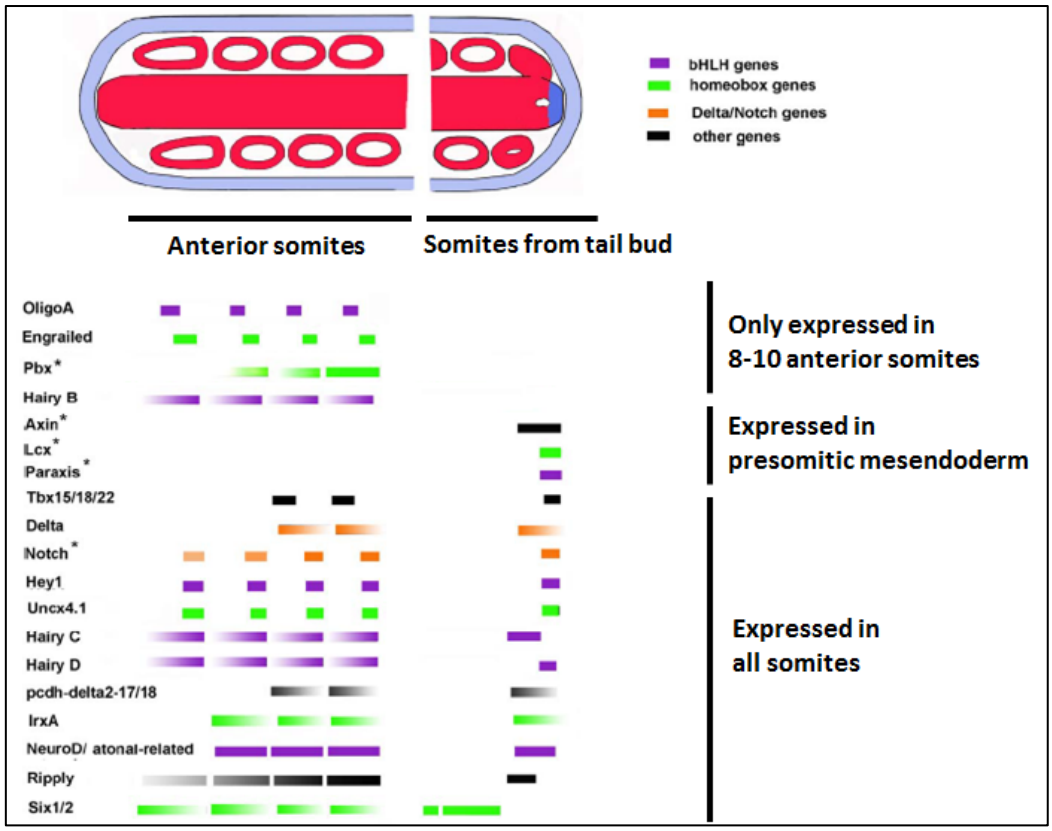


Figure 20. Scheme of segmental gene expression during amphioxus somitogenesis. Gene expression at midneurula stage (N2) in the anterior somites is on the left. Gene expression in the tail bud and the nascent posterior somites is on the right. Only *HairyB*, *Pbx*, *OligoA* and *Engrailed* are expressed in the enterocoelic somites (8-10 anterior somites). Expression of *Axin*, *Lcx* and *Paraxis* in the posterior presomitic mesoderm suggests a role to keep an undifferentiated state of the mesoderm. Vertebrate orthologues of *Tbx15/18/22*, *Delta*, *Notch*, *Hey1*, *Uncx4.1*, *HairyC*, *HairyD*, *Pcdhδ2-17/18*, *IrxA*, *NeuroD/atonal-related*, *Six1/2* and *Ripply* are involved in the segmental process suggesting a common role in amphioxus. Modified from (Beaster-Jones et al., 2008).

Moreover, in addition to the enterocoelic and schyzocoelic somites, several data suggest an additional division of the amphioxus somites, particularly within the anterior enterocoelic ones. Ontogenetically, the most anterior of these enterocoelic somites form simultaneously, whereas the most posterior form sequentially. Additionally, gene expression also differentiates these two somitic regions. Indeed, *Mox* is never expressed in the most anterior simultaneously formed somites, suggesting a functional difference with the posterior somites (Minguillon and Garcia-Fernandez, 2002). In addition, in our laboratory we have shown that inhibition of FGF or MAPK pathways at the blastula stage induces a complete loss of the most anterior enterocoelic somites, whereas formation of all the most posterior somites (enterocoelic and schyzocoelic) is independent of these two pathways (Bertrand et al., 2011). Therefore, we clearly establish the presence of three different somitic populations in amphioxus: (i) the most anterior enterocoelic, FGF-sensitive, Mox-negative, and Engrailed-positive somites; (ii) the posterior enterocoelic, FGF-insensitive, Mox- and Engrailed positive somites; and (iii) the posterior schizocoelic, FGF-insensitive, Engrailed-negative, and Mox-, Axin-, Lcx-, and Paraxis-positive somites (**Figure 21**).

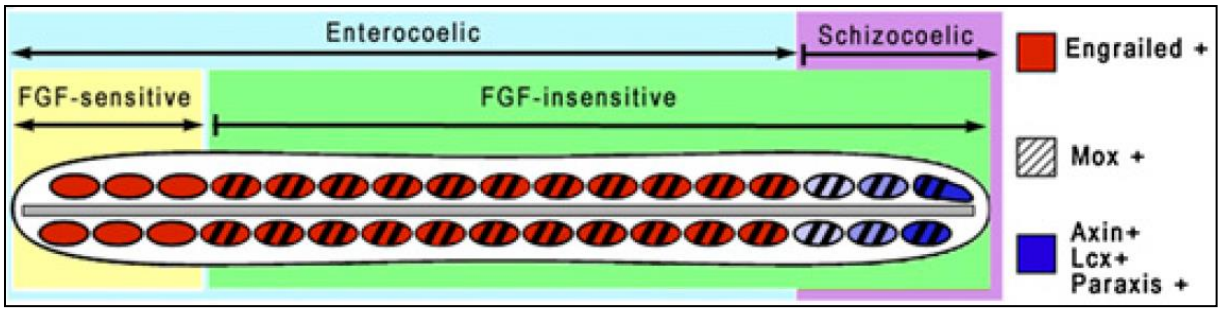


Figure 21. Posterior somite budding is not dependent on FGF signaling. The different somitic regions are depicted with different colors. The three most anterior somites formed by enterocoely are FGF-sensitive, whereas the posterior somites are FGF-insensitive. Modified from (Bertrand et al., 2011).

Concerning the "wavefront", in vertebrates, as I have explained before, it is characterized by an expression gradient of members of the FGF, RA and WNT families. In amphioxus, transcripts of *Wnts1*, *3*, *5*, *6*, and *8* are expressed in the posterior mesoderm at the gastrula and neurula stages and in the larval tail bud (Holland et al., 2000; Holland et al., 2005; Schubert et al., 2000; Schubert et al., 2001b) suggesting that members of the WNT family could play a role in maintaining an undifferentiated state of the presomitic mesendoderm before the segmentation. However, the small distance (10uM) in which these genes are expressed make difficult to determine whether there is or not a gradient of expression. Regarding the role of FGF signal, it was already shown that this signal is only involved in the formation of the most anterior somites and not in the posterior somites (Bertrand et al., 2011). Additionally, any study testing the role of the RA signal in amphioxus somitogenesis have been performed, leaving the question open about its ancestral role at the base of the chordates.

Taking together these known data concerning amphioxus somitogenesis, it is quite probable that the "clock and wavefront" system was selected specifically in vertebrates in parallel to the development of more complex somite-derived structures but that it was not required for somitogenesis in the ancestor of chordates. In addition, the specific role of FGF signal in the formation of the most-anterior somites in amphioxus also suggests that functional evolution of FGF signal was instrumental for the appearance of the vertebrates' head (Bertrand et al., 2011), something which has been the fundamental question of my PhD work, and will be discussed in depth later.

1.4 MYOGENESIS

In vertebrates, all the truncal muscles derive from the somites, and are controlled by a group of genes called the PSEDN network (comprising Pax-Six-Eya and Dach). Remarkably, the facial muscles possess a different developmental origin and are controlled by different master genes. Thus, in the following pages I will describe the development and genetic control of the trunk and facial muscles in vertebrates and I will compare these data with some known data of amphioxus myogenesis.

1.4.1 Genetic control of truncal myogenesis in vertebrates and amphioxus

1.4.1. 1 Truncal myogenesis in vertebrates: the Pax-Six-Eya-Dach network.

The conserved Pax-Six-Eya-Dach network (PSEDN) has been implicated in a variety of developmental processes in eyes, muscles, endocrine glands, placodes and pharyngeal pouches across bilaterians (Heanue et al., 1999; Kozmik et al., 2007; Schlosser, 2015). Interestingly, in vertebrates these network acts upstream of myogenic related factor genes (MRFs). Thus, *Pax3* and *Pax7*, two well-known myogenic genes, are expressed in the somites already formed, playing a direct role in the induction of myogenic cells, where *Pax3* plays a major role during primary myogenesis and *Pax7* in later phases of myogenesis (Maroto et al., 1997; Seale et al., 2000). Double mutant mice *Six1*$^{-/-}$*Six4*$^{-/-}$ and *Eya1*$^{-/-}$*Eya2*$^{-/-}$ show a loss of expression of *Pax3* in the hypaxial dermomyotome demonstrating that SIX transcription factors and their coactivators, control the expression of *Pax3* in the hypaxial dermomyotome (Grifone et al., 2007; Grifone et al., 2005). Additionally, studies have shown that SIX proteins can bind transcriptional regulatory sequences of *Myog* and *Myf5*, in conjunction with PAX3 protein in the epaxial dermomyotome, suggesting that both SIX and PAX3 can act in parallel to activate the myogenic program (**Figure 22**) (Giordani et al., 2007). Concerning the last member of the network, *Dachschund2* (*Dach2*), it plays a role together with *Eya2* to regulate myogenic differentiation. Additionally, *Eya2* acts together with *Six1* to regulate myogenesis (Heanue et al., 1999), showing a conserved role of the PSEDN.

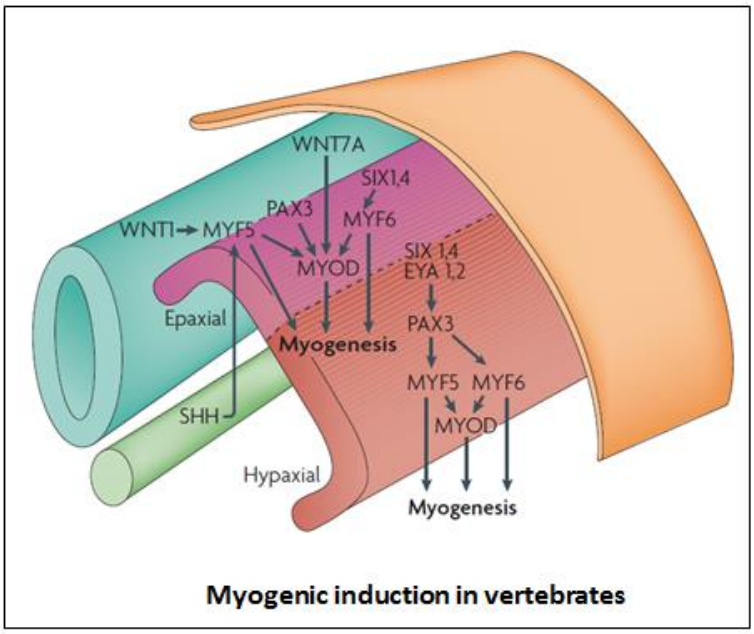


Figure 22. Myogenic induction in vertebrates. Schema of the signaling pathways and genetic interactions during the myogenic induction in vertebrate somites. In the epaxial muscles, PAX3, MYF5 and MYF6 induce *MyoD* expression independently. By contrast, in the hypaxial dermomyotome, PAX3 induces expression of *Myf5* directly, which in turn activates *MyoD* expression. WNT1 signalling from the dorsal neural tube induces myogenesis through direct activation of *Myf5*, whereas *Wnt7a* expression from the dorsal ectoderm preferentially activates *MyoD*. Hedgehog signalling pathway (SHH) also regulates myogenesis through the maintenance of *Myf5* expression. *Six1* and *Six4* regulates *Myf6* expression in the epaxial dermomyotome and together with their cofactors (*Eya1* and *Eya2*) induce the expression of *Pax3* in the hypaxial dermomyotome. Modified from (Bryson-Richardson and Currie, 2008).

1.4.1.2 Amphioxus myogenesis: the Pax-Six-Eya-Dach network.

Interestingly, and unlike vertebrates, where the complete segmentation of the somites occurs before the myogenic determination, in amphioxus, it seems that the onset of myogenesis and the segmental process are coupled (i.e. genes as *Delta*, *Notch*, *Tbx15/18/22*, *Hey1*, *Uncx4.1*, *HairyB*, *C* and *D*, *IrxB* and *Ripply* are expressed almost at the same time as genes involved in myogenic determination). Indeed, classical myogenic factors such as *MRF1* (*Myogenic Regulatory Factor 1*) and *MRF2*, which belong to a family of muscle-specific basic helix-loop-helix (bHLH) transcription factors and orthologues of *MyoD*, are associated in amphioxus with myogenic determination and are expressed in the paraxial mesoderm at the gastrula stage, later on at neurula stage *MRF1* is expressed in all the somites, whereas the *MRF2* expression become more restricted to the posterior somites. Finally at larva stage there is no longer expression of these transcription factors (Schubert et al., 2003).

Concerning the PSDEN network in amphioxus, orthologues of these genes show a dynamic expression pattern. Thus, at gastrula stage *Pax3/7* is expressed in the axial and paraxial mesoderm. At early neurula stage (N1) the expression is still detected in the paraxial mesoderm, and also new domains of expression are observed in the neural plate and endoderm. At mid-neurula stage (N2) a conspicuous signal is detected in the posterior axial

mesoderm (the nascent notochord), meanwhile the expression in the paraxial mesoderm begins to fade away except in the most anterior somites. At late neurula stage expression in the somites is no longer detectable except in the wall of the first somites on the left side (**Figure 23a-e**) (Holland et al., 1999). In amphioxus, the SIX family genes (specifically *Six1/2* and *Six4/5*) and their cofactor *Eya* are expressed in the paraxial mesoderm at the gastrula stage. Later, at mid-neurula stage *Six1/2*, *Six4/5* are expressed in all the presumptive somites excepting the most posterior (newly formed somites) where an expression of *Eya* is observed. At late neurula stage, when the somites start to be formed by schizocoely from the tailbud, *Six1/2* is expressed in all the somites, whereas *Six4/5* and *Eya* are only expressed in the most posterior two pairs. Finally at early larva stage *Six1/2*, *Six4/5* and *Eya* are expressed in the last three posterior somites (**Figure 23f-r**) (Kozmik et al., 2007). Concerning *Dach*, the homologue of *Dach2* in vertebrates, it is firstly expressed at late gastrula in the paraxial mesoderm, and this expression is detected in the forming somites until neurula stage (8-10 somites) when the somites are still formed by enterocoely. Later, the expression of *Dach* is no longer detected (**Figure 23s-v**) (Candiani et al., 2003). Up to date, any functional analyses of the PSEDN network have been done *in vivo* in amphioxus. However, *in vitro* it has been shown the protein-protein interaction between Six-Eya proteins as in other animals. But any interaction has been detected between Dach and Eya proteins as in *Drosophila* and some vertebrates like chicken (Chen et al., 1997; Heanue et al., 1999; Kozmik et al., 2007). Certainly, unravelling the specific role of the PSEDN network in myogenesis in amphioxus, could shed some light about the ancestral role of this network in the formation of muscles in the ancient chordates.

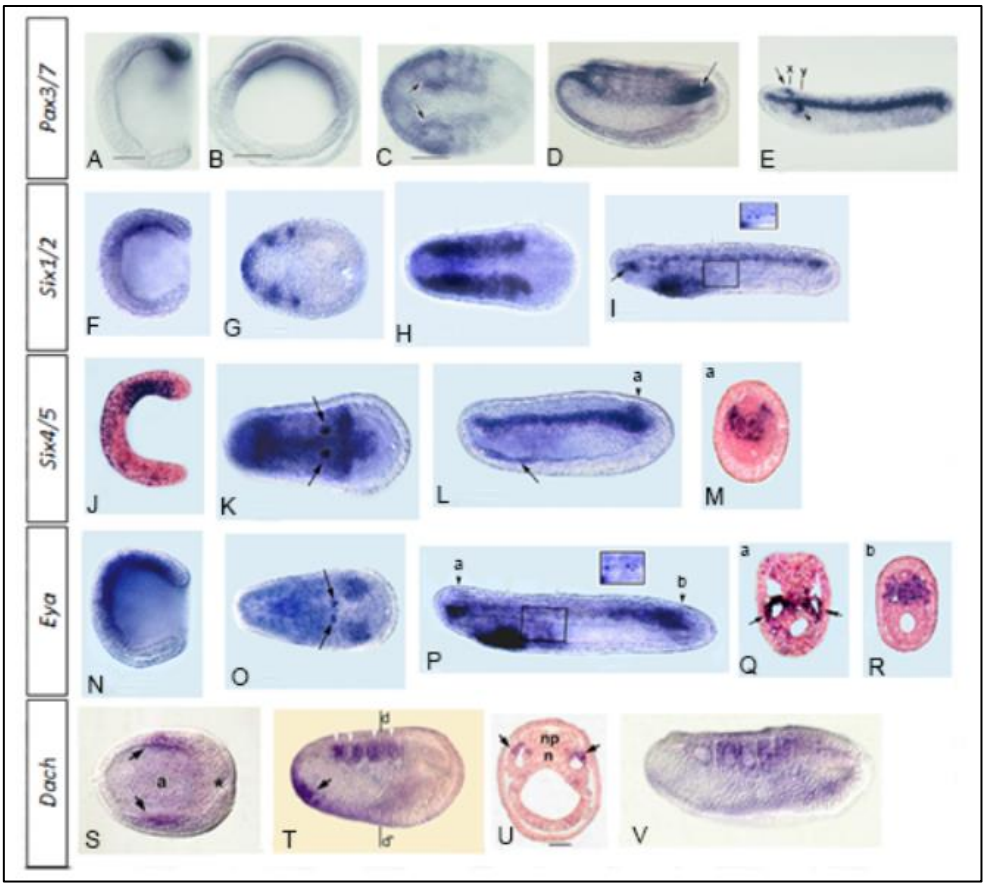


Figure 23. Expression pattern of the PSEDN in representatives stages of amphioxus. Lateral views for A, D, E, F, I, J, L, N, P, T and V. Blastoporal view for B. Dorsal view for C, G, H, K, O, and S. Cross-sections for J, M, Q and R. In all the panels (excepting cross-sections) anterior is to the left. **(A-E)** *Pax3/7* expression pattern. At early gastrula is observed a conspicuous expression in the axial and paraxial mesoderm (A-B), later on the expression is observed in the paraxial mesoderm (arrows in C). At mid-neurula stage is observed an expression in the nascent notochord **(F-I)** *Six1/2* expression pattern. First detected in the paraxial mesoderm and dorsomedially in the mesendoderm at gastrula stage (F), later from late gastrula transcripts are detected in all the somites except in the most posterior pair and in the pharyngeal endoderm **(J-M)** *Six4/5* expression pattern. At early gastrula stage *Six4/5* is detected in the ectoderm and dorsomedially in the mesendoderm (J), later at late neurula stage expression is detected in the nascent notochord, and in the somites except in the most posterior pair (K,L) a section through a is observed in M showing an expression in the notochord and posterior somites **(N-R)** *Eya* expression pattern. First detected in the dorsomeadilly mesendoderm at early gastrula stage (N), transcripts of *Eya* are later detected in the most posterior somites in late neurula and premouth stages (O, P). Cross-section in (a) and (b) shows the expression in the most anterior somites **(S-V)** *Dach* expression pattern. Transcripts of *Dach* are first detected at early neurula in the paraxial mesoderm and endoderm (arrow) (S). Cross-section through (d) shows mainly the expression of *Dach* in the dorsal part of the somites. At neurula stage *Dach* is detected in the first five somites and the forming somites. Later on *Dach* is no longer detected in the somites. Modified from (Candiani et al., 2003; Holland et al., 1999; Kozmik et al., 2007).

1.4.2 Formation of Facial Muscles in Vertebrates

1.4.2.1 Facial muscles in vertebrates

The head of vertebrates comprises at least six different groups of muscles with different developmental origins. Thus, the extra-ocular muscles (EOM) derive from the prechordal mesoderm whereas the masticatory and facial expression muscles both derive from the paraxial head mesoderm (Diogo et al., 2015). Other cell types also derive from the paraxial head mesoderm as the cardiomyocytes (Kelly et al., 2001; Mjaatvedt et al., 2001; Waldo et al., 2001), the posterior part of the neurocranium, and angiogenic cells (**Figure 24**) (Couly et al., 1992, 1993; Evans and Noden, 2006; Hacker and Guthrie, 1998; Noden, 1983). The other muscles that compose the head as the tongue and neck muscles (hypobranchial and cucullaris muscle groups) possess an embryonic origin that is still debated and might be derived mostly from cells provided by the most anterior somites or/and from the paraxial head mesoderm (Birchmeier and Brohmann, 2000; Czajkowski et al., 2014; Harel et al., 2009; Huang et al., 2001; Piekarski and Olsson, 2007; Theis et al., 2010). Thus, only three groups of muscles that compose the head are completely derived from the cranial mesoderm which comprises the prechordal mesoderm and the paraxial head mesoderm, these muscles are the EOM, the masticatory and the facial expression muscles.

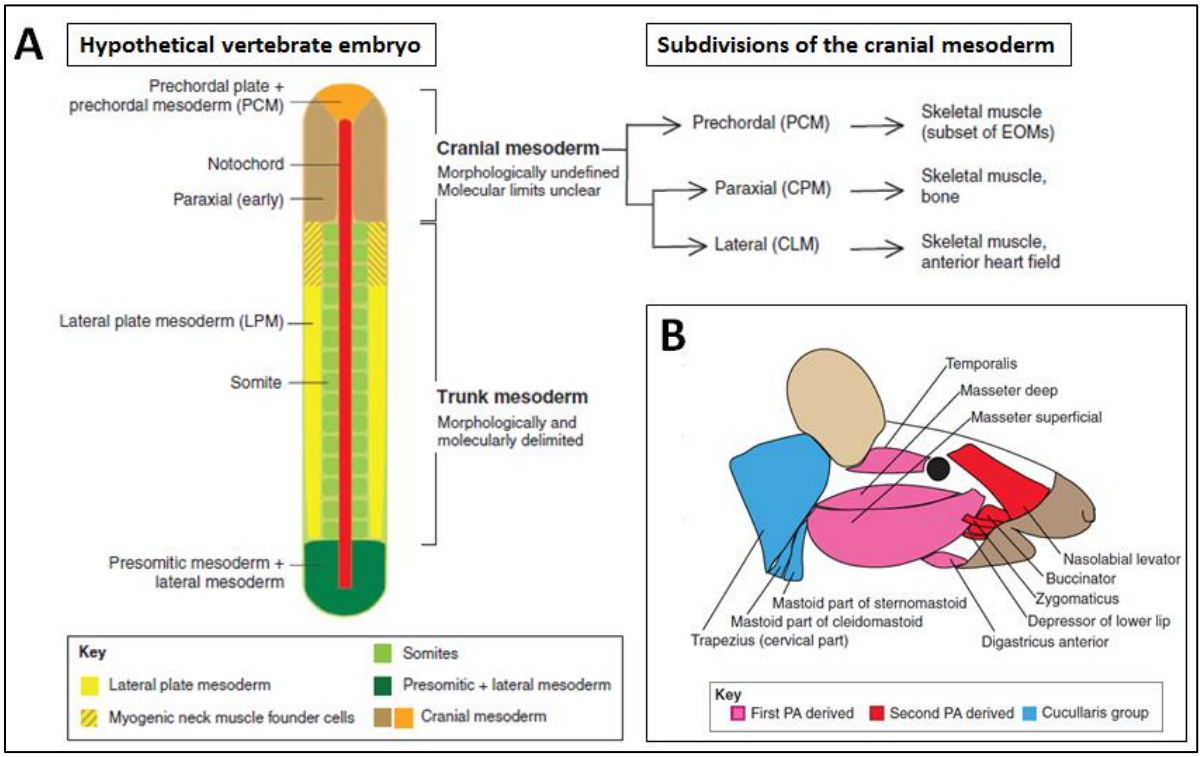


Figure 24. Head mesoderm and its derivatives. (A) Schema of a hypothetical 15-somites vertebrate embryo (dorsal view). The mesoderm that will give rise to the muscles along the embryo are colored according to the legend. Based on fate mapping and gene expression, the different subdivisions of the cranial mesoderm (head mesoderm) and its derivatives are shown (upright). **(B)** Schema of the vertebrate head muscles and their origins (legend). Abbreviations: (PCM) prechordal mesoderm, (LPM) lateral plate mesoderm, (EOMs) extra-ocular muscles, (CPM) cranial paraxial mesoderm, (CLM) cranial lateral mesoderm, (PA) pharyngeal arches. Modified from (Sambasivan et al., 2011).

1.4.2.2 Genetic programs for facial muscles formation in vertebrates

The activation of the head myogenesis program is different from the truncal myogenesis program. Nevertheless, in both cases the myogenic regulatory factors (MRFs), determine muscle identity and/or promote muscle differentiation (Buckingham, 2006; Sambasivan and Tajbakhsh, 2007). The different MRF genes (e.g., *Myf5*, *Mrf4*, *MyoD* and *Myogenin*) are able to trigger the differentiation of muscles in both head and trunk (Molkentin and Olson, 1996). However, the upstream regulators required for the activation of the trunk muscle program are different from those required for the head muscle program. Thus, genes as *Tbx1*, Pitx2, *Tcf21* (also known as *Capsulin*), *Msc* (*MyoR*) and *Lhx2* represent the upstream regulators in the head (Bothe et al., 2007; Bryson-Richardson and Currie, 2008; Grifone and Kelly, 2007; Sambasivan et al., 2011; Tzahor, 2009, 2015; Tzahor and Evans, 2011). Indeed, *Tbx1*, a T-box transcription factor, has an important role regulating the expression of *Myf5* and *MyoD* in all branchiomeric muscles (Dastjerdi et al., 2007; Grifone et al., 2008; Kelly et al., 2004). *Pitx2*, a bicoid-related homeodomain transcription factor, has a crucial role in specifying EOMs and it is a key factor in the development of branchiomeric muscles (Dong et al., 2006; Kitamura et al., 1999; Shih et al., 2007). Both *Tbx1* and *Pitx2* might act in a cooperative way activating the same target genes, explaining why myogenesis is still observed in branchiomeric muscles of double mutant mice Tbx1:Myf5 (Nowotschin et al., 2006; Sambasivan et al., 2009). Moreover, *Tcf21* and *Msc*, both bHLH transcriptional repressors, are required by *Myf5* to activate the muscle program in the first mandibular arch (Lu et al., 2002). Finally, *Lhx2*, a LIM-homeobox transcription factor, has been identified as a new player in the development of pharyngeal and cardiac muscle formation (**Figure 25**) (Harel et al., 2012). In the case of the trunk program, as was shown before, myogenesis is controlled by the PSEDN network. Concerning the role of PSEDN factors in head muscles formation, it has been shown that the expression of *Pax3* is absent in head muscle stem/progenitor cells (Hacker and Guthrie, 1998; Sambasivan et al., 2009; Tajbakhsh et al., 1997). Interestingly, *Six1*, a critical factor in trunk myogenesis, has been shown to be implicated in head muscle development in zebrafish (Lin et al., 2009) and is expressed in the anterior head mesoderm (Bothe et al., 2011). Thus, suggesting different roles for Pax and Six proteins in the formation of head muscles.

1.4.2.3 Other regulatory signals playing a role in head and trunk myogenesis

Other signaling pathways also play important roles in myogenesis and also differ between the head and trunk myogenesis. For example, bone morphogenetic protein (BMP) and Wnt/β-catenin signaling pathways appear as inhibitors of myogenesis in the head paraxial mesoderm, and the inhibitors of these signaling pathways (e.g. Noggin, Gremlin and Frzb) are inducers of head myogenesis (Tzahor et al., 2003). On the contrary, the Wnt signal emanating from the dorsal neural tube plays an inductive role in truncal myogenesis (Munsterberg et al., 1995; Stern et al., 1995; Tajbakhsh et al., 1998). The role of the Fibroblast Growth Factor (FGF) signaling pathway in myogenesis has not completely been clarified. Thus, FGF signal induces the expression of myogenic genes in the primary myotome triggering the appearance of muscle progenitors (Delfini et al., 2009). Moreover, in fish *fgf8* is required for *MyoD* expression in the somites (Groves et al., 2005; Reifers et al., 1998). However, it has been shown that in some myoblast cultures, the FGF signal induce

differentiation, whereas in other cultures FGF signal acts as a repressor (Clegg et al., 1987; Olwin and Rapraeger, 1992; Seed and Hauschka, 1988). Regarding the role of the FGF signal in the head, experiments using FGF8 loaded beads grafted into the head mesoderm of chick embryos resulted in the downregulation of *Myf5* and *Pitx2* expression and in the upregulation of the expression of *MyoR* and *Tbx1*. Beads loaded with the FGF signaling inhibitory molecule SU5402 slightly upregulated *Pitx2* and suppressed *MyoR* and *Tbx1* expression. In addition, the inhibition of FGF signaling pathway using retinoic acid resulted in the downregulation of *MyoR* and *Tbx1* expression (Bothe et al., 2011; von Scheven et al., 2006). Thus, these experiments suggest a direct role of FGF signaling for the muscle master regulators *Tbx1* and *Pitx2*, activating *Tbx1* and suppressing *Pitx2* expression in the head mesoderm. Finally, the ligand of the Notch pathway, *Delta 1* (*Dll1*) is not required in the head myogenic progenitors neither for satellite cell population nor for Pax7 expression, this is in an opposite way to its role in somites-derived muscle (**Figure 25**) (Czajkowski et al., 2014).

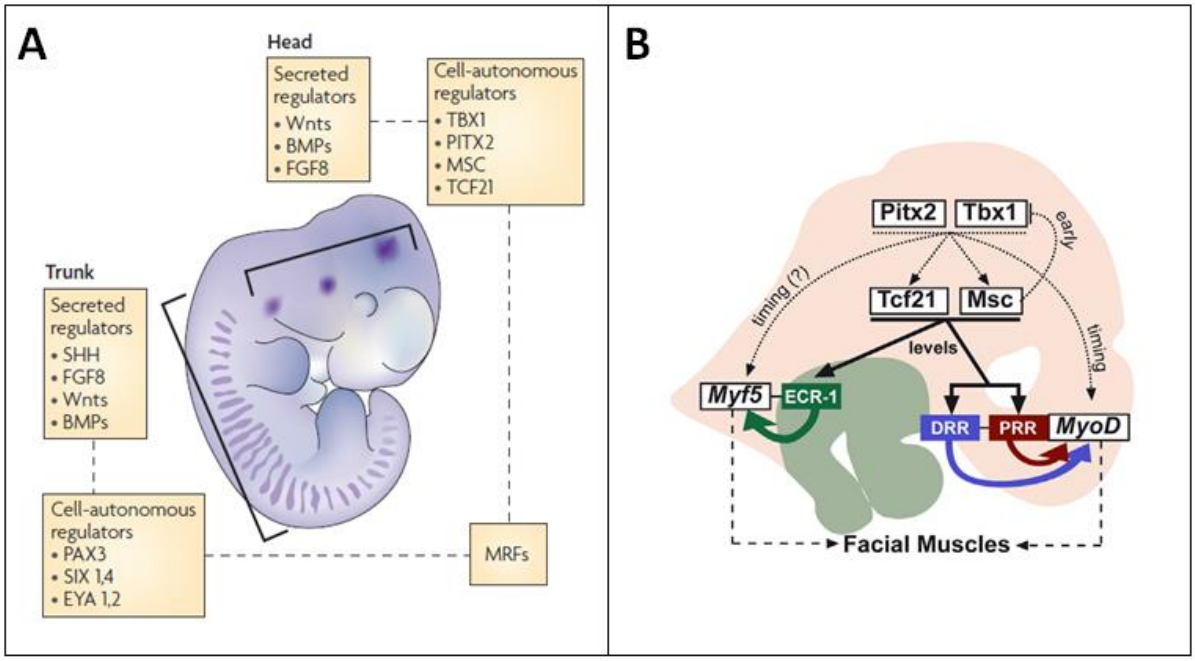


Figure 25. Regulatory signals and genetic programs involved in muscle development in head and trunk. (A) Different studies have highlighted the diverse roles played by the intercellular signaling pathways such as the Wnt/β-catenin, BMP or FGF signals in the control of head and trunk myogenesis. In addition, the gene regulatory network for the activation of the myogenic program is completely different in head and trunk. Thus, transcription factors as *Tbx1*, *Pitx2*, *M*sc (*MyoR*) and *Tcf21* are upstream of MRFs during head myogenesis, whereas transcription factors as *Pax3*, *Six1*,*4* and the cofactors of the last *Eya1*,*2* are key regulators of trunk myogenesis. Modified from (Bryson-Richardson and Currie, 2008). **(B)** In the branchial arches *Pitx2* and *Tbx1* act at the top of the cascade in the myogenic program. They turn on the transcription factors *Tcf21* and *Msc* that bind *cis*-regulatory regions of *Myf5* (ECR-1) or *MyoD* (DRR and PRR). Modified from (Moncaut et al., 2012).

1.4.3 Orthologues of vertebrate head muscles formation in amphioxus

As discussed earlier, the upstream regulators of the myogenic program in the vertebrate's head are different from those of the trunk. In amphioxus, to date, there is no study describing the expression pattern of neither *Lhx2* nor *Msc* (*MyoR*) (although orthologous of these genes are found in the genome of amphioxus (personal information)). The LIM-homeobox family (i.e. *Lhx1/5*, *Lhx3/4*, *Lmx*, *Lhx2/9* (*Apterous*), *Lhx6/8* (*Arrowhead*) and *Islet*) play a conserved role in neuronal specification in metazoans (Srivastava et al., 2010). In amphioxus, members of the LIM-homeobox family as *Lhx3*, *Islet*, and *Lim1/5* (*Lhx1/5*) might be involved in processes such as hindbrain segmentation or neuronal specification but not in myogenesis (Jackman et al., 2000; Langeland et al., 2006; Wang et al., 2002). Presumably *Lhx2* in amphioxus could be involved in neural specification. The expression of orthologues of upregulators of the myogenic head program in vertebrates as *Tbx1* or *Pitx2*, have been described in amphioxus. Thus, *Tbx1/10* is expressed in the endoderm of the gill slits and in the ventral part of the first 10-12 somites at the neurula stage (**Figure 26a-b**) (Mahadevan et al., 2004). In amphioxus, *Pitx* (the orthologue of *Pitx1, 2* and *3* from vertebrates) is expressed in the left side of the mesoderm, ectoderm and endoderm at the neurula stage suggesting a role in left-right asymmetry patterning, as it is the case for *Pitx2* in vertebrates. However, no expression is detected in somites during their formation (**Figure 26c-e**) (Boorman and Shimeld, 2002). Thus, orthologues of the upstream regulators of the myogenic program of the head in vertebrates, are not expressed at early stages in amphioxus, suggesting that they are not involved in the determination of the most anterior somites in amphioxus.

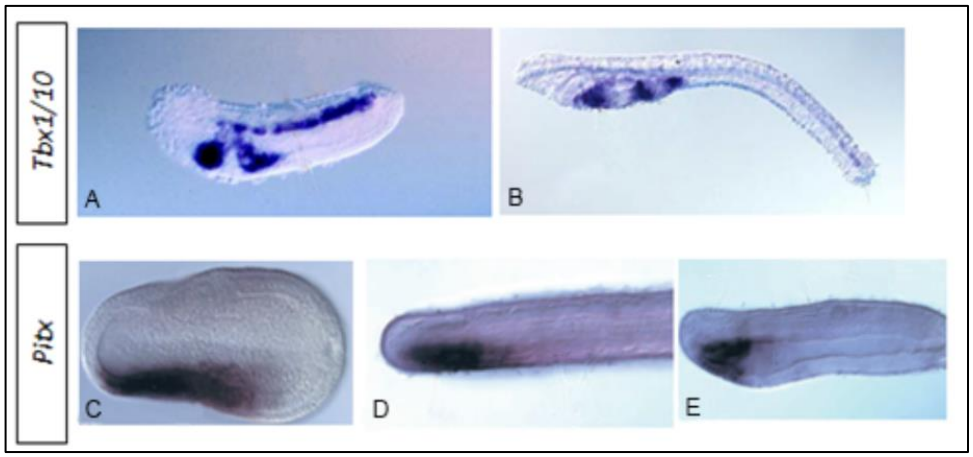


Figure 26. Expression pattern of *Tbx1/10* and *Pitx* in amphioxus. (A-B) First detected at mid-neurula stage, *Tbx1/10* transcripts are observed in the ventral half of the somites and in the ventral branchial arch mesoderm and endoderm. At larva stage transcripts are detected in the first three branchial arch whereas expression in the somites is no longer observed **(C-E)** *Pitx* expression is observed in the left anterior mesoderm, ectoderm and endoderm at neurula stage, this expression continues through the development of amphioxus suggesting a role in the left-right asymmetry of amphioxus. Modified from (Boorman and Shimeld, 2002; Mahadevan et al., 2004)

1.4.4 Final remarks

As we have seen, a major difference exists in the control of myogenesis in the truncal and head muscles in vertebrates. The muscles of the head are triggered by different master regulators such as *Tbx1*, *Pitx2*, *Lhx2*, *Tcf21*, and *Msc*, whereas for the truncal muscles the Pax-Six-Eya-Dach network have a major role activating the muscle differentiation program. Nevertheless, for both the head and truncal muscles the differentiation program is triggered by members of the same gene family, the MRF genes. In amphioxus, it seems that the muscles differentiation program for the whole body is activated by the Pax-Six-Eya-Dach network. However, to date there is no study showing the specific role of this network in the formation of the muscle and somites in amphioxus.

Concerning somitogenesis, in vertebrates it has been very well studied, thus the "clock and wavefront" model seems to be valid for all the vertebrates studied so far. Hence, whether or not this model is conserved across chordates, brought different laboratories to perform comparative studies using amphioxus as a model system. In their studies, they look at the gene expression pattern of several genes involved in the "clock and wavefront" model. Notably, several genes involved in the segmental process in vertebrates are expressed in amphioxus posterior somites, suggesting a conserved function of these genes in both vertebrates and amphioxus. Also, they found that members of the WNT family are expressed in the tail bud of amphioxus suggesting a similar role in amphioxus as in vertebrates (keeping an undifferentiated state of the mesendoderm). However, it is difficult to determine whether there is or not a gradient, due to the tiny length of the amphioxus tailbud compared with the PSM of vertebrates. Concerning the role of retinoic acid, even if different studies have been published on the role of retinoic acid during amphioxus development (Escriva et al., 2002; Schubert et al., 2004; Schubert et al., 2005), any functional study has been performed to test its role during amphioxus somitogenesis, leaving this question open.

Finally, although the FGF signaling pathway is not required for posterior somitogenesis in amphioxus as it is the case in vertebrates, recent studies performed in our laboratory have shown a direct role of this signal for the formation of the anterior somites. This observation has extraordinary evolutionary implications, since it suggests that changes in the function of FGF signaling played a crucial role during evolution for the origin of the vertebrates' head. In the next chapter, I will discuss the implications of this assumption.

1.5 FGF SIGNALING PATHWAY AND ANTERIOR SOMITOGENESIS IN AMPHIOXUS

1.5.1 FGF signaling pathway

Firstly identified in vertebrates 40 years ago, the Fibroblast Growth Factors (FGFs) form a family of extracellular signaling peptides that are crucial to activate different processes during the development of many metazoan organisms (Itoh and Ornitz, 2011). In the presence of heparin sulphate proteoglycan (HSPG), FGFs act through binding to a dimeric form of its tyrosine kinase receptor (FGFR), triggering the transphosphorylation and activation of the intracellular TK domain of the receptor and finally activate several intracellular cascades (i.e. Ras/MAPK, PI3K/Akt and PLCγ/PKC). In vertebrates, because of the whole genome duplications (2R or 3R) there are between 22 to 27 FGFs and 4 FGFRs (Oulion et al., 2012b). It seems that the activation of Erk1/2 MAP kinases is a common response for all the 4 FGFRs whereas p38 and Jun kinases could be activated in a cell-type specific manner (**Figure 27**) (Mason, 2007). In contrast with these complexity in vertebrates, amphioxus only possesses eight FGFs and one FGFR (Oulion et al., 2012b).

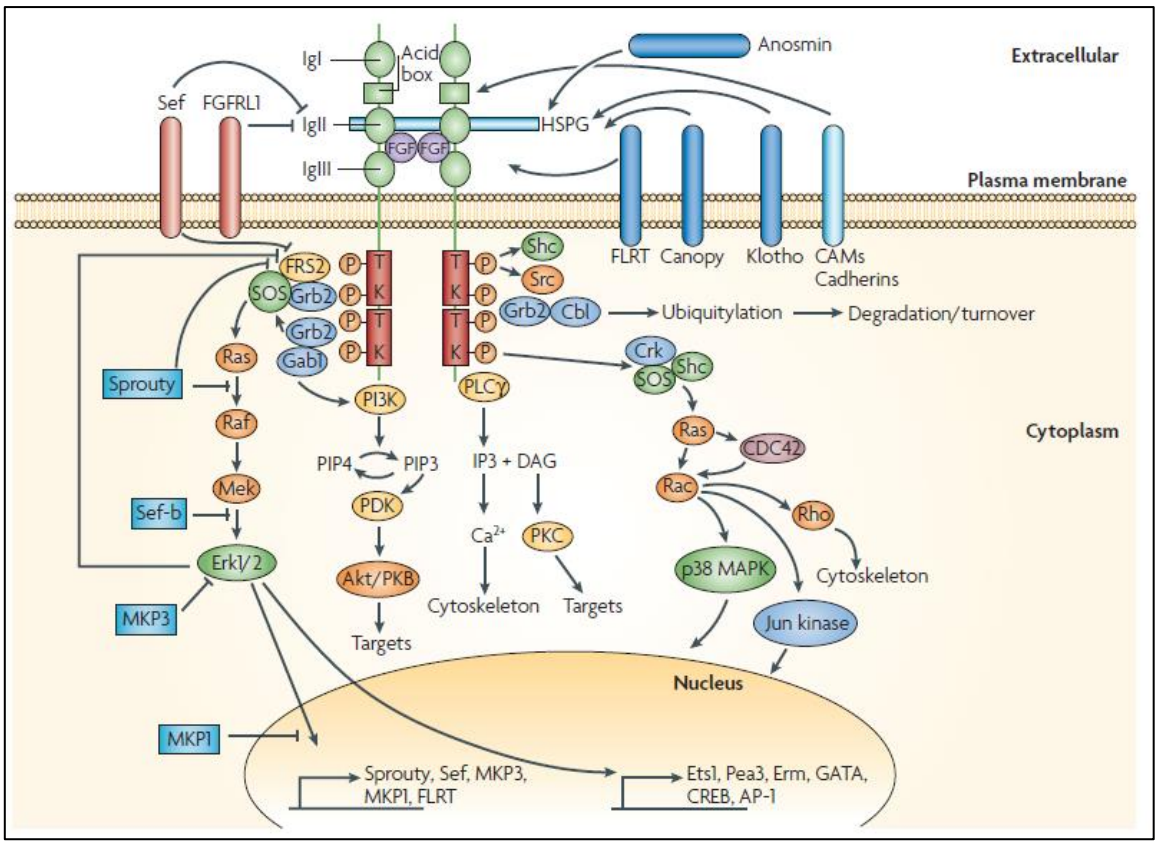


Figure 27. Schematic representation of the FGF signaling pathway. Once FGFs interact with their receptor they trigger a multiplicity of intracellular signaling cascades (i.e. Ras/MAPK, PI3K/Akt and PLCγ/PKC). The Erk-MAP kinase pathway has been most widely implicated in FGF developmental functions to date and is activated by Ras downstream of an FRS2-SOS-Grb2 complex. Abbreviations: (CAM) cell adhesion molecule, (CREB) cyclic AMP response element binding protein, (FLRT) fibronectine leucine-rich transmembrane proteins, (FRS) FGF receptor substrate, (HSPG) heparin sulphate proteoglycan, (Ig) immunoglobulin, (IP3) inositol triphosphate, (MAPK) mitogen-activated protein kinase, (MPK) MAPK phosphatase, (PI3K) phosphatidylinositol-3-kinase, (PIP4)

phosphatidylinositol-4-phosphate, (PKB) protein kinase B, (PLCγ) phospholipase Cγ, (SOS) son of *sevenless*, (TK) tyrosine kinase. Modified from (Mason, 2007).

1.5.2 Expression of FGFs in amphioxus

In vertebrates, *fgf8* and *fgf4* play important roles in somitogenesis. Indeed, deleting the expression of these genes leads to the loss of expression of most PSM genes, including cycling genes, Wnt pathway genes and markers of undifferentiated PSM, suggesting a major role for *fgf8* and *fgf4* to keep an undifferentiated state of the PSM (Boulet and Capecchi, 2012; Naiche et al., 2011; Niwa et al., 2011). In amphioxus, it has been shown a dynamic expression during embryonic development for five (*FGF8/17/18*, *FGF9/16/20*, *FGFA*, *FGFE* and *FGFC*) of the eight FGFs described (Bertrand et al., 2011). Interestingly, *FGF8/17/18*, the orthologue of *fgf8* of vertebrates, is expressed in the posterior dorsal mesendoderm at the gastrula stage. Later on, the expression in the mesoderm fades away rapidly and is no longer visible at midneurula stage (N2). The orthologue of *fgf4* of vertebrates in amphioxus is *FGFE* and its expression is detected by *in situ* hybridization only at the midlate neurula stage (N2) in the first left somite. The other FGF ligands do not show any expression in the mesendoderm that will give rise the posterior somites (**Figure 28**). Taken together, these results suggest that FGF signal is not playing a crucial role for the formation of the posterior somites. Finally, in amphioxus, FGFR is expressed ubiquitously at all developmental stages except in the epidermis, and with a higher expression level in the mesoderm (Bertrand et al., 2011).

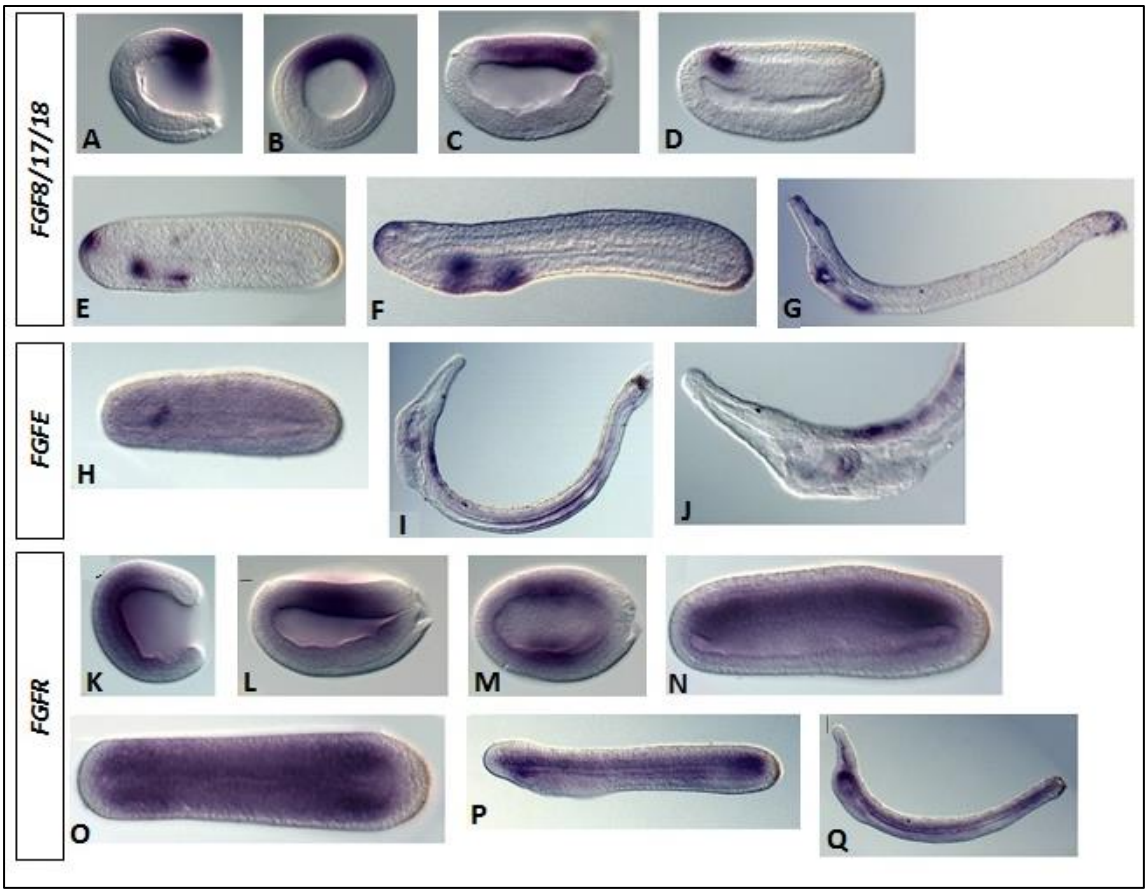


Figure 28. *FGF8/17/18*, *FGFE* and *FGFR* expression in amphioxus. (A-G) Dynamic expression of *FGF8/17/18* in amphioxus. (A,B) posterior dorsal mesendoderm expression at the gastrula stage. This expression fades at early neurula stage (C). Then a transient expression is observed in the cerebral vesicle (D). (E,F). Pharyngeal expression is observed at late neurula (E), premounth (F) and larva stage (G). **(H-J)** *FGFE,* the amphioxus orthologue of *fgf4* in vertebrates, shows an expression restricted to the first left somite at the mid neurula stage (H). At the larva stage, the expression is observed in the neural tube, the gut, and the club-shaped gland. **(K-Q)** (K) Gastrula stage embryo showing expression in the anterior mesendoderm. (L,M) At early neurula stage expression is observed in the paraxial mesoderm. (N,O) At midlate neurula stage expression is observed in the mesoderm, particularly in the most anterior and most posterior somites. (P) In premouth stage embryo, expression is observed in the notochord, the posterior somites and the anterior pharyngeal endoderm. (Q) Finally, at larva stage, expression is observed in the notochord and the anterior pharyngeal endoderm. Anterior is to the left in all embryos. Blastoporal view (E), dorsal views (M,O), lateral views for the rest. Modified from (Bertrand et al., 2011).

1.5.3 FGF signaling pathway controls anterior somitogenesis in amphioxus

Inhibition of the FGF signaling pathway using the FGFR inhibitor SU5402 at the blastula stage in amphioxus embryos leads to a loss of the anterior somites. Indeed, embryos treated with this inhibitor showed a loss of expression in the anterior paraxial mesoderm of genes as *Brachyury2* (involved in mesoderm formation and differentiation of the notochord (Holland et al., 1995)), *Delta* (involved in segmentation in vertebrates (Rasmussen et al., 2007)), *Snail* (establishes a muscle/notochord boundary in *Ciona* (Fujiwara et al., 1998)), and *MRF1* (determined muscle identity and/or muscle differentiation (Buckingham, 2006)).

Interestingly, this treatment does not affect the expression of genes such as *Neurogenin* (neuron differentiation), *Chordin* (axial dorsal mesendoderm marker) and *Nodal* (in amphioxus at gastrula stage it is expressed in the paraxial mesoderm (Yu et al., 2002)) suggesting that Nodal pathway is not under the control of the FGF signalling pathway (**Figure 29**). In addition, the work performed in our laboratory showed that for the formation of the anterior somites, the FGF signal acts through the Ras/MAPK pathway. Thus, embryos treated at the same stages with the MAPK pathway inhibitor U0126 have a similar phenotype as the embryos treated with the inhibitor SU5402 (Bertrand et al., 2011) (**Figure 29**). In the same work, my colleagues showed that formation of the posterior somites is not controlled by the FGF signal. In fact, treatments at blastula stage with SU5402 do not affect the formation of the posterior somites. Moreover, later treatments (late gastrula stage) do not inhibit the formation of any somite, anterior nor posterior, although their morphology look impaired. Thus, these results indicated a crucial role of the FGF signal before the gastrula stage for the formation of the anterior somites (Bertrand et al., 2011).

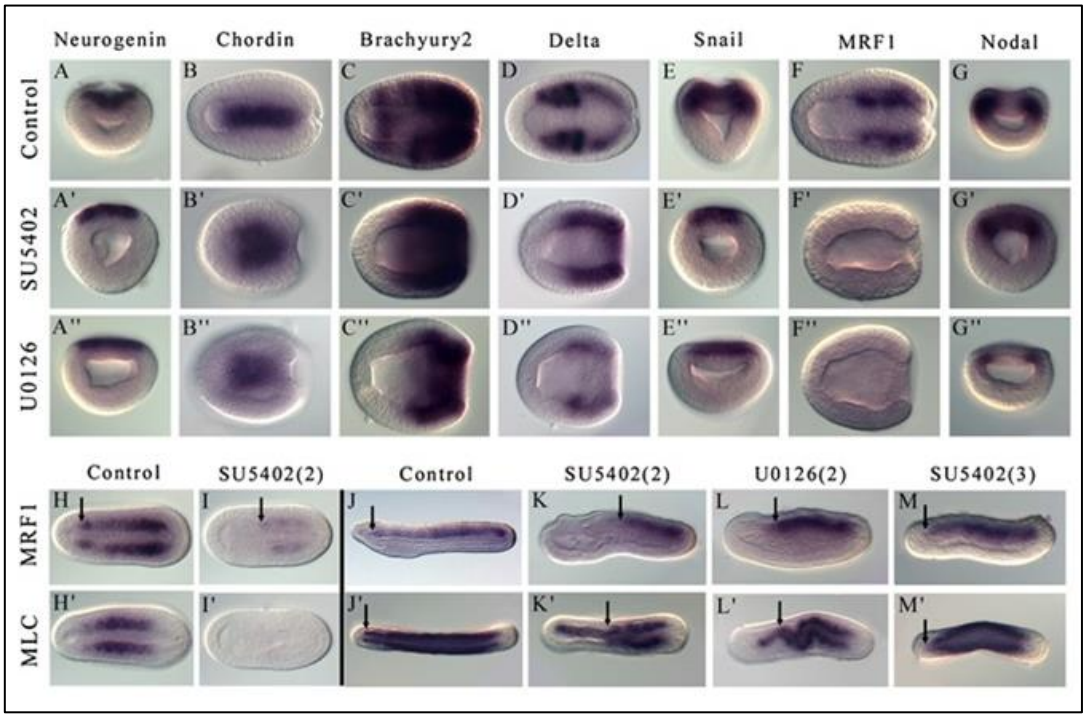


Figure 29. FGF and MAPK signaling pathway inhibition induce the loss of the most anterior somites. Expression patterns by whole-mount *in situ* hybridization of *Neurogenin* (A-A''), *Chordin* (B-B''), *Brachuyry2* (C-C''), *Delta* (D-D''), *Snail* (E-E''), *MRF1* (F-F''), *Nodal* (G-G''), and *MLC* (H'-L') after treatments with SU5402 (50 µm) or with U0126 (25 µm). Embryos treated at blastula stage were fixed at the late gastrula stage (A-G''), at the midneurula stage (H, H' and I, I'), or at the premouth stage (J-L and J'-L'). Embryos treated at late gastrula stage were fixed at premouth (M,M'). A-A'', E-E'', and G-G'' are blastopore views. B-B'', C-C'', D-D'', F-F'', H'H', I-I' and J'-M' are dorsal views. J-M are lateral views. The most anterior limit of MRF1 and MLC is labeled by a black arrow. Anterior is to the left in dorsal and lateral views, and dorsal is to the top in side and blastoporal views. Extracted from (Bertrand et al., 2011).

1.5.4 Final remarks

It is tempting to speculate that in order to evolve a new structure, it is necessary to first lose the ancestral functional structure. Particularly, in the case of the evolution of the vertebrate's head, we can imagine that loss of the anterior somites in the ancestor of vertebrates liberated the developmental constraints imposed by these structures allowing the appearance of new structures. In other words, the appearance of the head in vertebrates had to be preceded by the loss of anterior somites but preserving the structures developed from the ectoderm and endoderm. As I have shown in this chapter, the formation of the most anterior somites in amphioxus is controlled by the FGF signal, and the inhibition of the FGF signal does not affect other germ layers as the ectoderm or endoderm. Thus, an interesting hypothesis has been suggested, in which functional evolution of the FGF signal during early development in the ancestor of vertebrates, played a major role in the evolution of the head through the loss of anterior somites.

Finally, these first studies in amphioxus, both about anterior and posterior somitogenesis, opened many questions. Some of these questions are:

1) Is there a role for RA in somitogenesis in amphioxus? If this role exists, is there any opposition between RA/FGF signaling pathways in amphioxus somitogenesis as in vertebrates? Does RA play a role in the control of the amphioxus asymmetric somites?

2) Why the anterior somites but not the posterior somites are controlled by the FGF signal? How the boundary between the anterior FGF-sentive somites and the posterior FGF-insensitive somites is established along the antero-posterior axis of amphioxus?

3) Since in vertebrates, different master genes trigger the expression of the MRF genes in the head and in the trunk, can we expect in amphioxus any difference between the anterior and posterior somites regarding genes that control the expression of MRF genes?

And finally, probably one of the most important questions would be:

4) Which gene regulatory network is triggered by the FGF signaling for the control of anterior somitogenesis in amphioxus?

2. Introduction of Article 1

In this first article, I tried to understand the implication of different classical signals, playing a role in vertebrates' somitogenesis, such as FGF, RA, Wnt or Nodal, in the control of amphioxus somitogenesis. I also tried to understand whether a functional crosstalk exists in amphioxus between the FGF signal and Hox genes to control the limit of the anterior FGF-sensitive somites. Therefore, the *Hox* genes (that encode transcription factors) are known for their role in patterning the antero-posterior body axis (Lewis, 1978). Thus, in vertebrates, *Hox* genes play a critical role in the patterning of the axial skeleton (Wellik, 2007). *Hox* genes are organized in clusters and are expressed in a collinear way so that the *Hox* genes located at the 3' end within the cluster are expressed before and more anteriorly and those at 5' end are expressed later and more posteriorly along the anteroposterior axis of the embryonic trunk. In vertebrates, because of the two rounds of whole genome duplication there are four *Hox* gene clusters (Holland and Garcia-Fernandez, 1996). Interestingly, changes in the expression of the *Hox* genes can lead to shifts in the normal development of the embryo. Thus, for instance, retinoic acid (RA) added ectopically, activates *Hox* expression anteriorly inducing a posteriorization of the embryo and changes in the axial pattern (Kessel, 1992).

Amphioxus possesses only one *Hox* cluster (from *Hox1* to *Hox15*), showing also a collinear expression (Holland et al., 2008a; Pascual-Anaya et al., 2012). Remarkably, in amphioxus, the posterior limit of the anterior-most somites (the FGF sensitive somites) coincides with the anterior limit of expression of *Hox1* (**Figure 30**) (Bertrand et al., 2011; Wada et al., 1999). Therefore, it might be possible that the anterior *Hox1* limit could act as a frontier to differentiate between the anterior-most FGF-sensitive somites, and the posterior FGF-insensitive somites. In amphioxus, RA can activate the expression of *Hox* genes and like in vertebrates the embryo shows a posteriorization after RA treatment. Moreover, embryos treated with the RAR antagonist BMS009 show an anteriorization (Escriva et al., 2002). Thus, to test if *Hox* genes are acting as a frontier between these two types of somites (FGF-sensitive and insensitive somites), we shifted the expression of *Hox1* anteriorly or posteriorly with RA or the RAR antagonist, respectively. Then, we expected two possible outcomes;

1) The limit between FGF-sensitive and insensitive somites moves forward or backward, meaning that anterior *Hox* genes define the limit between FGF-sensitive and FGF-insensitives somites.

2) The limit between FGF-sensitive and insensitive somites do not moves, meaning that anterior *Hox* genes do not define the limit between FGF-sensitive and FGF-insensitives somites.

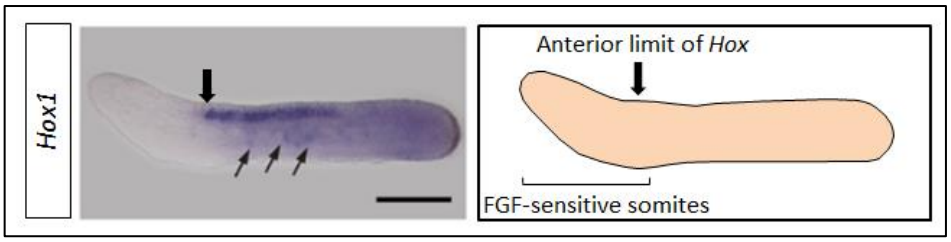


Figure 30. *Hox1* expression in a premouth embryo. Anterior limit of Hox1 coincides with the posterior limit of FGF-sensitive somites. Modified from (Pascual-Anaya et al., 2012)

Our results showed that the limit between the FGF-sensitive and FGF-insensitive somites is not controlled by Hox genes. Furthermore, in order to better understand the evolution of somitogenesis, we also tried to decipher the role of RA and FGF during the formation of the posterior somites. In vertebrates, posterior somitogenesis is controlled in part through the opposition between the posterior FGF/Wnt proliferation signal and the most anterior RA differentiation signal. We were able to show that it is not the case in amphioxus in which RA and FGF signals are not implicated in posterior somitogenesis and in which both signals do not seem to interact through a negative crosstalk. We also showed that the asymmetry of amphioxus somitogenesis is controlled by the Nodal signal in amphioxus and that this left/right asymmetry signal cannot be buffered by RA in contrary to vertebrates.

2.1 Evolution of the role of RA and FGF signals in the control of somitogenesis in Chordates

PLOS | ONE

RESEARCH ARTICLE

Evolution of the Role of RA and FGF Signals in the Control of Somitogenesis in Chordates

Stéphanie Bertrand[1]*, Daniel Aldea[1], Silvan Oulion[1¤a], Lucie Subirana[1], Angel R. de Lera[2], Ildiko Somorjai[1¤b], Hector Escriva[1]*

1 UPMC Univ Paris 06, UMR 7232, BIOM, Observatoire Océanologique de Banyuls sur Mer, F-66650, Banyuls/Mer, France, **2** Departamento de Química Orgánica, Facultade de Química, CINBIO, Universidade de Vigo, and Instituto de Investigación Biomédica de Vigo (IBIV), Vigo, Spain

¤a Current Address: Département Forme, Institut des Sciences de l'Evolution, UMR 5554, CNRS/Université Montpellier II, Place Eugène Bataillon, Montpellier, France
¤b Current Address: Gatty Marine Laboratory, Scottish Oceans Institute, East Sands, University of St Andrews, St Andrews, Scotland
* stephanie.bertrand@obs-banyuls.fr (SB); hector.escriva@obs-banyuls.fr (HE)

CrossMark
click for updates

OPEN ACCESS

Citation: Bertrand S, Aldea D, Oulion S, Subirana L, de Lera AR, Somorjai I, et al. (2015) Evolution of the Role of RA and FGF Signals in the Control of Somitogenesis in Chordates. PLoS ONE 10(9): e0136587. doi:10.1371/journal.pone.0136587

Editor: Thomas Brand, Heart Science Centre, Imperial College London, UNITED KINGDOM

Received: April 20, 2015

Accepted: August 5, 2015

Published: September 15, 2015

Copyright: © 2015 Bertrand et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper.

Funding: The laboratory of HE is supported by the ANR BLAN 1716 01. IMLS´s laboratory is currently supported by MASTS (Marine Alliance for Science and Technology Scotland).

Competing Interests: The authors have declared that no competing interests exist.

Abstract

During vertebrate development, the paraxial mesoderm becomes segmented, forming somites that will give rise to dermis, axial skeleton and skeletal muscles. Although recently challenged, the "clock and wavefront" model for somitogenesis explains how interactions between several cell-cell communication pathways, including the FGF, RA, Wnt and Notch signals, control the formation of these bilateral symmetric blocks. In the cephalochordate amphioxus, which belongs to the chordate phylum together with tunicates and vertebrates, the dorsal paraxial mesendoderm also periodically forms somites, although this process is asymmetric and extends along the whole body. It has been previously shown that the formation of the most anterior somites in amphioxus is dependent upon FGF signalling. However, the signals controlling somitogenesis during posterior elongation in amphioxus are still unknown. Here we show that, contrary to vertebrates, RA and FGF signals act independently during posterior elongation and that they are not mandatory for posterior somites to form. Moreover, we show that RA is not able to buffer the left/right asymmetry machinery that is controlled through the asymmetric expression of Nodal pathway actors. Our results give new insights into the evolution of the somitogenesis process in chordates. They suggest that RA and FGF pathways have acquired specific functions in the control of somitogenesis in vertebrates. We propose that the "clock and wavefront" system was selected specifically in vertebrates in parallel to the development of more complex somite-derived structures but that it was not required for somitogenesis in the ancestor of chordates.

Introduction

Segmentation along the antero-posterior body axis is a morphological feature found in several metazoan lineages. In vertebrates, segmentation is conspicuous in the paraxial mesoderm, which forms transient bilateral symmetric blocks during the somitogenesis process [1]. The

somites, which will give rise to the axial skeleton, the skeletal muscles of the trunk and part of the dermis, form in an antero-posterior succession through segmental epithelialisation of the mesenchymal presomitic mesoderm (PSM). During elongation of the vertebrate embryo, a pool of proliferating cells that are continuously added to the caudal zone is maintained in the most posterior part, the tailbud [2]. Although the precise structure of vertebrate tailbuds varies from one species to another, vertebrates share a similar anatomy and the global mechanisms controlling posterior elongation and somitogenesis seem to be conserved. An elegant paradigm was first proposed by Cooke and Zeeman in 1976 to explain the regular formation of segments during somitogenesis termed the "clock and wavefront" model [3]. Molecular evidence for this hypothesis came more than twenty years later and our understanding of how somitogenesis is controlled in vertebrates has been highly improved since then [4, 5]. Our current understanding of the "clock and wavefront" model relies on the specific interactions of several signalling pathways, including the retinoic acid (RA), the Fibroblast Growth Factor (FGF), the Wnt (Wingless/INT-1) and the Notch pathways. These interactions permit the synchronized activation of segmentation genes in the PSM in response to the "segmentation clock" [5]. This clock is defined by periodic waves of expression of genes of the FGF, Wnt, and Notch signalling pathways that are travelling along the PSM [6]. The position of the "wavefront" or "determination front" is defined by the posterior FGF/Wnt pathway which is antagonized by the RA pathway in the rostral region of the PSM [7]. As a consequence of the interaction between the clock and the wavefront, the cells of the PSM that pass the determination border during one oscillation of the clock define a pre-patterned somite [5].

Vertebrates, together with their sister group the tunicates and cephalochordates (i.e. amphioxus), form the chordate superphylum [8, 9]. They share morphological features considered synapomorphies of this clade. Particularly they show, at least transiently during embryonic development, a notochord localized ventral to a dorsal hollow nerve tube. Chordates are also characterized by segmented muscles present on both sides of the main body axis. In tunicates, these muscles are only found in the tail of the tadpole are not formed through the antero-posterior successive segmentation of an unsegmented paraxial mesoderm, but develop directly from muscle cells that are produced early during development and get subsequently rearranged on both sides of the tail midline [10]. In the cephalochordate amphioxus, segmented muscles are derived from somites that form in an anterior-to-posterior sequence, from the most rostral part of the embryo to the most caudal part. However, in contrast to vertebrates, the segmented muscles show a clear asymmetry with the left muscle fibres more anterior than the right ones. The most anterior early-arising somites appear as bilateral pairs by means of enterocoelic evagination of the paraxial dorsal wall of the archenteron, whereas the posterior somites form from the tailbud by schizocoely alternatively on the left and right sides of the embryo [11]. In spite of the morphological differences between amphioxus and vertebrate somitogenesis processes, developing amphioxus somites express homologs of many genes involved at each step of vertebrate paraxial mesoderm segmentation [12]. The only functional evidence of how somitogenesis is controlled in amphioxus came from the analysis of the role of the FGF signalling pathway during embryogenesis [13]. We showed that the formation of the most anterior somites is under the control of FGF whereas posterior somites continue to form in spite of FGFR inhibition [13]. This result suggests that important differences exist in the control of somitogenesis between vertebrates and cephalochordates. However, we still do not know how the formation of somites is governed during posterior elongation in amphioxus and what mechanisms underly the difference in the control of the formation of the most anterior and posterior somites.

In this work, we address these two questions. We show that although *Hox* genes control antero-posterior patterning in amphioxus, and their most anterior limit of expression coincides

Results

RA/Hox1 and FGF pathways act independently during the formation of the anterior somites

We previously showed that inhibiting the FGF signalling pathway at the blastula stage induces a specific loss of the most anterior somites in amphioxus embryos [13]. Strikingly, the anterior paraxial mesoderm region in which somites do not form corresponds to the region where *Hox* genes are not expressed. Thus, we wondered if the necessity for FGF signalling is linked to the anterior limit of *Hox1* expression [22, 24]. To answer this question we treated embryos with RA or BMS009 (an antagonist of the retinoic acid receptor from amphioxus [25]) at the blastula stage and fixed them when the pharynx starts to enlarge (stage L1, 40 h.p.f. at 19°C). As expected, these treatments allowed us to move the anterior border of *Hox1* expression forward or backward, respectively (Fig 1A–1C), whereas the formation of anterior somites is not affected (Fig 1G–1I). We also inhibited the FGF signalling pathway with SU5402 (an inhibitor of FGF receptor [26]) and undertook double treatments with RA and SU5402 or with BMS009 and SU5402. When embryos are treated with both RA and SU5402, the anterior limit of *Hox1* expression is shifted anteriorly compared to embryos only treated with SU5402 (Fig 1D, 1E and 1M). However, the region without somites is comparable to what is observed in embryos in which only FGF signalling is inhibited as highlighted by *Myosin Light Chain-alkali* (*MLC*) expression (Fig 1J, 1K and 1M). Indeed, in the anterior region, we still observe *MLC* expression but it is restricted to the notochord which in SU5402-treated embryos is not straight as in wild-type animals. Moreover, in embryos treated with BMS009 and SU5402, we observe that the anterior limit of *Hox1* expression is shifted posteriorly (Fig 1C, 1F and 1M) but the region without somites is still comparable to what we observe in SU5402 treated embryos (Fig 1J, 1L and 1M). Altogether our data suggest that the anterior border of *Hox1* expression does not define the limit between somites that depend upon FGF signalling for their formation and somites whose formation is not controlled by FGF. Moreover, these data show that there is no cross-talk between RA and FGF signalling for the formation of the most anterior somites in amphioxus.

FGF and RA signalling pathways act independently during posterior elongation

To test whether the RA signalling pathway is involved in the posterior elongation process, we interfered with this signal at the late neurula stage N3 (27 h.p.f. at 19°C), when 8 to 9 somite pairs are formed. We treated embryos with RA or BMS009. Embryos were fixed at stage L0 and at the larval stage L3 (60 h.p.f. at 19°C) and we analyzed the expression of several marker genes to assess the phenotype of the embryos. The activation of RA signalling induces a reduction of the pharynx size, whereas BMS009 treatment leads to an enlargement of the pharyngeal region as observed when embryos are treated at earlier stages. The formation of new somites is normal with embryos at stage L0 having 11 to 12 somites expressing *MyoD related factor 1 (MRF1)* similarly to control animals (Fig 2A–2C). However, *Xlox* (Fig 2D–2F) and *Tbx6/VegT* (Fig 2G–2I) gene expression show that the antero-posterior patterning is affected in an opposite way by the two treatments in both the endoderm and the neural tube. Indeed, we observe that the expression fields of *Xlox* and *Tbx6/16* are pushed posteriorly in RA treated embryos whereas the posterior domain of expression is pushed anteriorly in BMS009 treated embryos, showing that even at late developmental stages RA signalling is still controlling AP patterning.

In vertebrates, the position of the "wavefront" in the "clock and wavefront" model for somite formation is controlled by an opposition between RA signalling coming from the most

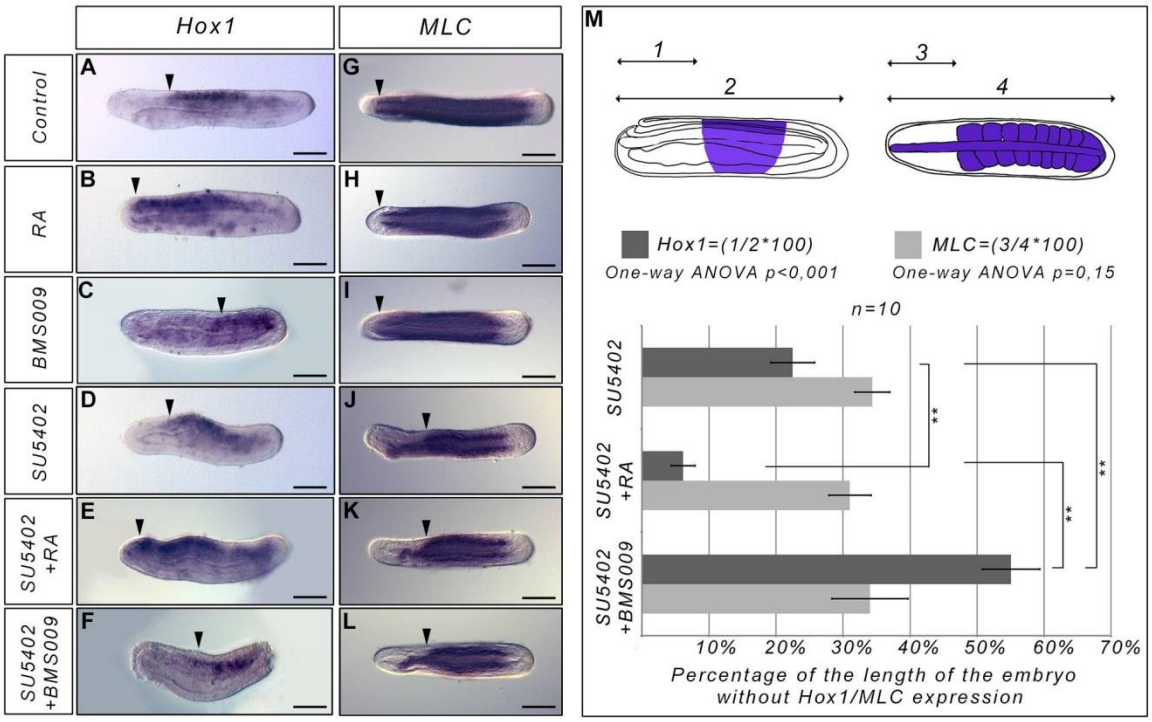


Fig 1. The anterior limit of *Hox1* expression does not define a functional boundary between anterior FGF-sensitive and posterior FGF-insensitive somites. Expression of *Hox1* (lateral views, anterior to the left) and *MLC* (dorsal views, anterior to the left) at the L0 stage in control embryos (A, G), in RA treated embryos (B, H), in BMS009 treated embryos (C, I), in SU5402 treated embryos (D, J), in SU5402 and RA treated embryos (E, K) and in SU5402 and BMS009 treated embryos (F, L). All the treatments were performed at the blastula stage. The arrowheads indicate the anterior limit of Hox1 expression (A-F) or the anterior limit of the embryonic region with formed somites (G-L). (M) Graph presenting the percentage of the length of the embryo without *Hox1* expression (dark grey) or without somites in the anterior region (light grey). The schematic embryos show how this percentage was calculated. (1) corresponds to the length of the anterior region without Hox1 expression, (2) corresponds to the total length of the embryo. (3) corresponds to the length of the anterior region without somite MLC expression. (4) corresponds to the total length of the embryo. Dark blue regions in the schematic embryos correspond to the territories expressing *Hox1* (lateral view, anterior to the left) or *MLC* (dorsal view, anterior to the left) in SU5402 treated embryos. A one-way ANOVA analysis was undertaken and the result indicates that the means of the region without somites is not significantly different between the three treatments (SU5402, SU5402+RA and SU5402+BMS009) whereas the means of the region without *Hox1* expression is significantly different between the three treatment conditions (SU5402, SU5402+RA and SU5402+BMS009). **P<0.003 (corrected p-value); Two samples Student t-test, n = 10 embryos. Error bars indicate s.e.m. Scale bars = 50μm.

doi:10.1371/journal.pone.0136587.g001

posterior somites, and FGF signalling coming from the tailbud. To test whether we could observe any cross-talk between these pathways in amphioxus during posterior elongation, we looked at the expression of FGF and RA target genes when we interfere with these pathways at the N3 stage. We looked at the expression of *ER81/Erm/Pea3*, *Sprouty*, and *FGFR* as targets of the FGF pathway. After SU5402 treatment, we observe a complete loss of *ER81/Erm/Pea3* expression (Fig 3A and 3B) and a partial loss of *Sprouty* expression (Fig 3E and 3F). Indeed, *Sprouty* is still expressed in the notochord whereas the endodermal expression is lost after treatment (Fig 3E and 3F). On the other hand, following FGFR inhibition, *FGFR* expression is similar to what is observed in control animals (Fig 3I and 3J). When embryos are treated with RA or BMS009, the expression of all these genes is similar to what is observed in wild-type embryos, except in the pharyngeal region which is enlarged in BMS009 treated embryos and

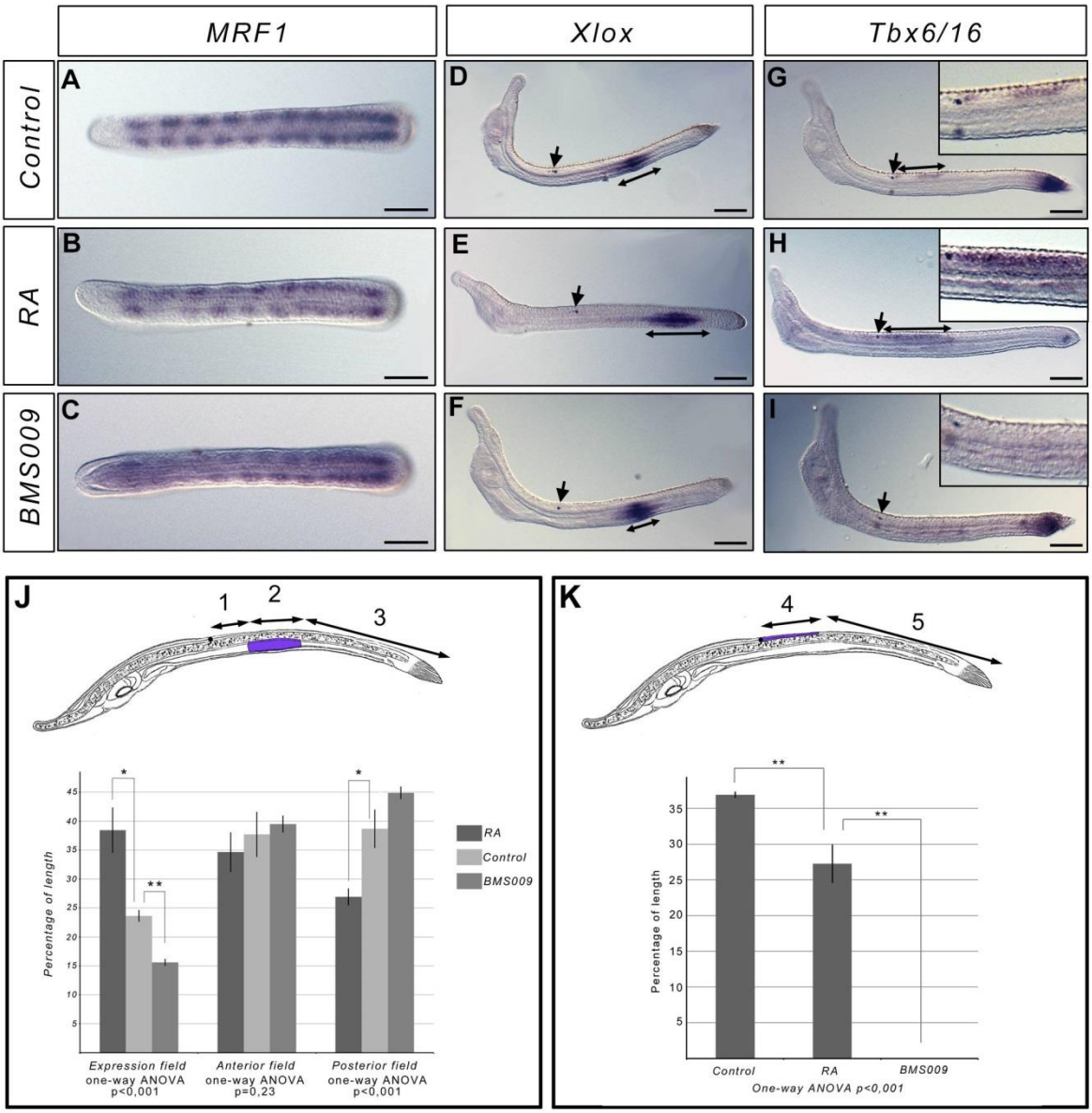


Fig 2. Interfering with RA signalling during posterior elongation does not affect somitogenesis. Expression of *MRF1* in L0 stage control embryos (A), and in embryos treated at the N3 stage with RA (B) or BMS009 (C) (dorsal views, anterior to the left). The expression shows that the number of formed somites is identical in treated and control embryos. Expression of *Xlox* and *Tbx6/16* in L3 stage control embryos (D, G), and in embryos treated at the N3 stage with RA (E, H) or BMS009 (F, I) (side views, anterior to the left). The arrows indicate the position of the pigment spot. The double arrow lines indicate the size of the domain expressing *Xlox* or *Tbx6/16*. Enlargement of the photograph at the level of the pigment spot is presented for *Tbx6/16 in situ* hybridization on the top left of the panels. Scale bars = 50μm. Morphometric analysis of the expression domains of *Xlox* (J) and *Tbx6/16* (K). Schematic larva with the domain of expression highlighted in blue-violet are presented (side view, anetrior to the left). (1) corresponds to the length of the embryo, posterior to the pigment spot, without *Xlox* expression. (2) corresponds to the length of the embryo with *Xlox* expression. (3) corresponds to the length of the posterior field of the embryo without *Xlox* expression. The percentage of the length of the field with *Xlox* expression was calculated as 2/(1+2+3)*100, the percentage of length of the anterior field without expression as 1/(1+2+3)*100 and the percentage of length of the posterior field without expression as 3/(1+2+3)*100 (J). One-way ANOVA analysis indicates that the the means of the percentage of length of the field with *Xlox* expression between the three conditions (control embryos, RA-treated embryos and BMS009-treated embryos) are significantly different as well as the means of the percentage of length of the posterior field without expression of *Xlox* (J). (4) corresponds to the length of the field posterior to the pigment spot showing *Tbx6/16* expression. (5) correspond to the

length of the posterior field of the embryo without Tbx6/16 expression. The percentage of length with *Tbx6/16* expression was calculated as 4/(4+5)*100 (K). One-way ANOVA analysis indicates that the the means of the percentage of the length with *Tbx6/16* expression are significantly different between the three conditions (control embryos, RA-treated embryos and BMS009-treated embryos). **P<0.005 (corrected p-value); *P<0,025 (corrected p-value); t-test, n = 3 embryos. Error bars indicate s.e.m.

doi:10.1371/journal.pone.0136587.g002

reduced in RA treated neurulae (Fig 3C, 3D, 3G, 3H, 3K and 3L). We then looked at the expression of two targets of the RA signalling pathway, the *ParaHox* genes *Cdx* and *Xlox* [27]. We

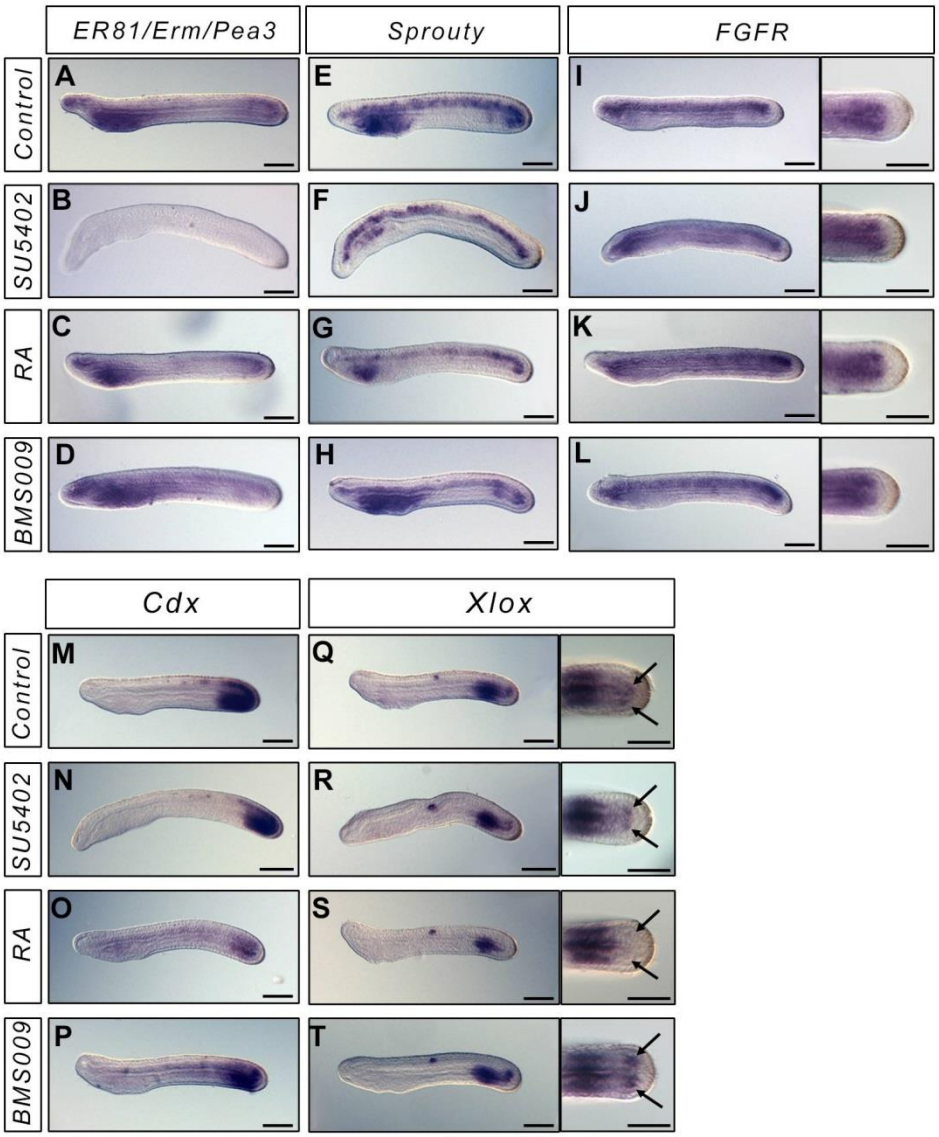


Fig 3. FGF and RA signals do not cross-talk during posterior elongation. Expression of the FGF signalling pathway genes *ER81/Erm/Pea3*, *Sprouty*, *FGFR*, and of the ParaHox genes *Cdx* and *Xlox* in L1 control embryos (A, E, I, M, Q), and in embryos treated at the N3 stage with SU5402 (B, F, J, N, R), RA (C, G, K, O, S) or BMS (D, H, L, P, T) (lateral views, anterior to the left). For *FGFR* and *Xlox*, dorsal views of the tailbud region are also presented. Scale bars = 50μm.

doi:10.1371/journal.pone.0136587.g003

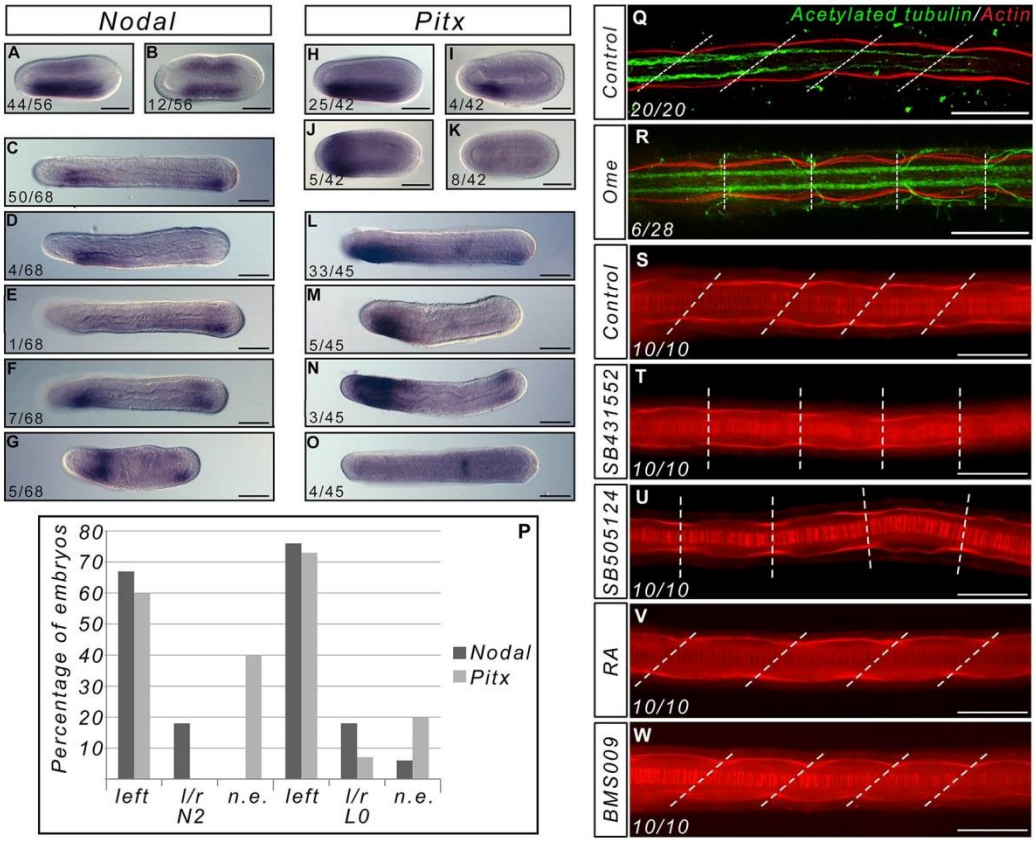


Fig 4. The asymmetry of somites depends on asymmetric *Nodal* expression and is not altered by RA signal pertubations. Expression of *Nodal* in Omeprazole treated embryos at the N2 (A, B) and L0 (C-G) stages (dorsal views, anterior to the left). In wild type embryos expression of *Nodal* is similar to (A) and (C), and expression of *Pitx* is similar to (H) and (L). Expression of *Pitx* in Omeprazole treated embryos at the N2 (H-K) and L0 (L-O) stages (dorsal views, anterior to the left). The number of embryos showing these patterns of expression is indicated on each panel. (P) Graph presenting the number of embryos at the N2 and L0 stage presenting expression in the posterior mesoderm of *Nodal* or *Pitx* only on the left side, on both sides, or on neither side (l/r = left/right, n. e. = not expressed) (P). (Q-R) Texas Red X-Phalloidin (red) and anti-acetylated tubulin (green) labelled control L3 larvae (Q) and Omeprazole treated larvae showing symmetric somites (R). The number of larvae showing symmetric somites after Omeprazole treatment is indicated. Texas Red X-Phalloidin labelling in control (S), SB431552-treated (T), SB505124-treated (U), RA-treated (V) and BMS-treated (W) L3 stage larvae. The pictures represent dorsal views, with anterior to the left, focused on the central region of the larva. The dotted lines join the boundaries between somites on the right and left sides. Scale bars = 50μm.

doi:10.1371/journal.pone.0136587.g004

SB431552, we obtained larvae with symmetric somites (Fig 4T and 4U) and pharynges [29]. On the other hand, we show that following RA or BMS009 treatments, even when starting treatments as early as the blastula stage, somites are still asymmetric in the larvae, similarly to the control animals (Fig 4V and 4W). Interestingly, the pharynx of RA treated embryos is absent as previously described, showing that the asymmetry of the somites is not the result of mechanical constraints imposed by pharynx morphology.

Discussion

A/P patterning and FGF dependency of somite formation

Hox genes are known to regulate patterning of embryonic structures along the antero-posterior axis in bilaterians [35]. Moreover, the most anterior limit of *Hox* gene expression in

amphioxus, which is the anterior limit of *Hox1* expression [22, 24], corresponds to the limit between somites whose formation is dependent upon FGF signalling and those whose formation does not depend upon this signal [13]. To test whether *Hox1* plays a role in defining a functional boundary between the most anterior and the posterior somites, we shifted its expression anteriorly or posteriorly by interfering with the RA signalling pathway. We show that moving the anterior limit of *Hox1* expression has no effect on the size of the territory in which somites are absent when the FGF signal is inhibited. This means that even if the Hox code might be controlling the antero-posterior identity of the somites in amphioxus, it is not defining their dependency upon an FGF signal. Moreover our data suggest that Hox/RA and FGF signals have independent functions during early embryogenesis in amphioxus.

The role of RA, FGF, and their negative cross regulation in the control of somitogenesis

In the present view of the "clock and wavefront" model for somitogenesis [3], the boundary between the newly formed somites and the presomitic mesoderm (PSM) is defined by the posterior FGF signal which is antogonized by the RA signal coming from the formed somites and anterior PSM [7, 36]. The FGF signal in the tailbud is thought to be necessary to maintain a pool of undifferentiated cells, whereas RA is promoting the differentiation of PSM cells. FGF and RA signals show opposite gradients and are mutually inhibiting each other. Indeed, FGF8 activates the expression of *Cyp26*, which codes for the enzyme degrading RA, and inhibits the expression of *Raldh2*, the gene coding for the enzyme responsible for the synthesis of RA [7]. On the other hand, RA can restrict the expression of *fgf8* in chicken [7] and mouse [37, 38], or activates the expression of *MKP3*, which codes for a phosphatase that blocks the MAPK cascade activation in *Xenopus* [39]. This negative crosstalk between FGF and RA is not only important during somitogenesis, but has been recruited in many developmental processes such as the antero-posterior patterning of the heart field [40], the timing of emigration of trunk neural crest cells [41], or limb induction [42]. We show that, in amphioxus, the RA signalling pathway is not implicated in somitogenesis, neither through interaction with the FGF signal for anterior somite formation, nor during posterior elongation. Indeed, activating or inhibiting the RA pathway at early stages or during posterior elongation leads to normal number and shape of formed somites. This is a major difference with what is observed in all vertebrates studied thus far. Indeed, RA depletion induces the formation of smaller somites in quail [43], chicken [7], mouse [44], and *Xenopus* [39], whereas RA signalling activation induces caudal truncation in mouse [45] and zebrafish [46], as well as abnormal somite size and disorganized somite boundaries in *Xenopus* [39].

In this work we also show that the cross regulation between the FGF and RA signalling pathways is absent in amphioxus. This raises questions about how and when the involvement of both pathways in somitogenesis evolved in the chordate clade, as well as of crosstalk between them. In tunicates, the sister group of vertebrates, the role of RA during development seems to be different in the different lineages. Ascidians like *Ciona intestinalis* possess a unique retinoic acid receptor (RAR); upon activation of the RA signalling pathway, *Hox1* expression shifts anteriorly in the epidermis and nervous system [47, 48] suggesting a conserved role for RA in controlling the expression of Hox genes in chordates. However, the larvacean *Oikopleura dioica* possesses no RAR [49] and no homeotic transformation is observed when embryos are treated with RA, suggesting that the formation of the chordate body plan can be achieved without the contribution of a classical RA signalling pathway [49]. In any case, although tunicate tadpoles have segmented muscles, these are not formed through a process comparable to somitogenesis [10]. Moreover, there is no evidence for a putative role of RA in the formation of

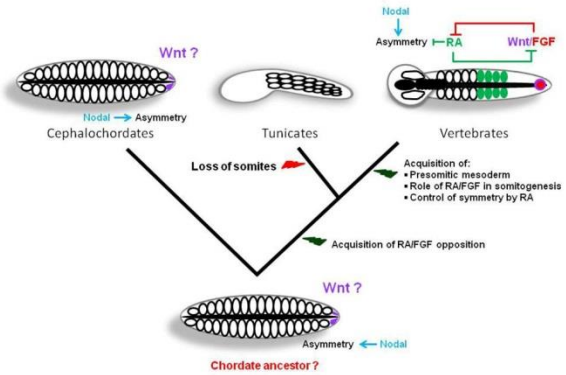


Fig 5. Evolutionary scenario for somitogenesis in chordates. Evolutionary relationships among the three chordate clades are presented, as well as a schematic view of the morphology of embryos of the putative ancestor of chordates, cephalochordates, tunicates and vertebrates (all dorsal views except tunicates for which a lateral view is schematized) during posterior elongation. We propose that the ancestral chordate embryo was morphologically close to amphioxus and that the asymmetry of somite formation was under the control of Nodal. After the divergence of cephalochordates, an opposition between the RA and FGF pathways was acquired. In tunicates, the somitogenesis process was lost, probably as an adaptation to a reduced number of embryonic cells. In the vertebrate lineage, the opposition between RA and FGF gained importance in parallel to the acquisition of the PSM as an intermediate zone between both signals. The recruitment of RA in the control of somitogenesis permitted the acquisition of symmetry through the buffering of the left/right machinery controlled by Nodal. The Wnt pathway, indicated with a question mark, is a good candidate as a signal ancestrally controlling posterior elongation in chordates although up to now there are no functional data supporting this hypothesis.

doi:10.1371/journal.pone.0136587.g005

somites, and the tailbud for the schizocoelic somites). Moreover, during formation of the posterior somites, there is no intermediate tissue between the tailbud and the last formed somites in cephalochordates (i.e. there is no PSM). Finally, somites are asymmetric in amphioxus whereas they form symmetrically in vertebrates.

Interestingly, the PSM defines a territory where RA and FGF gradients are negatively regulating each other, and we show in this study that such opposition is absent in amphioxus. With respect to the role of RA, we show that, in contrast to vertebrates, this signal is not implicated at all during the somitogenesis process in amphioxus: it does not interact with FGF, and it is not able to buffer the lateralizing signal of the left/right machinery. Considering the fact that a negative cross-regulation between RA and FGF was described in the tunicate *Ciona intestinalis* [52], and hypothesising that the control of somitogenesis in the ancestor of chordates was likely similar to that in amphioxus, we can propose a scenario for the evolution of somitogenesis in chordates (Fig 5). The ancestor of olfactores acquired a functional cross-talk between RA and FGF that gained greater significance during embryonic development in the vertebrate lineage, probably due to the acquisition of three RARs *versus* one, and of four FGFRs and twenty-two ligands *versus* a unique FGFR and seven to eight ligands in non-vertebrate chordates [58]. In vertebrates, the RA signal was recruited in the control of somitogenesis together with FGF in parallel to the formation of an intermediate structure between both signals, the PSM. Moreover, the implication of RA in somitogenesis also permitted the acquisition of the symmetrical character of this process through the buffering of the left/right signal. The main signal controlling posterior elongation in the chordate ancestor is however still unknown. The expression of several genes coding for actors of the Wnt signalling pathway in the tailbud during amphioxus posterior elongation [11, 12, 59–62], together with data in vertebrates, make this signal a good candidate.

Several lines of evidence support the hypothesis that the ancestor of chordates had amphioxus-like somitogenesis. From paleontological data it has been suggested that the ancestor of chordates was segmented from the most anterior to the most posterior part of the body like amphioxus [63]. Although there is up to now no data suggesting that these segments were asymmetric, the fact that in amphioxus the asymmetry of somitogenesis is controlled by the Nodal signalling pathway and that this signal is also controlling left/right asymmetries in other bilaterians [53], combined with evidence that the default state in vertebrates is asymmetry [33, 34, 37] support this proposition. Our scenario also implies that at least part of the "clock and wavefront" system for the control of somite formation was acquired in the vertebrate lineage. Recent experiments in quail have shown that non-somitic mesoderm is able to form synchronously several somites without any external cues except deprivation of BMP signal [64]. The authors suggest that the "clock and wavefront" system only serves to control the timing of somite formation and their rostro-caudal patterning. In amphioxus, we show here that there is no wavefront comparable to the one of vertebrates. Indeed, modification of FGF signalling does not alter posterior somitogenesis. Moreover, the RA pathway is not acting as an opposing signal to the FGF pathway. Furthermore, even if segmentation genes are expressed during somitogenesis [12] there is up to now no evidence for a clock. If our hypothesis is true, the recruitment of RA and FGF in the control of somitogenesis was key to the evolution of vertebrate morphology. Indeed, it allowed the acquisition of an important plasticity in the formation of the posterior part of the embryo [65] and also probably the appearance of lateral locomotor structures like fins or limbs, since such morphological traits would probably not have been selected during evolution in animals with asymmetric somites.

Acknowledgments

We would like to thank Jordi Garcia Fernandez and Juan Pascual-Anaya for providing the plasmid to prepare the RNA probe for the *Hox1 in situ* hybridization and Gwenaël Piganeau for helpful comments. The laboratory of HE is supported by the ANR BLAN 1716 01. IMLS.´s laboratory is currently supported by MASTS (Marine Alliance for Science and Technology Scotland).

Author Contributions

Conceived and designed the experiments: SB IS HE. Performed the experiments: SB DA SO LS IS. Analyzed the data: SB DA SO LS IS HE. Contributed reagents/materials/analysis tools: ARdL. Wrote the paper: SB IS HE.

References

1. Dequeant ML, Pourquie O. Segmental patterning of the vertebrate embryonic axis. Nat Rev Genet. 2008; 9(5):370–82. PMID: 18414404. doi: 10.1038/nrg2320
2. Wilson V, Olivera-Martinez I, Storey KG. Stem cells, signals and vertebrate body axis extension. Development. 2009; 136(10):1591–604. PMID: 19395637. doi: 10.1242/dev.021246
3. Cooke J, Zeeman EC. A clock and wavefront model for control of the number of repeated structures during animal morphogenesis. J Theor Biol. 1976; 58(2):455–76. PMID: 940335.
4. Aulehla A, Herrmann BG. Segmentation in vertebrates: clock and gradient finally joined. Genes Dev. 2004; 18(17):2060–7. PMID: 15342488.
5. Hubaud A, Pourquie O. Signalling dynamics in vertebrate segmentation. Nat Rev Mol Cell Biol. 2014; 15(11):709–21. PMID: 25335437. doi: 10.1038/nrm3891
6. Pourquie O. Vertebrate segmentation: from cyclic gene networks to scoliosis. Cell. 2011; 145(5):650–63. PMID: 21620133. doi: 10.1016/j.cell.2011.05.011

7. Diez del Corral R, Olivera-Martinez I, Goriely A, Gale E, Maden M, Storey K. Opposing FGF and retinoid pathways control ventral neural pattern, neuronal differentiation, and segmentation during body axis extension. Neuron. 2003; 40(1):65–79. PMID: 14527434.

8. Delsuc F, Brinkmann H, Chourrout D, Philippe H. Tunicates and not cephalochordates are the closest living relatives of vertebrates. Nature. 2006; 439(7079):965–8. PMID: 16495997.

9. Satoh N, Rokhsar D, Nishikawa T. Chordate evolution and the three-phylum system. Proc Biol Sci. 2014; 281(1794):20141729. PMID: 25232138. doi: 10.1098/rspb.2014.1729

10. Passamaneck YJ, Hadjantonakis AK, Di Gregorio A. Dynamic and polarized muscle cell behaviors accompany tail morphogenesis in the ascidian Ciona intestinalis. PLoS One. 2007; 2(8):e714. PMID: 17684560.

11. Schubert M, Holland LZ, Stokes MD, Holland ND. Three amphioxus Wnt genes (AmphiWnt3, AmphiWnt5, and AmphiWnt6) associated with the tail bud: the evolution of somitogenesis in chordates. Dev Biol. 2001; 240(1):262–73. PMID: 11784062.

12. Beaster-Jones L, Kaltenbach SL, Koop D, Yuan S, Chastain R, Holland LZ. Expression of somite segmentation genes in amphioxus: a clock without a wavefront? Dev Genes Evol. 2008; 218(11–12):599–611. PMID: 18949486. doi: 10.1007/s00427-008-0257-5

13. Bertrand S, Camasses A, Somorjai I, Belgacem MR, Chabrol O, Escande ML, et al. Amphioxus FGF signaling predicts the acquisition of vertebrate morphological traits. Proc Natl Acad Sci U S A. 2011; 108(22):9160–5. PMID: 21571634. doi: 10.1073/pnas.1014235108

14. Schubert M, Yu JK, Holland ND, Escriva H, Laudet V, Holland LZ. Retinoic acid signaling acts via Hox1 to establish the posterior limit of the pharynx in the chordate amphioxus. Development. 2005; 132 (1):61–73. PMID: 15576409.

15. Schubert M, Holland ND, Laudet V, Holland LZ. A retinoic acid-Hox hierarchy controls both anterior/posterior patterning and neuronal specification in the developing central nervous system of the cephalochordate amphioxus. Dev Biol. 2006; 296(1):190–202. PMID: 16750825.

16. Fuentes M, Benito E, Bertrand S, Paris M, Mignardot A, Godoy L, et al. Insights into spawning behavior and development of the European amphioxus (Branchiostoma lanceolatum). J Exp Zool B Mol Dev Evol. 2007; 308(4):484–93. PMID: 17520703.

17. Fuentes M, Schubert M, Dalfo D, Candiani S, Benito E, Gardenyes J, et al. Preliminary observations on the spawning conditions of the European amphioxus (Branchiostoma lanceolatum) in captivity. J Exp Zool B Mol Dev Evol. 2004; 302(4):384–91. PMID: 15287102.

18. Hirakow R, Kajita N. Electron microscopic study of the development of amphioxus, Branchiostoma belcheri tsingtauense: The gastrula. Journal of Morphology. 1991; 207(1):37–52.

19. Hirakow R, Kajita N. Electron microscopic study of the development of amphioxus, Branchiostoma belcheri tsingtauense: the neurula and larva. Kaibogaku Zasshi. 1994; 69(1):1–13. PMID: 8178614.

20. Holland LZ, Holland PWH, Holland ND. Revealing homologies between body parts of distantly related animals by in situ hybridization to developmental genes: Amphioxus vs. vertebrates. In: Ferraris JD, Palumbi S, editors. *Molecular Zoology*: *Advances*, *Strategies*, *and Protocols*. New York: Wiley-Liss; 1996.

21. Somorjai I, Bertrand S, Camasses A, Haguenauer A, Escriva H. Evidence for stasis and not genetic piracy in developmental expression patterns of Branchiostoma lanceolatum and Branchiostoma floridae, two amphioxus species that have evolved independently over the course of 200 Myr. Dev Genes Evol. 2008; 218(11–12):703–13. PMID: 18843503. doi: 10.1007/s00427-008-0256-6

22. Pascual-Anaya J, Adachi N, Alvarez S, Kuratani S, D'Aniello S, Garcia-Fernandez J. Broken colinearity of the amphioxus Hox cluster. Evodevo. 2012; 3(1):28. PMID: 23198682. doi: 10.1186/2041-9139-3-28

23. Wu HR, Chen YT, Su YH, Luo YJ, Holland LZ, Yu JK. Asymmetric localization of germline markers Vasa and Nanos during early development in the amphioxus Branchiostoma floridae. Dev Biol. 2011; 353(1):147–59. PMID: 21354126. doi: 10.1016/j.ydbio.2011.02.014

24. Holland PW, Holland LZ, Williams NA, Holland ND. An amphioxus homeobox gene: sequence conservation, spatial expression during development and insights into vertebrate evolution. Development. 1992; 116(3):653–61. PMID: 1363226.

25. Escriva H, Holland ND, Gronemeyer H, Laudet V, Holland LZ. The retinoic acid signaling pathway regulates anterior/posterior patterning in the nerve cord and pharynx of amphioxus, a chordate lacking neural crest. Development. 2002; 129(12):2905–16. PMID: 12050138.

26. Mohammadi M, McMahon G, Sun L, Tang C, Hirth P, Yeh BK, et al. Structures of the tyrosine kinase domain of fibroblast growth factor receptor in complex with inhibitors. Science. 1997; 276(5314):955–60. PMID: 9139660.

27. Osborne PW, Benoit G, Laudet V, Schubert M, Ferrier DE. Differential regulation of ParaHox genes by retinoic acid in the invertebrate chordate amphioxus (Branchiostoma floridae). Dev Biol. 2009; 327 (1):252–62. PMID: 19103191. doi: 10.1016/j.ydbio.2008.11.027

28. Yu JK, Holland LZ, Holland ND. An amphioxus nodal gene (AmphiNodal) with early symmetrical expression in the organizer and mesoderm and later asymmetrical expression associated with left-right axis formation. Evol Dev. 2002; 4(6):418–25. PMID: 12492142.

29. Soukup V, Yong LW, Lu TM, Huang SW, Kozmik Z, Yu JK. The Nodal signaling pathway controls left-right asymmetric development in amphioxus. Evodevo. 2015; 6:5. PMID: 25954501. doi: 10.1186/2041-9139-6-5

30. Morii M, Takeguchi N. Different biochemical modes of action of two irreversible H+,K(+)-ATPase inhibitors, omeprazole and E3810. J Biol Chem. 1993; 268(29):21553–9. PMID: 8408006.

31. Levin M, Thorlin T, Robinson KR, Nogi T, Mercola M. Asymmetries in H+/K+-ATPase and cell membrane potentials comprise a very early step in left-right patterning. Cell. 2002; 111(1):77–89. PMID: 12372302.

32. Shimeld SM, Levin M. Evidence for the regulation of left-right asymmetry in Ciona intestinalis by ion flux. Dev Dyn. 2006; 235(6):1543–53. PMID: 16586445.

33. Vermot J, Pourquie O. Retinoic acid coordinates somitogenesis and left-right patterning in vertebrate embryos. Nature. 2005; 435(7039):215–20. PMID: 15889094.

34. Kawakami Y, Raya A, Raya RM, Rodriguez-Esteban C, Izpisua Belmonte JC. Retinoic acid signalling links left-right asymmetric patterning and bilaterally symmetric somitogenesis in the zebrafish embryo. Nature. 2005; 435(7039):165–71. PMID: 15889082.

35. Pearson JC, Lemons D, McGinnis W. Modulating Hox gene functions during animal body patterning. Nat Rev Genet. 2005; 6(12):893–904. PMID: 16341070.

36. Dubrulle J, McGrew MJ, Pourquie O. FGF signaling controls somite boundary position and regulates segmentation clock control of spatiotemporal Hox gene activation. Cell. 2001; 106(2):219–32. PMID: 11511349.

37. Vermot J, Gallego Llamas J, Fraulob V, Niederreither K, Chambon P, Dolle P. Retinoic acid controls the bilateral symmetry of somite formation in the mouse embryo. Science. 2005; 308(5721):563–6. PMID: 15731404.

38. Kumar S, Duester G. Retinoic acid controls body axis extension by directly repressing Fgf8 transcription. Development. 2014; 141(15):2972–7. PMID: 25053430. doi: 10.1242/dev.112367

39. Moreno TA, Kintner C. Regulation of segmental patterning by retinoic acid signaling during Xenopus somitogenesis. Dev Cell. 2004; 6(2):205–18. PMID: 14960275.

40. Sirbu IO, Zhao X, Duester G. Retinoic acid controls heart anteroposterior patterning by down-regulating Isl1 through the Fgf8 pathway. Dev Dyn. 2008; 237(6):1627–35. PMID: 18498088. doi: 10.1002/dvdy.21570

41. Martinez-Morales PL, Diez del Corral R, Olivera-Martinez I, Quiroga AC, Das RM, Barbas JA, et al. FGF and retinoic acid activity gradients control the timing of neural crest cell emigration in the trunk. J Cell Biol. 2011; 194(3):489–503. PMID: 21807879. doi: 10.1083/jcb.201011077

42. Zhao X, Sirbu IO, Mic FA, Molotkova N, Molotkov A, Kumar S, et al. Retinoic acid promotes limb induction through effects on body axis extension but is unnecessary for limb patterning. Curr Biol. 2009; 19 (12):1050–7. PMID: 19464179. doi: 10.1016/j.cub.2009.04.059

43. Maden M, Graham A, Zile M, Gale E. Abnormalities of somite development in the absence of retinoic acid. Int J Dev Biol. 2000; 44(1):151–9. PMID: 10761860.

44. Niederreither K, Subbarayan V, Dolle P, Chambon P. Embryonic retinoic acid synthesis is essential for early mouse post-implantation development. Nat Genet. 1999; 21(4):444–8. PMID: 10192400.

45. Sakai Y, Meno C, Fujii H, Nishino J, Shiratori H, Saijoh Y, et al. The retinoic acid-inactivating enzyme CYP26 is essential for establishing an uneven distribution of retinoic acid along the anterio-posterior axis within the mouse embryo. Genes Dev. 2001; 15(2):213–25. PMID: 11157777.

46. Martin BL, Kimelman D. Brachyury establishes the embryonic mesodermal progenitor niche. Genes Dev. 2010; 24(24):2778–83. PMID: 21159819. doi: 10.1101/gad.1962910

47. Nagatomo K, Fujiwara S. Expression of Raldh2, Cyp26 and Hox-1 in normal and retinoic acid-treated Ciona intestinalis embryos. Gene Expr Patterns. 2003; 3(3):273–7. PMID: 12799071.

48. Nagatomo K, Ishibashi T, Satou Y, Satoh N, Fujiwara S. Retinoic acid affects gene expression and morphogenesis without upregulating the retinoic acid receptor in the ascidian Ciona intestinalis. Mech Dev. 2003; 120(3):363–72. PMID: 12591605.

49. Canestro C, Postlethwait JH. Development of a chordate anterior-posterior axis without classical retinoic acid signaling. Dev Biol. 2007; 305(2):522–38. PMID: 17397819.

50. Davidson B, Shi W, Beh J, Christiaen L, Levine M. FGF signaling delineates the cardiac progenitor field in the simple chordate, Ciona intestinalis. Genes Dev. 2006; 20(19):2728–38. PMID: 17015434.

51. Imai KS, Levine M, Satoh N, Satou Y. Regulatory blueprint for a chordate embryo. Science. 2006; 312 (5777):1183–7. PMID: 16728634.

52. Pasini A, Manenti R, Rothbacher U, Lemaire P. Antagonizing retinoic acid and FGF/MAPK pathways control posterior body patterning in the invertebrate chordate Ciona intestinalis. PLoS One. 2013; 7(9): e46193. PMID: 23049976.

53. Namigai EK, Kenny NJ, Shimeld SM. Right across the tree of life: the evolution of left-right asymmetry in the Bilateria. Genesis. 2014; 52(6):458–70. PMID: 24510729. doi: 10.1002/dvg.22748

54. Watanabe H, Schmidt HA, Kuhn A, Hoger SK, Kocagoz Y, Laumann-Lipp N, et al. Nodal signalling determines biradial asymmetry in Hydra. Nature. 2014; 515(7525):112–5. PMID: 25156256. doi: 10.1038/nature13666

55. Boorman CJ, Shimeld SM. Pitx homeobox genes in Ciona and amphioxus show left-right asymmetry is a conserved chordate character and define the ascidian adenohypophysis. Evol Dev. 2002; 4(5):354–65. PMID: 12356265.

56. Le Petillon Y, Oulion S, Escande ML, Escriva H, Bertrand S. Identification and expression analysis of BMP signaling inhibitors genes of the DAN family in amphioxus. Gene Expr Patterns. 2013; 13(8):377–83. PMID: 23872339. doi: 10.1016/j.gep.2013.07.005

57. Onai T, Yu JK, Blitz IL, Cho KW, Holland LZ. Opposing Nodal/Vg1 and BMP signals mediate axial patterning in embryos of the basal chordate amphioxus. Dev Biol. 2010; 344(1):377–89. PMID: 20488174. doi: 10.1016/j.ydbio.2010.05.016

58. Oulion S, Bertrand S, Escriva H. Evolution of the FGF Gene Family. Int J Evol Biol. 2012; 2012:298147. PMID: 22919541. doi: 10.1155/2012/298147

59. Lin HC, Holland LZ, Holland ND. Expression of the AmphiTcf gene in amphioxus: insights into the evolution of the TCF/LEF gene family during vertebrate evolution. Dev Dyn. 2006; 235(12):3396–403. PMID: 17013891.

60. Qian G, Li G, Chen X, Wang Y. Characterization and embryonic expression of four amphioxus Frizzled genes with important functions during early embryogenesis. Gene Expr Patterns. 2013; 13(8):445–53. PMID: 24012522. doi: 10.1016/j.gep.2013.08.003

61. Holland LZ, Holland NN, Schubert M. Developmental expression of AmphiWnt1, an amphioxus gene in the Wnt1/wingless subfamily. Dev Genes Evol. 2000; 210(10):522–4. PMID: 11180802.

62. Schubert M, Holland LZ, Holland ND. Characterization of an amphioxus wnt gene, AmphiWnt11, with possible roles in myogenesis and tail outgrowth. Genesis. 2000; 27(1):1–5. PMID: 10862149.

63. Morris SC, Caron JB. Pikaia gracilens Walcott, a stem-group chordate from the Middle Cambrian of British Columbia. Biol Rev Camb Philos Soc. 2012; 87(2):480–512. PMID: 22385518. doi: 10.1111/j.1469-185X.2012.00220.x

64. Dias AS, de Almeida I, Belmonte JM, Glazier JA, Stern CD. Somites without a clock. Science. 2014; 343(6172):791–5. PMID: 24407478. doi: 10.1126/science.1247575

65. Keyte AL, Smith KK. Heterochrony and developmental timing mechanisms: changing ontogenies in evolution. Semin Cell Dev Biol. 2014; 34:99–107. PMID: 24994599. doi: 10.1016/j.semcdb.2014.06.015

3. Introduction Article 2

Next-generation sequencing (NGS) technology has improved largely the studies in different fields of biology. Specially, this technology has facilitated the use of non-model organisms in the Evo-Devo field (da Fonseca et al., 2016). In this second paper we took advantage of this new technology to shed light on how FGF signaling controls the formation of the anterior somites in amphioxus. We previously showed that treatment with the inhibitor of the FGF receptor (SU5402) at the blastula stage (5hpf) (early treatment) results in the loss of the most anterior somites. However, embryos treated at the late gastrula stage (14hpf) (late treatment) possess all the somites although they form incorrectly (Bertrand et al., 2011). Thus, to uncover the genes that are under the FGF control for the formation of the anterior somites, we performed a comparative RNA-seq analysis between RNAs extracted from early and late treated embryos. A list of candidate genes controlling somitogenesis was hence obtained and we confirmed their putative role in anterior somitogenesis by undertaking *in situ* hybridization.

From the list of genes that were specifically downregulated after early treatment and were expressed in the paraxial mesoderm, we chose several transcription factors for undertaking functional analyses. As knock-down techniques are not available in *B. lanceolatum*, we decided to construct constitutive activator or repressor forms of these transcription factors. To do so, we fused the repressor domain Engrailed or the activator domain VP16 with the DNA-binding domain of the different transcription factors, as previously described (Mayor et al., 2000). The microinjection of the mRNA coding for these chimeras into the unfertilized eggs of amphioxus was performed as described by Yu et al. (Yu et al., 2004) and then after the fertilization, phenotypes were analyzed.

All the obtained data indicate that the FGF signal seems to act through the Ets family factor *ER81/Erm/Pea3* to control anterior somitogenesis. Moreover, *Six1/2* would be the master gene controlling anterior somites formation whereas *Pax3/7* would be implicated in the control of the formation of the posterior ones. These results have an important value in the context of the existing debate about the evolution of the head of vertebrates. Indeed, amphioxus anterior and posterior somitogenesis are under the control of genes that are orthologues of those controlling trunk myogenesis in vertebrates. Moreover, orthologues of the genes controlling head mesoderm myogenesis in vertebrates (*Tbx1* and *Pitx2*) are not expressed in amphioxus at the good time and place, suggesting that vertebrates secondarily recruited these genes for the control of head myogenesis. All this work is presented as a manuscript and additional data and general conclusions of my work will be treated in the last chapter of the thesis.

3.1 Anterior somitogenesis in amphioxus sheds lights on the origin of the vertebrates' head

Title: Anterior somitogenesis in amphioxus sheds light on the origin of the vertebrates' head

Daniel Aldea[1], Lucie Subirana[1], Celine Keime[2], Jose Luis Gomez-Skarmeta[3], Sylvain Marcellini[4], Stephanie Bertrand[1a], Hector Escriva[1a]

1 UPMC Univ Paris 06, UMR 7232, BIOM, Observatoire Océanologique de Banyuls sur Mer, F-66650, Banyuls/Mer, France
2 IGBMC (Institut de Génétique et de Biologie Moléculaire et Cellulaire), INSERM, U964, CNRS, UMR7104, Université de Strasbourg, 67404 Illkirch, France.
3 Centro Andaluz de Biología del Desarrollo (CABD), Consejo Superior de Investigaciones Científicas/Universidad Pablo de Olavide, Seville, Spain
4 Laboratory of Development and Evolution, Department of Cell Biology, Faculty of Biological Sciences, University of Concepcion, Concepción, Chile.

a To whom correspondence should be addressed at: Laboratoire Arago, Avenue du Fontaulé, 66650, Banyuls-sur-Mer, France. E-mail: stephanie.bertrand@obs-banyuls.fr, hescriva@obs-banyuls.fr. Phone: +33(0)468887390, Fax: +33(0)468887393

Abstract

The appearance of new body structures during animal evolution is a central question in Evolutionary Developmental Biology (Evo-Devo). Particularly, the origin of the vertebrates' head has attracted the attention of researchers for many years. In this context, one striking question concerns the evolution of the unsegmented mesoderm of the vertebrates' head from a completely segmented amphioxus-like ancestor. It is extensively assumed that paraxial mesoderm of the hypothetical chordate ancestor, as that of extant cephalochordates, was segmented from its most anterior to its most posterior part. Interestingly, it has been shown that formation of the anterior-most somites in the cephalochordate amphioxus is controlled by the FGF signaling, suggesting that functional evolution of this signaling pathway during early development in the vertebrate ancestor could have played a key role in the loss of mesoderm segmentation in the vertebrate's head.

Here, using a comparative RNA-seq approach, we looked for "putative genes", controlled by the FGF signal during early development in amphioxus, playing a role in the control of anterior somitogenesis. Then, we functionally tested whether some of these genes directly control anterior somitogenesis in amphioxus. Our results show a functional compartmentalization of the paraxial mesoderm in amphioxus, where *Six1/2* and the FGF effector *ER81/Erm/PEA3* are major players in the anterior somitogenesis in amphioxus while *Pax3/7* controls posterior somitogenesis. This regulatory cascade in amphioxus resembles that for the control of trunk somitogenesis in vertebrates and diverges from the gene cascades controlling the formation of the vertebrate head muscles. Our results strengthen the hypothesis that changes in the FGF function during early development were instrumental for the loss of anterior somites, releasing developmental constraints in the anterior part of the embryo and allowing a secondary acquisition of head muscles in the ancestor of vertebrates.

Introduction

The vertebrates' head appeared more than 500 million years ago. This body structure is a sophisticated novelty, and its appearance is associated with the transition from a filter-feeding to a predator life-style. Indeed, the head concentrates sensory organs and neural structures that coordinate movements by integrating sensorial stimuli to efficiently capture preys with specialized mandibular structures. In the last decades, embryological studies helped to unravel the developmental origin of the different components of the head. Thus, it is known that most of these structures derive from the neural crest cells (NCCs), the neurogenic placodes and the cranial mesoderm (CM) (Diogo et al., 2015; Le Douarin and Dupin, 2012; Schlosser, 2008, 2015). While some studies suggest an early origin of the neural crest cells and placodes in non-vertebrate chordates (Bronner, 2015; Schlosser, 2015), the evolutionary origin of the cranial mesoderm remains poorly understood.

An important difference between the head and the trunk mesoderm in vertebrates is that the trunk is segmented into transitional epithelial somites during embryonic development, while segmentation of the CM is still debated. Historically, various authors considered the vertebrates' head mesoderm as segmented (revised in (Holland et al., 2008b)) similar to the anterior paraxial mesoderm of amphioxus. Indeed, although the exact anatomy of the chordate ancestor is unknown, it is extensively assumed that it possessed a completely segmented body from the most anterior to the most posterior part, similar to extant cephalochordates. Particularly, the presence of head cavities in elasmobranch (rays and shark) embryos has been used as a strong argument for the segmentation of head mesoderm in vertebrates (Balfour, 1876). This led Goodrich to propose that mesoderm, branchial arches and cranial nerves were integrated into single metameres (Goodrich, 1918; Goodrich, 1930). This hypothesis led some authors to study early divergent vertebrates such as lampreys as a proxy for the intermediate organism between amphioxus and gnathostomes. The first studies in lampreys also described head cavities and suggested they were homologues of shark head cavities (Damas, 1944; Koltzoff, 1902). However, later studies have shown that real head cavities are absent in lampreys and regional segmentation in the lamprey's head may be induced by adjacent structures rather than being the result of a real segmentation (Kuratani et al., 1999). Thus head cavities would represent a gnathostome synapomorphy (revised in (Kuratani, 2008)).

Molecular data in chick embryos show that CM is formed after two waves of expression of the segmental gene *hairy* (Jouve et al., 2002), indicating that the CM is formed at best by

two segments. Thus, the first wave would correspond to the prechordal cranial mesoderm (PCM) that gives rise to the extra-ocular muscles (EOM) and the second one would correspond to the entire cranial paraxial mesoderm (CPM) that gives rise to the masticatory and facial expression muscles (Diogo et al., 2015; Sambasivan et al., 2011). However, besides the diverse embryological origins of putative head segments (if any) in vertebrates and expression of segmental genes, muscle formation in the head and muscle formation in the trunk are controlled by different master genes in vertebrates. Therefore, *Pitx2* plays a crucial role in the formation of the head muscles derived from both the prechordal cranial mesoderm (PCM) and the cranial paraxial mesoderm (CPM), whereas *Tbx1* play a pivotal role in the formation of the muscles derived from the CPM (Sambasivan et al., 2011). Conversely, *Pax3* plays a critical role in the formation of truncal muscles and is expressed in the forming somites of vertebrates (Buckingham and Relaix, 2007; Schubert et al., 2001a). Moreover, members of the SIX family together with their cofactors EYA are involved in the formation of truncal muscles in vertebrates (Grifone et al., 2007; Grifone et al., 2005).

The phylum Chordata comprises the vertebrates, urochordates (or tunicates) and cephalochordates (Bertrand and Escriva, 2011). In the sister group of vertebrates, the urochordates, somites have been secondarily lost, whereas in the cephalochordate amphioxus somites are present all along the antero-posterior axis of the body. These somites in amphioxus give rise to the muscles of the body, to the ventral mesoderm and to the dermis and connective tissues (Mansfield et al., 2015). In amphioxus, the anterior 8-10 somites pinch off from dorsolateral furrows in the wall of the archenteron by enterocoely and then posterior somites are added directly one at a time from the epithelium of the taildbud by schizocoely (Beaster-Jones et al., 2008; Mansfield et al., 2015). Remarkably, it has been shown that the FGF signal plays a pivotal role in the formation of the three most anterior somites of amphioxus (Bertrand et al., 2011). This result together with some gene expression patterns indicate that the body of amphioxus could be divided into three different compartments, (i) the three most anterior enterocoelic, FGF-sensitive, *Mox*-negative, and *Engrailed*-positive somites; (ii) the posterior enterocoelic, FGF-insensitive, *Mox*- and *Engrailed*-positive somites; and (iii) the posterior schizocoelic, FGF-insensitive, *Engrailed*-negative, and *Mox*-, *Axin*-, *Lcx*-, and *Paraxis*-positive somites. Despite these differences between amphioxus and vertebrates, conservation of gene expression between cephalochordates and gnathostomes has been used as a support for the head segmentation hypothesis in vertebrates (Holland et al.,

2008b), particularly for *Gbx* and anterior Hox or *Tbx15/18/22*, *Six1/2*, *Tbx1/10* and *Delta*. However, most of these genes are expressed posterior to the putative homologues of head mesoderm segments (e.g. Hox), or late during embryonic development, after the three most anterior somites have formed (e.g. Tbx1/10).

In this work we hypothesized that, if the vertebrate head mesoderm is homologue to the three most anterior somites of amphioxus, the control of myogenesis should be conserved. Thus, we functionally studied the master genes controlling anterior somitogenesis in amphioxus which are downstream of the FGF signal. We first performed a comparative RNA-seq analysis revealing several transcription factors whose expression is controlled by FGF only during the formation of the most anterior somites in amphioxus, and whose vertebrate orthologues are involved in trunk myogenesis. Functional analyses using constitutive repressors of these transcription factors, such as the FGF target *ER81/Erm/PEA3* or *Six1/2*, show that they are pivotal for the formation of the most anterior somites in amphioxus. Conversely, the transcription factor *Pax3/7* shows a central role in the formation of posterior somites. Overall, our work reveals that the body of amphioxus could be divided in two ancient compartments, where the most anterior FGF-sensitive somites are *Six1/2* dependent whereas the posterior FGF-insensitive somites are *Pax3/7* dependent. These results show that head muscles in vertebrates are formed using gene regulatory networks that are not implicated in anterior somitogenesis in amphioxus, and that anterior muscles in amphioxus are formed using similar gene regulatory networks as trunk muscles in vertebrates. Altogether our results allow us to suggest that the vertebrate ancestor lost its anterior somites through changes in the FGF function during early development, and that muscles of the head were secondarily gained during evolution through the recruitment of other transcription factors.

Material and Methods

Animals, embryo collection, drug treatments

Ripe adults from the Mediterranean amphioxus species (*Branchiostoma lanceolatum*) were collected at the Racou beach near Argelès-sur-Mer, France, (latitude 42° 32' 53" N and longitude 3° 03' 27" E) with a specific permission delivered by the Prefect of Region Provence Alpes Côte d'Azur. *Branchiostoma lanceolatum* is not a protected species. Gametes were collected by heat stimulation as previously described (Fuentes et al., 2007; Fuentes et al., 2004). Prior to pharmacological treatments, and before hatching, embryos were transferred to new Petri dishes with a known final volume of sea water. SU5402 (Calbiochem 572631) was dissolved in dimethylsulfoxide (DMSO) at 10^{-2}M and added to cultures of embryos at a final concentration of 25 µM at the blastula stage (5 hours post fertilization (hpf) at 19°C) or at the gastrula stage (15,5hpf at 19°C). Control embryos were raised simultaneously with equivalent concentrations of DMSO in filtered sea water. Embryos were either fixed in PFA4%-MOPS as previously described (Holland et al., 1996) or frozen in liquid nitrogen.

Sequencing and RNA-seq analysis

Total RNA was extracted from embryos 3, 6 or 9 hours post treatment (hpt) as described in supplementary figure 1 using the RNeasy Plus Mini Kit (QIAGEN) after disrupting and homogenizing the sample with the TissueLyser (QIAGEN). Samples were sequenced using Illumina (1x54 bases) technology at the "Plateforme BIOPUCES et SEQUENÇAGE", IGBMC, Illkirch, France. Transcripts were mapped onto the reference transcriptome (Oulion et al., 2012a) with bwa v0.6.1 (Li and Durbin, 2010), and following the subsequent set of parameters; a seed of 27 bases, with a maximum of 2 differences in the seed, 4 mismatched allowed, a maximum of 4 gap extensions and default parameters for all others. The read counts were normalized across libraries with the method proposed by Anders and Huber and implemented in DESeq v1.10.1. (Anders and Huber, 2010). Resulting p-values were adjusted for multiple testing by using the Benjamini and Hochberg (1995) method (Benjamini and Hochberg, 1995). Contigs were then clustered by temporal expression profile using the STEM software (Ernst and Bar-Joseph, 2006). A False Discovery Rate (FRD) <0,05 and Log2 Fold Change > 1 was chosen for analysis of significantly differentially expressed contigs between SU5402-treated and control embryos in early and late treatments.

Cloning and in situ hybridization

For *B. lanceolatum* genes not previously published, sequences were recovered from its reference transcriptome (Oulion et al., 2012a) by TBLASTN using sequences from *B. floridae* as queries. Specific primers were then designed for RT-PCR amplification of partial coding regions. We also designed specific primers to amplify the 3'UTR regions of *ER81/Erm/PEA3*, *Six1/2* and *Pax3/7*. The complete list of primers used is presented in Table 2. A mix of total RNA of *B. lanceolatum* extracted from embryos at different developmental stages was used as a template for retro-transcription. Amplification was performed using Advantage 2 Polymerase kit (Clontech) and a touch-down PCR program with annealing temperature ranging from 65 to 40°C. Amplified fragments were cloned using the pGEM-T Easy system (Promega) and sub-cloned in pBluescript II KS+ for probe synthesis. All the clones were confirmed by Sanger sequencing. For GFP, probe was synthesized from a pcDNA3-spacer-GFP-NX plasmid (gift from Angela Nieto and Jose Manuel Mingot). Whole mount *in situ* hybridizations were performed as described in Somorjai et al. (2008) (Somorjai et al., 2008).

Constructs, in vitro mRNA synthesis and microinjection

All the vectors were constructed using the pCS2+ expression vector backbone. Constitutive activator forms of *Pax3/7* (VP16-Pax3/7), *Six1/2* (VP16-Six1/2) and *ER81/Erm/Pea3* (VP16-EEP) were created by fusing the coding sequence of the 81 aa activation domain of VP16 protein (Friedman et al., 1988) to the N-terminal side of the DNA binding domain coding sequence of *Pax3/7* or *ER81/Erm/Pea3* and to the N-terminal side of the full-length coding sequence of *Six1/2*. Constitutive repressor forms of *Pax3/7* (Eng-Pax3/7), *Six1/2* (Eng-Six1/2) and *ER81/Erm/Pea3* (Eng-EEP) were created by fusing the coding sequence of the repressor domain of the engrailed protein (Jaynes and O'Farrell, 1991) to the N-terminal side of the DNA binding domain coding sequence of *Pax3/7* or *ER81/Erm/Pea3* and to the N-terminal side of the full-length sequence of *Six1/2*. All the constructs were confirmed by Sanger sequencing. The vectors were linearized and *in vitro* transcription was performed using the mMESSAGE mMACHINE SP6 Transcription Kit (AM1340). Transgenesis vector was kindly provided by Sylvain Marcellini laboratory. We changed the minimal promotor of this vector for one harboring the *B. lanceolatum* minimal promotor of beta-actin protein as was described by Feng et al. in the Chinese specie *B. belcheri* (Feng et al., 2014). We then cloned our putative enhancers from genomic DNA of *B.*

lanceolatum using the following primers Forward: 5'-CTTGTACACGGGGTCCTCTC-3' and Reverse: 5'-GGAGACAAACCGCTCTCTTG-3'. The sequence was confirmed by Sanger sequencing. Microinjections of plasmids and mRNA were carried out as described in Yu et al. (Yu et al., 2004).

.

Results

Comparative RNA-seq analyses reveal candidate genes downstream of the FGF signal during anterior somitogenesis

It has been previously shown that inhibition of the FGF signal during early development in amphioxus induces the loss of the most anterior somites (i.e. treatment with the FGFR inhibitor SU5402 at blastula stage (5 hpf at 19°C)). However, embryos treated with SU5402 at late gastrula stage (15,5 hpf at 19°C) do not lose the most anterior somites. These results indicate a specific role of the FGF signaling pathway during early development for the formation of the most anterior somites (Bertrand et al., 2011).

In order to understand how the FGF signal controls the formation of these somites, we performed a comparative RNA-seq analysis among transcriptomes of control *versus* treated embryos at blastula (FGF sensitive somitogenesis) and gastrula (FGF non sensitive somitogenesis) stages. Moreover, we undertook a dynamic analysis by sampling embryos at 3, 6 and 9 hours post treatment (hpt) in both cases (Fig. sup. 1). These analyses reveal an important number of contigs which expression is controlled directly or indirectly by the FGF signal at both developmental stages. Thus, in total the expression of 1677 and 2716 contigs were downregulated or upregulated during the early phase of development, respectively (i.e. blastula stage), whereas the expression of 1001 and 478 contigs were downregulated or upregulated during the later phase, respectively (i.e. gastrula) (Fig. 1a).

We took advantage of the dynamic expression profile following the treatments for each contig (see Fig. sup. 1) to select putative master genes in the hierarchy between the FGF signal and the observed phenotype. For this purpose, we clustered the contigs according to their temporal expression profile and we selected contigs that were downregulated at 3hpt after early treatment but not at 6 or 9 hpt nor after late stage treatment (Fig. 1b). A detailed analysis of these contigs reveals that many of them correspond to genes already described as being implicated in amphioxus somitogenesis (Beaster-Jones et al., 2008), but also to orthologues of genes involved in myogenesis or somitogenesis in vertebrates (Table 1). Additionally, we identified genes that were not described previously in amphioxus and that might have a role in somitogenesis such as an orthologue of *Lim domain only protein 4* gene (*LMO4*), which in vertebrates acts as a cofactor of Snail (Ferronha et al., 2013) and which is expressed in somites in mice and zebrafish (Kenny et al., 1998; Lane et al., 2002), or the

orthologue of the *Smad interacting protein 1* gene (*SIP1*), that has been recently shown to be involved in vertebrate somitogenesis (Kok et al., 2010; Maruhashi et al., 2005).

From this list of genes, we choose to concentrate further on orthologues of 24 genes described in vertebrates as having a role in somitogenesis or myogenesis, on genes highly downregulated, and on genes with an unknown role in amphioxus (Table 1). To validate our list of candidates as possible master genes at the top of the cascade downstream of FGF for the formation of the most anterior somites, we performed *in situ* hybridization on both control and treated embryos before the first somites form at gastrula (11 hpf at 19°C) and late gastrula (14 hpf at 19°C) stages. Thus, we broadly observed a downregulation of the expression in the paraxial mesoderm at gastrula and late gastrula stages when embryos were treated with SU5402 at the blastula stage (5hpf) (Fig. 2 and Fig. sup. 2) confirming the data obtained by RNA-seq. As previously described (Bertrand et al., 2011), the effectors of the FGF signaling pathway *Er81/Erm/Pea3* and *Dusp6/7/9* were strongly downregulated after treatment (Fig. 2 a-h). Additionally, the expression of myogenic genes as *Six1/2*, *Six4/5*, *Eya*, *Pax3/7*, *FoxC*, *MRF2* and *Mef2* was lost in the paraxial mesoderm that gives rise to the anterior somites (Fig. 2 i-j'). Interestingly, whereas the expression in the paraxial mesoderm is downregulated for genes such as *Six4/5* or *Eya*, the expression in the axial mesoderm remains unaffected (Fig. sup. 2 p,t), suggesting that different regulatory signals control their expression in axial and paraxial mesoderm. Furthermore, embryos treated with SU5402 lose the anterior stripe of expression of the segmentation genes *Delta*, *HairyB*, *HairyC*, *HairyD*, *Ripply*, *Hey1/2*, *Uncx4.1* and *Tbx15/18/22* (Fig. 2 k'-p''). Moreover, the expression of the transcription factor *Snail* and its putative cofactor *LMO4* was downregulated in the paraxial mesoderm (Fig. 2 q''-x''). Remarkably, the expression domain in the neural plate of *Snail* remains unaltered as previously mentioned (Bertrand et al., 2011) (Fig. sup. 2 r'',t''). Likewise, the expression in the paraxial mesoderm of *Twist1/2* and *SIP1* was downregulated, whereas the expression in the axial mesoderm was unaffected (Fig. 2 y''-f''' and Fig. sup. 2 y''-f'''). Concerning members of the Gli superfamily, *Zic* showed a mild downregulation in the paraxial mesoderm (Fig. 2 g'''-j''') whereas *Gli* expression was highly downregulated in the paraxial mesoderm after treatment (Fig. 2 k'''-n'''). *Frizzled4*, a receptor of the Wnt signaling pathway, was also downregulated in the paraxial mesoderm (Fig. 2 o'''-r'''). For the other 68 genes selected, we failed to observe any specific signal using *in situ* hybridization or they were not expressed in the anterior paraxial mesoderm (Table 1).

The FGF signal acts through ER81/Erm/Pea3 for the formation of the most anterior somites

In our previous work we showed that the FGF signal acts through the Ras/MAPK pathway for the formation of the anterior somites in amphioxus (Bertrand and Escriva, 2011). Indeed, embryos treated at 5 hpf with the inhibitor of the ERK1/2 pathway (U0126) present a similar phenotype to the embryos treated with the inhibitor of the FGF receptor (SU5402) (Bertrand et al., 2011). From the variety of intracellular effectors of the FGF signaling pathway acting in the Ras/MAPK cascade (Wasylyk et al., 1998), our comparative RNA-seq data confirmed our previous finding showing the downregulation of the gene coding for the Ets family transcription factor *ER81/Erm/PEA3* (Bertrand et al., 2011). These results led us to propose *ER81/Erm/PEA3* as the candidate gene linking the FGF signal and the control of anterior somitogenesis. To investigate the specific role of *ER81/Erm/PEA3* in the control of anterior somitogenesis in amphioxus, we generated protein chimeras by fusing the Engrailed repressor domain (Jaynes and O'Farrell, 1991) or the VP16 activation domain (Friedman et al., 1988) to the DNA binding domain (DBD) of *ER81/Erm/PEA3* (Fig. 3a). We then microinjected the mRNA coding for these two chimeras into amphioxus unfertilized eggs. Remarkably, we found that embryos injected with mRNA coding for Engrailed-ER81/Erm/PEA3 (Eng-EEP) lost the most anterior somites (Fig. 3c-d), whereas embryos injected with the mRNA coding for VP16-ER81/Erm/PEA3 did not show any observable phenotype alteration (Fig. sup. 3 a-b) when compared to control embryos at the late neurula stage (Fig. 3 b-c). This result suggests that the FGF signal acts through its nuclear effector *ER81/Erm/PEA3* for the formation of the most anterior somites in amphioxus.

Six1/2 and Pax3/7 play different roles in amphioxus somitogenesis along the anteroposterior axis

Among all the candidate genes highlighted by the transcriptomic approach, two of them called our attention, *Six1/2* and *Pax3/7*, because of their known implication in trunk myogenesis in vertebrates (revised in (Buckingham and Rigby, 2014)). To investigate the putative role of these transcription factors (*Six1/2* and *Pax3/7)* in amphioxus we modified their *in vivo* function through microinjection in unfertilized eggs of mRNAs coding for the following protein chimeras: Engrailed-Six1/2 (Eng-Six1/2), Engrailed-Pax3/7 (Eng-Pax3/7), VP16-Six1/2 (VP16-Six1/2) and VP16-Pax3/7 (VP16-Pax3/7) (Fig. 3a). Embryos injected with mRNAs coding for VP16-Six1/2 or VP16-Pax3/7 did not show any observable phenotypic alteration (Fig. sup. 3 c-f). However, embryos injected with the mRNA coding for

the constitutive repressor Eng-Six1/2 lost the most anterior somites (Fig. 3 e-f) and embryos injected with the mRNA coding for Eng-Pax3/7 were shorter than controls with an observed expression of the muscle gene marker *MLC* (Myosin Light Chain) in the anterior paraxial mesoderm but not in the posterior part of the body. This suggests that these embryos failed to form the posterior somites (Fig. 3 g-h). We then wanted to corroborate this result through the microinjection of both the constitutive repressor forms of Eng-Six1/2 and Eng-Pax3/7. Embryos microinjected with Eng-Six1/2 and Eng-Pax3/7 were shortened and both the most anterior somites and the most posterior somites were not forming. However, the muscle marker gene *MLC* was still expressed in the central somites (Fig. 3 i). Likewise, embryos microinjected with Eng-Pax3/7 and then treated with the FGFR inhibitor SU5402 presented a similar phenotype (Fig. 3 j).

Functional relationship between Six1/2, Pax3/7 and ER81/Erm/PEA3 during anterior somitogenesis in amphioxus

The constitutive repression of *ER81/Erm/PEA3* and *Six1/2* target genes induced the loss of anterior somites whereas the repression of *Pax3/7* target genes induced the loss of posterior somites. Remarkably, these three genes are coexpressed in the paraxial mesoderm during amphioxus embryogenesis (Fig. 4 a-c), and we decided to study the possible hierarchical relationships between them. For this purpose we analyzed the expression of each of these genes at the late gastrula stage when the paraxial mesoderm is already committed to the formation of the most anterior somites and at the late neurula stage when the anterior somites are already formed, in embryos injected with each constitutive repressor chimera. Thus, embryos injected with the constitutive repressor form of *ER81/Erm/PEA3* showed a loss of *Six1/2*, *Pax3/7* and *MRF2* (*Myogenic Regulatory Factor*) expression (Fig. 4 f-h). This phenotype is similar to that obtained in embryos treated with SU5402 (see above). However, embryos injected with the constitutive repressor form of *Six1/2* expressed *ER81/Erm/PEA3* normally (Fig. 4 i), but *Six1/2* (using a probe for the 3'UTR), *Pax3/7* and *MRF2* were downregulated in the paraxial mesoderm next to the invaginating mesoderm (Fig. 4 j-l). Then, embryos injected with the constitutive repressor form of *Pax3/7* showed a normal endogenous expression of *ER81/Erm/PEA3*, *Six1/2*, and *MRF2*, consistent with our previous observations where anterior somites are formed normally in embryos microinjected with the Eng-Pax3/7 mRNA (Fig. 4 m-p).

We then studied the phenotype induced by the different microinjections at late neurula stage, when the anterior-most somites are already formed and posterior somites are added

one-at-time in wild type embryos. Embryos injected with Eng-EEP, which loses the anterior-most somites as previously shown (Fig. 3 c-d), exhibited a loss of expression in the most anterior somites territory of *Six1/2* and *FoxC*, a transcription factor related to myogenesis and mesoderm specification (Aldea et al., 2015; Amin et al., 2010; Andrikou et al., 2013) (Fig. 5 l,n). *Pax3/7* is not expressed in the anterior somites at this developmental stage in wild type embryos. However, it is expressed in the neural tube, and this expression domain was maintained in Eng-EEP injected embryos or in SU5402 treated embryos (Fig. 5 c,h,m). Likewise, embryos injected with Eng-Six1/2 did not show any signal in the most anterior somites using the probes *Six1/2* and *FoxC* (Fig. 5 q, s), whereas the expression of the genes *ER81/Erm/PEA3* and *Pax3/7* (Fig. 5 p, r) seemed not to be altered. On the other hand, when we examined the expression of *ER81/Erm/PEA3*, *Six1/2*, and *FoxC* in embryos injected with Eng-Pax3/7, we detected transcripts of *ER81/Erm/PEA3*, *Six1/2* and *FoxC* in the formed somites (i.e. anterior somites) (Fig. 5 u-x).

As we have shown, inhibition of the FGF signal at the blastula stage inhibits the formation of the anterior-most somites but does not affect the formation of the notochord in amphioxus (Bertrand et al., 2011). Indeed, genes that are normally expressed in both axial and paraxial mesoderm, such as *Twist1/2* or *SIP1*, were downregulated in the paraxial mesoderm in SU5402 treated embryos, whereas the expression in the axial mesoderm remained unaffected (Fig. sup. 2b’’’, f’’’). To test if the notochord was well-formed in embryos injected with the chimeras we used *Brachyury2* as a marker gene (Holland et al., 1995). In a similar way to embryos treated with SU5402, embryos injected with Eng-EEP or with Eng-Six1/2 showed *Brachyury2* (*Bra2*) expression along the antero-posterior axis of the embryo (Fig. 5 e,j,o,t). Remarkably, in Eng-Pax3/7 injected embryos, *Bra2* was very weakly expressed in both the notochord and the posterior tailbud (Fig. 5 y), in agreement with the proposed role of *Pax3/7* in the formation of the notochord and posterior somites in amphioxus (Holland et al., 1999).

ER81/Erm/PEA3 response elements are present in cis-regulatory regions of genes implicated in myogenesis

Gene expression is controlled through *cis*-regulatory elements both in time and space. Moreover, changes in these elements were instrumental in the origin and evolution of morphological novelties in eukaryotes (Acemel et al., 2016; Wray et al., 2003). While discovery of these elements was extremely difficult and laborious in the past, recent advances in high-throughput technologies now allow their easy identification. Thus the Assay for

transposase-Accessible Chromatin with high-throughput sequencing (ATAC-seq) technology allows to map the chromatin accessibility genome-wide, including open chromatin regions and enhancers (Buenrostro et al., 2015). We took advantage of this technique to search for open chromatin regions in the proximities of candidate transcription factors playing a role in somitogenesis during early development. Since we are interested in anterior somitogenesis, we performed ATAC-seq on amphioxus embryos at three different developmental stages: 8hpf (at 19°C, gastrula stage when the paraxial mesoderm is already committed to the formation of anterior somites), 15hpf (at 19°C, late gastrula stage, when segmentation by enterocoely of anterior somites starts) and 36hpf (at 19°C, late neurula stage, when anterior somites are already formed and posterior elongation starts adding somites in the tailbud region by schizocoely). Interestingly, we found a peak that is "turned on" at 8hpf and for which the accessibility to the chromatin is no longer observable at 15hpf and 36hpf (Fig. 6 a). This peak is located in the first intron of the gene *Zic* (one of the downregulated candidate genes (Table 1)). Moreover, the genomic sequence in which this ATAC-seq peak is localized possesses an Ets-family transcription factor DNA binding site. We cloned 1 Kb of genomic DNA harboring this putative enhancer in the 5' region of a construct containing the minimal promotor of amphioxus beta-actin (Feng et al., 2014) upstream of the GFP reporter gene (Fig. 6b) and flanked by the two Tol2 integration sites. Then, we microinjected this construction together with the mRNA coding for the Tol2 transposase into unfertilized amphioxus eggs. After fertilization, embryos were fixed at 10 hpf. To better observe the GFP expression we performed *in situ* hybridization using a RNA probe against GFP. Remarkably, the GFP expression recapitulates the normal expression pattern of *Zic* (Fig. 6 c-f). A similar approach led us to find another genomic region in the proximity of *Six1/2* (Fig. 6 g), where a potential DNA binding site for Ets family transcription factors is also present. This region harbors an ATAC-seq peak at 8hpf and 15hpf. Further analysis with a reporter plasmid displaying this putative sequence must be undertaken to corroborate the function of the putative enhancer *in vivo*.

Discussion

Amphioxus somites are divided into three ontogenetically and functionally different groups

Based on anatomic and functional data, previous studies showed that amphioxus somites can be divided into three different groups: (i) the most anterior enterocoelic, FGF-sensitive, *Mox*-negative, and *Engrailed*-positive somites; (ii) the posterior enterocoelic, FGF-insensitive, *Mox*- and *Engrailed*-positive somites; and (iii) the posterior schizocoelic, FGF-insensitive, *Engrailed*-negative, and *Mox*-, *Axin*-, *Lcx*-, and *Paraxis*-positive somites (Bertrand et al., 2011). Here, we studied the implication of the FGF signal in the control of the most-anterior somites and we confirmed this three-part division. Moreover, we demonstrate that the most anterior somites are FGF signalling dependent for their formation through a gene regulatory cascade in which *ER81/Erm/PEA3* and *Six1/2* are major players, whereas the posterior schizocoelic somites are *Pax3/7* dependent. The question of the regulatory mechanisms controlling the formation of the posterior enterocoelic somites, which are still formed even when *Six1/2* and *Pax3/7* target genes are repressed (see Fig. 3 i), remains an open question.

Anterior mesoderm segmentation and myogenesis are concomitant processes controlled by the FGF signal in amphioxus

The FGF signal plays an essential role in vertebrate somitogenesis during the posterior elongation process as part of the clock and wavefront system (Hubaud and Pourquie, 2014). In amphioxus, we have recently shown that, in contrary to vertebrates, the FGF signal does not play a significant role during posterior elongation and addition of new somites from the tailbud (Bertrand et al., 2015) whereas, it plays a pivotal role for the formation of the most anterior somites (Bertrand et al., 2011). How the FGF signal controls anterior but not posterior somitogenesis in amphioxus was an evident question we tried to answer through a comparative transcriptomic approach. The results obtained pointed out an important amount of genes whose expression is directly or indirectly controlled by the FGF signal specifically during anterior somitogenesis.

In amphioxus, embryonic development is relatively fast, and the first somites can already be observed as early as 17 hpf. However, the paraxial mesoderm is already committed to form the anterior somites at the gastrula stage (11 hpf). Consequently, we observed that most of the genes that emerged as early targets of the FGF signaling pathway during early somitogenesis in our comparative RNA-seq analysis were related with both the presomitic

mesoderm segmentation and myogenesis processes in vertebrates. Thus, we observed that orthologues of genes implicated in vertebrate myogenesis such as *Six1/2*, *Six4/5*, *Eya*, *Pax3/7*, *FoxC*, *MRF2* or *Mef2*, which are expressed before segmentation (Aldea et al., 2015; Kozmik et al., 2007; Schubert et al., 2003), as well as orthologues of genes involved in vertebrate segmentation such as *Delta* or their *Hairy* effectors, *Ripply*, *Hey1/2*, *Uncx4.1* and *Tbx15/18/22*, which are also expressed before segmentation (Beaster-Jones et al., 2008; Minguillon et al., 2003; Rasmussen et al., 2007), were downregulated in the paraxial mesoderm (Fig. 2). Moreover, other orthologues of genes involved in vertebrate somitogenesis such as *Gli*, or genes that have not been described before in amphioxus, with putative roles in somitogenesis, such as *LMO4* or *SIP1*, were also downregulated in the paraxial mesoderm. Interestingly, *Snail* a transcription factor involved in epithelial-to-mesenchymal transition also shows a downregulation in the paraxial mesoderm but not in the neural plate. Taken together, these data confirm the essential role of the FGF signal in the control of anterior somitogenesis during early development in amphioxus. In addition, contrary to vertebrate somitogenesis where segmentation occurs before myogenesis (Bryson-Richardson and Currie, 2008), both processes seem to be coupled during amphioxus anterior somitogenesis.

FGF signaling controls anterior somitogenesis/myogenesis in amphioxus directly through its effector gene ER81/Erm/PEA3

In zebrafish and chick, *Erm* and *Pea3* are the effectors of the FGF signal during somitogenesis (Brent and Tabin, 2004; Raible and Brand, 2001). It has been shown in a previous study (Bertrand et al., 2011), and confirmed by our RNA-seq approach, that the amphioxus orthologue of these genes, *ER81/Erm/PEA3*, is completely downregulated following inhibition of the FGF signal. These data suggested that in amphioxus *ER81/Erm/PEA3* could be the effector of the FGF signal during the control of anterior somitogenesis. To functionally test this hypothesis, we injected the mRNA coding for the constitutive repressor form of *ER81/Erm/PEA3* (Eng-EEP). We observed that the embryos lost the most anterior somites similarly to embryos treated with SU5402 supporting our proposition.

We then used an ATAC-seq approach in order to find putative enhancers that would be active during early development of amphioxus and containing Ets family transcription factor binding sites. We found that amphioxus Zic contains a peak in its first intron with a clear Ets family transcription factor binding site that could represent an active enhancer. Moreover, this

peak is specific of early development since it is only present at 8hpf and disappears in later stages. Thus, we performed transient transgenesis in amphioxus by microinjecting a reporter plasmid in which expression of GFP is controlled by the putative enhancer of Zic. Our results clearly show that this enhancer directs gene expression to the paraxial mesoderm at the gastrula stage. In vertebrates, Zic genes are involved in different processes during embryogenesis such as neuroectodermal development, neural crest induction, somite segmentation and myogenesis (Houtmeyers et al., 2013). Particularly, in mice, *Zic2* and *Zic3* play an essential role in paraxial mesoderm segmentation (Inoue et al., 2007). Moreover, *Zic1* and *Zic2* act together with *Pax3* and *Gli2* to activate *Myf5*, a key transcription factor for myogenesis in mice (Himeda et al., 2013; Pan et al., 2011). Similarly, in ascidians, *Macho-1*, that encodes a transcription factor of the Zic family, plays a central role in muscle-determination (Imai et al., 2004; Sawada et al., 2005). In amphioxus, *Zic* is expressed in the most anterior somites, in the posterior tailbud and during neural development (Gostling and Shimeld, 2003) and our data suggests that it could be implicated in anterior somitogenesis through an activation by the FGF-*ER81/Erm/PEA3* cascade.

Anterior an posterior somitogenesis in amphioxus are controlled by different regulatory logics involving the same players

In vertebrates, SIX1 is essential for myogenesis as highlighted by the impaired somitogenesis phenotype observed in *Six1*$^{-/-}$ mice mutants (Relaix et al., 2013). In addition, it has been shown that SIX1 acts together with SIX4 to activate *Myf5* and participates to the direct transcriptional activation of the key transcription factor gene in myogenesis of vertebrates, *Myod* (Giordani et al., 2007; Grifone et al., 2005; Relaix et al., 2013; Wu et al., 2014). Another key player of myogenesis in vertebrates is PAX3. In mice, *Pax3* is expressed in the presomitic mesoderm and then in all newly formed somites (Schubert et al., 2001a) and *Pax3* mutants show impaired somitogenesis and trunk muscle formation (Tremblay et al., 1998). Interestingly, PAX3 binds together with SIX1/4 and its EYA cofactors to the enhancer that drives the early expression of *Myf5* in the hypaxial somites (Daubas and Buckingham, 2013).

In amphioxus, *Six1/2* expression is first detected at the gastrula stage in the invaginating mesendoderm and in the paraxial mesoderm. Then, starting from early neurula stage, transcripts are detected in all presumptive somites excepting the two most posterior ones (Kozmik et al., 2007). Expression of *Pax3/7* in amphioxus is quite dynamic. Thus, it is first detected at the early gastrula stage in the axial and paraxial mesoderm and later on, at the

neurula stage, in the paraxial mesoderm, in the endoderm, in the lateral edges of the neural plate and in the nascent notochord (Holland et al., 1999). According to their gene expression pattern and their role in vertebrates, it has been proposed that *Six1/2* and *Pax3/7* might play crucial roles in myogenesis and/or somitogenesis in amphioxus (Holland et al., 1999; Kozmik et al., 2007). However, no functional study supporting this proposition has been performed until now.

Here, in the comparative transcriptomic approach, we showed that, when the FGF signal is blocked, both *Six1/2* and *Pax3/7* are highly downregulated, specifically during early development. This result, together with the known role of these genes in vertebrates led us to functionally study their implication during early development in amphioxus. Our experimental approach consisted in the overexpression of constitutive repressor chimeras of each transcription factor in the embryo and the study of the induced phenotype.

Injection of Eng-Six1/2 induces the loss of anterior somites, a similar phenotype as embryos treated with the FGFR inhibitor SU5402, indicating its direct role in the control of anterior somitogenesis in amphioxus. However, injection of Eng-Pax3/7 does not induce the loss of the anterior somites but the embryos do not form the posterior somites. These results show that formation of anterior and posterior somites in amphioxus are controlled differently, with *Six1/2* as a master gene for anterior somitogenesis and *Pax3/7* for posterior somitogenesis.

Moreover, we were able to establish a functional hierarchy among these genes and the FGF effector *ER81/Erm/PEA3*. Thus, repression of target genes of *ER81/Erm/PEA3* (through injection of Eng-EEP induces loss of gene expression in the paraxial mesoderm of amphioxus gastrula for *Six1/2*, *Pax3/7* and *MRF2*, demonstrating the high position of this gene in the cascade controlling anterior somitogenesis. However, while repression of *Six1/2* target genes induces the loss of anterior somites and the loss of *Pax3/7* and *MRF2* expression in the paraxial mesoderm, the expression of *ER81/Erm/PEA3* is not lost, confirming the higher position of *ER81/Erm/PEA3* in the cascade.

Conversely, repression of target genes of *Pax3/7* does not induce the loss of *ER81/Erm/PEA3*, *Six1/2* nor *MRF2* expression in the gastrula but posterior somites fail to form as well as posterior notochord, suggesting that Pax3/7 is the master gene controlling the formation of these two structures (Fig. 3 e-f).

Amphioxus somitogenesis sheds light on the evolution of the vertebrate's head

Skeletal muscle determination and differentiation in vertebrates is driven by members of the myogenic regulatory family (MRF), such as *MyoD*, *Myf5*, or *MRF4*. These genes are implicated in the control of myogenesis both in the head and the trunk. However, their expression is controlled by different upstream regulatory cascades in these two body regions. In the trunk, it has been shown that *Pax3* and members of the SIX family together with their cofactors EYA, are major regulators of myogenesis, while in the head, *Pitx2* plays a crucial role in the formation of the head muscles derived from both the prechordal cranial mesoderm (PCM) and the cranial paraxial mesoderm (CPM), whereas *Tbx1* play a pivotal role in the formation of the muscles derived from the CPM (Sambasivan et al., 2011).

Here, we have shown that amphioxus anterior somitogenesis is controlled by the FGF signal through its effector *ER81/Erm/PEA3,* which in turn controls *Six1/2* expression. Moreover, we have shown a compartmentalization of the amphioxus somites along the anteroposterior axis. Only the most anterior somites are controlled by *Six1/2* and the FGF signal, while posterior somites are controlled by *Pax3/7*, which probably also plays a minor role in anterior somitogenesis since it is expressed in the paraxial mesoderm of the amphioxus gastrulae. Interestingly, *Pitx* and *Tbx1/10* are not expressed during early development of amphioxus when the most anterior somites form. Indeed, *Pitx* is expressed asymmetrically starting at the midneurula stage in somites and endoderm suggesting it is implicated in left/right axis patterning (Boorman and Shimeld, 2002) whereas *Tbx1/10* is expressed in the ventral part of the somites but at late stages, when the anterior somites are already formed (Mahadevan et al., 2004). Moreover, injection of *Tbx1/10* morpholinos does not seem to affect somitogenesis, but mainly lead to the formation of a shorter pharynx (Koop et al., 2014). Altogether these results indicate that muscles of the head of vertebrates are not formed using the same program as anterior muscles in amphioxus.

In the context of the intense debate about the origin of the vertebrate's head, our results are particularly interesting. Indeed, since amphioxus anterior somitogenesis is controlled by a similar regulatory cascade as trunk myogenesis in vertebrates, and major regulators of head myogenesis in vertebrates do not play a similar role in amphioxus, it is tempting to propose the following evolutionary scenario. An amphioxus-like vertebrate ancestor, completely segmented from its most anterior to its most posterior part, formed its anterior somites using a FGF-*ER81/Erm/PEA3*-*Six1/2* gene regulatory cascade and its most posterior somites using a *Pax3/7* regulatory cascade. Then, during evolution, the role of the FGF signal in the control of

anterior somitogenesis changed, inducing the loss of anterior somites and liberating the developmental constraints imposed by them in this part of the body. Secondarily, de novo muscles appeared which formation was regulated by new regulatory cascades (*Pitx2* and *Tbx1*) that were coopted for this purpose.

Acknowledgements

D.A. holds a fellowship from "Becas Chile". S.B. is supported by the Institut Universitaire de France. H.E. laboratory was supported by the Agence National de la Recherche (ANR BLAN 1716 01). We would like to thank Angela Nieto and Jose Manuel Mingot for providing the pcDNA3-spacer-GFP-NX plasmid and Naohito Takatori for providing the Tol2 system vectors.

Figure legends

Figure 1. Comparative RNA-seq analysis. (a) Venn diagram comparing the sets of significantly differentially expressed contigs between SU5402-treated and control embryos in early and late treatments. The total number of upregulated contigs and the total number of downregulated contigs are indicated in red and green, respectively. A False Discovery Rate (FRD) <0,05 and Log2 Fold Change > 1 was chosen for this analysis. (b) Line graph showing the profile of contigs highly downregulated at 3 hpt after early treatment, and which expression is not affected at later stages or after late treatment. Fold change is showed relative to control and expressed in Log2.

Figure 2. *In situ* hybridization of candidate gene in control and SU5402-treated embryos at 11 hpf and 14 hpf of. Dorsal views in all panels**.** Vertebrate orthologues of effectors of the FGF signaling pathway as *ER81/Erm/Pea3* (*EEP*) (a-d) and *Dusp6/7/9* (e-h) are downregulated in treated embryos (b,d,f,h). Orthologues of genes that in vertebrates are involved in myogenesis as *Six1/2* (i-l), *Six4/5* (m-p), *Eya* (q-t), *Pax3/7* (u-x), *FoxC* (y-b'), *MRF2* (c'-f') and *Mef2* (g'-j') are downregulated in the paraxial mesoderm after SU5402 treatment (j,l,n,p,r,t,v,x,z,b',d',f',h',j'). Orthologues of genes that in vertebrates are involved in somite segmentation as *Delta* (k'-n'), *HairyB* (o'-r'), *HairyC* (s'-v'), *HairyD* (w'-z'), *Ripply* (a''-d''), *Hey1/2* (e''-h''), *Uncx4.1* (i''-l''), *Tbx15/18/22* (m''-p'') are expressed in stripes in control embryos whereas this pattern is lost after SU5402 treatment. *Snail* (q''-t'') and the Snail co-factor *LMO4* (u''-x'') expression is lost in the paraxial mesoderm following SU5402 treatment. Similarly for *Twist1/2* (y''-b''') and *Smad interacting protein 1* (*SPI1*) (c'''-f''') is observed a downregulation in the paraxial mesoderm but not in the axial mesoderm. *Zic* is mildly downregulated after treatment (g'''-j'''), whereas *Gli* shows a strong downregulation (k'''-n'''). Finally, *Frizzled4* (*Fz4*) is downregulated in the paraxial mesoderm of treated embryos (o'''-r'''). Scale bar = 100 μm.

Figure 3. Role of *ER81/Erm/Pea3, Six1/2* and *Pax3/7* in amphioxus development. (A) Chimeras were constructed by fusing the Engrailed repressor domain at the N-terminal extremity of the full-length or DNA binding domain of *ER81/Erm/PEA3*, *Six1/2* or *Pax3/7*. **(B)** *In situ* hybridization of MLC (Myosin Light Chain) in control and treated embryos fixed at the late neurula stage. Control embryos showed somites from the most anterior to the most posterior part of the body (a,b). Anterior somites were lost (double head arrows) in Eng-EEP (c,d) and Eng-Six1/2 (e,f) injected embryos whereas in Eng-Pax3/7 injected embryos the posterior somites were absent (g,h). In Eng-Six1/2+Eng-Pax3/7 (i) injected embryos and in Eng-Pax3/7 injected embryos and then treated with SU5402 at blastula stage (j) the most anterior somites as well as the most posterior somites were lost whereas the middle somites (3) were still formed. Lateral views for a, c, e and g. Dorsal views for b, d, f, h, i, and j. Scale bar = 250 μm.

Figure 4. Relationships between *ER81/Erm/PEA3, Six1/2* and *Pax3/7* during anterior somitogenesis. Expression at the late gastrula stage of *ER81/Erm/PEA3*, *Six1/2*, *Pax3/7* and *MRF2* using coding regions or 3'UTR probes in control embryos (a-d), and embryos injected with mRNA of Eng-EEP (e-h), Eng-Six1/2 (i-l) or Eng-Pax3/7 (m-p). Compared with the expression in the paraxial mesoderm of *Six1/2* (arrow) and *Pax3/7* (arrowhead) (b,c), in EEP-Eng injected embryos this expression is clearly downregulated for *Six1/2* (f) and *Pax3/7* (double arrowhead) (g), as well in Six1/2-Eng injected embryos is observed a downregulation of the expression for *Six1/2* (double arrow) (j) and *Pax3/7* (double arrowhead) (k). Dorsal views in all panels. Scale bar = 100 μm.

Figure 5. Expression of *ER81/Erm/PEA3, Six1/2, Pax3/7, FoxC* and *Bra2* at the late neurula stage. (a-e) Expression pattern in control embryos. A loss of *ER81/Erm/PEA3*, *Six1/2* and *FoxC* expression (arrows) in the most anterior somites region was observed in SU5402 treated-embryos (f,g,i) whereas the expression of *Pax3/7* and *Bra2* was similar to control embryos (h,j). The expression of *Six1/2* and *FoxC* was lost in the most anterior somites region in embryos injected with Eng-EEP or Eng-Six1/2 (l,n,q,s) whereas the expression of *Pax3/7* and *Bra2* was not affected (m,o,r,t). In Eng-Pax3/7 injected embryos only the anterior somites are formed and a downregulation of the expression of *Brachyury* is observed in the notochord and tailbud (u-y) observed. Lateral views in all panels and anterior is to the left in all embryos. Scale bar = 250 μm.

Figure 6. Putative cis-regulatory regions implicated in the control of anterior somitogenesis in amphioxus. (a) Genomic landscape of *Zic* and ATAC-seq profiles at 8hpf, 15hpf and 36hpf. The putative regulatory region that was further tested is framed. (b) Construction used to test the Zic regulatory region. (c,d) GFP *in situ* hybridization in embryos injected with the reporter construct. (e, f) Endogenous Zic expression. (c, e) are lateral views. (d, f) are blastoporal views. (g) Genomic landscape of *Six1/2* and ATAC-seq profiles at 8hpf, 15hpf and 36hpf. Putative regulatory region is framed. Scale bar = 100 μm (c-f) for *in situ* hybridization. Scale bar in Kb in each figure for the genomic size.

Table legend

Table 1. Candidate genes studied. Name of each gene, corresponding contig code and Log2 fold change for each timepoint analyzed are presented.

Table 2. Primers for all the genes that were cloned. Name of each gene and forward and reverse sequence for each of these primers.

Supplementary figure legends

Supplementary figure 1. Experimental approach for the comparative RNA-seq analysis. Embryos were treated with SU5402, an inhibitor of the FGF receptor at 5 hpf (at 19°C, blastula stage) and at 15,5 hpf (at 19°C, late gastrula stage). Total RNA extraction was undertaken 3h, 6h and 9h after either early or late treatment, and sequenced using Illumina technology. hpf: hours post fertilization, (E3, E6, E9): total RNA extraction for early treatment, (L3, L6, L9): total RNA extraction for late treatment.

Supplementary figure 2. *In situ* hybridization of downregulated genes at 11 hpf and 14 hpf in control and SU5402-treated embryos. Blastoporal views in all panels**.** Vertebrate orthologues of effectors of the FGF signalling pathway as *ER81/Erm/Pea3* (*EEP*) (a-d) and *Dusp6/7/9* (e-h) are downregulated in treated embryos (b,d,f,h). Orthologues of genes that in vertebrates are involved in myogenesis as *Six1/2* (i-l), *Six4/5* (m-p), *Eya* (q-t), *Pax3/7* (u-x), *FoxC* (y-b'), *MRF2* (c'-f') and *Mef2* (g'-j') are downregulated in the paraxial mesoderm after SU5402 treatment (j,l,n,p,r,t,v,x,z,b',d',f',h',j'). Orthologues of genes that in vertebrates are involved in somite segmentation as *Delta* (k'-n'), *HairyB* (o'-r'), *HairyC* (s'-v'), *HairyD* (w'-z'), *Ripply* (a''-d''), *Hey1/2* (e''-h''), *Uncx4.1* (i''-l''), *Tbx15/18/22* (m''-p'') are expressed in stripes in control embryos whereas this pattern is lost after SU5402 treatment. *Snail* (q''-t'') and the Snail co-factor *LMO4* (u''-x'') expression is lost in the paraxial mesoderm following SU5402 treatment. Similarly for *Twist1/2* (y''-b''') and *Smad interacting protein 1* (*SPI1*) (c'''-f''') is observed a downregulation in the paraxial mesoderm but not in the axial mesoderm. *Zic* is mildly downregulated after treatment (g'''-j'''), whereas *Gli* shows a strong downregulation (k'''-n'''). Finally, *Frizzled4* (*Fz4*) is downregulated in the paraxial mesoderm of treated embryos (o'''-r'''). Scale bar=100 μm.

Supplementary figure 3. Embryos injected with the chimaera VP16-ER81/Erm/PEA3, VP16-Six1/2 and VP16-Pax3/7 did not show any observable alteration. Lateral views in a,c,e. Dorsal views in b,d,g. MLC *in situ* hybridization in all the embryos. VP16-EEP (a-b), VP16-Six1/2 (c-d), VP16-Pax3/7 (e-f). Scale bar = 250 μm.

References

Acemel, R.D., Tena, J.J., Irastorza-Azcarate, I., Marletaz, F., Gomez-Marin, C., de la Calle-Mustienes, E., Bertrand, S., Diaz, S.G., Aldea, D., Aury, J.M.*, et al.* (2016). A single three-dimensional chromatin compartment in amphioxus indicates a stepwise evolution of vertebrate Hox bimodal regulation. Nature genetics *48*, 336-341.

Aldea, D., Leon, A., Bertrand, S., and Escriva, H. (2015). Expression of Fox genes in the cephalochordate Branchiostoma lanceolatum. Front Ecol Evol *3:80*.

Amin, N.M., Shi, H., and Liu, J. (2010). The FoxF/FoxC factor LET-381 directly regulates both cell fate specification and cell differentiation in C. elegans mesoderm development. Development *137*, 1451-1460.

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. Genome biology *11*, R106.

Andrikou, C., Iovene, E., Rizzo, F., Oliveri, P., and Arnone, M.I. (2013). Myogenesis in the sea urchin embryo: the molecular fingerprint of the myoblast precursors. EvoDevo *4*, 33.

Balfour, F.M. (1876). The Development of Elasmobranch Fishes. Journal of anatomy and physiology *10*, 376 372-411.

Beaster-Jones, L., Kaltenbach, S.L., Koop, D., Yuan, S., Chastain, R., and Holland, L.Z. (2008). Expression of somite segmentation genes in amphioxus: a clock without a wavefront? Dev Genes Evol *218*, 599-611.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. . Journal of the Royal Statistical Society *57(B) :289-300*.

Bertrand, S., Aldea, D., Oulion, S., Subirana, L., de Lera, A.R., Somorjai, I., and Escriva, H. (2015). Evolution of the Role of RA and FGF Signals in the Control of Somitogenesis in Chordates. PloS one *10*, e0136587.

Bertrand, S., Camasses, A., Somorjai, I., Belgacem, M.R., Chabrol, O., Escande, M.L., Pontarotti, P., and Escriva, H. (2011). Amphioxus FGF signaling predicts the acquisition of vertebrate morphological traits. Proceedings of the National Academy of Sciences of the United States of America *108*, 9160-9165.

Bertrand, S., and Escriva, H. (2011). Evolutionary crossroads in developmental biology: amphioxus. Development *138*, 4819-4830.

Boorman, C.J., and Shimeld, S.M. (2002). Pitx homeobox genes in Ciona and amphioxus show left-right asymmetry is a conserved chordate character and define the ascidian adenohypophysis. Evolution & development *4*, 354-365.

Brent, A.E., and Tabin, C.J. (2004). FGF acts directly on the somitic tendon progenitors through the Ets transcription factors Pea3 and Erm to regulate scleraxis expression. Development *131*, 3885-3896.

Bronner, M.E. (2015). Evolution: On the crest of becoming vertebrate. Nature *527*, 311-312.

Bryson-Richardson, R.J., and Currie, P.D. (2008). The genetics of vertebrate myogenesis. Nature reviews Genetics *9*, 632-646.

Buckingham, M., and Relaix, F. (2007). The role of Pax genes in the development of tissues and organs: Pax3 and Pax7 regulate muscle progenitor cell functions. Annual review of cell and developmental biology *23*, 645-673.

Buckingham, M., and Rigby, P.W. (2014). Gene regulatory networks and transcriptional mechanisms that control myogenesis. Developmental cell *28*, 225-238.

Buenrostro, J.D., Wu, B., Chang, H.Y., and Greenleaf, W.J. (2015). ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. Current protocols in molecular biology / edited by Frederick M Ausubel [et al] *109*, 21 29 21-29.

Damas, H. (1944). Recherches sur le développement de Lampetra fluviatilis L.: contribution à l'étude de la céphalogenèse des vertébrés. Arch Biol *55:1–248*.

Daubas, P., and Buckingham, M.E. (2013). Direct molecular regulation of the myogenic determination gene Myf5 by Pax3, with modulation by Six1/4 factors, is exemplified by the -111 kb-Myf5 enhancer. Developmental biology *376*, 236-244.
Diogo, R., Kelly, R.G., Christiaen, L., Levine, M., Ziermann, J.M., Molnar, J.L., Noden, D.M., and Tzahor, E. (2015). A new heart for a new head in vertebrate cardiopharyngeal evolution. Nature *520*, 466-473.
Ernst, J., and Bar-Joseph, Z. (2006). STEM: a tool for the analysis of short time series gene expression data. BMC bioinformatics *7*, 191.
Feng, J., Li, G., Liu, X., Wang, J., and Wang, Y.Q. (2014). Functional analysis of the promoter region of amphioxus beta-actin gene: a useful tool for driving gene expression in vivo. Molecular biology reports *41*, 6817-6826.
Ferronha, T., Rabadan, M.A., Gil-Guinon, E., Le Dreau, G., de Torres, C., and Marti, E. (2013). LMO4 is an essential cofactor in the Snail2-mediated epithelial-to-mesenchymal transition of neuroblastoma and neural crest cells. The Journal of neuroscience : the official journal of the Society for Neuroscience *33*, 2773-2783.
Friedman, A.D., Triezenberg, S.J., and McKnight, S.L. (1988). Expression of a truncated viral trans-activator selectively impedes lytic infection by its cognate virus. Nature *335*, 452-454.
Fuentes, M., Benito, E., Bertrand, S., Paris, M., Mignardot, A., Godoy, L., Jimenez-Delgado, S., Oliveri, D., Candiani, S., Hirsinger, E.*, et al.* (2007). Insights into spawning behavior and development of the European amphioxus (Branchiostoma lanceolatum). J Exp Zool B Mol Dev Evol *308*, 484-493.
Fuentes, M., Schubert, M., Dalfo, D., Candiani, S., Benito, E., Gardenyes, J., Godoy, L., Moret, F., Illas, M., Patten, I.*, et al.* (2004). Preliminary observations on the spawning conditions of the European amphioxus (Branchiostoma lanceolatum) in captivity. J Exp Zool B Mol Dev Evol *302*, 384-391.
Giordani, J., Bajard, L., Demignon, J., Daubas, P., Buckingham, M., and Maire, P. (2007). Six proteins regulate the activation of Myf5 expression in embryonic mouse limbs. Proceedings of the National Academy of Sciences of the United States of America *104*, 11310-11315.
Goodrich, E. (1918). On the development of the segments of the head in Scyllium. . Q J Microsc Sci *63:1–30*.
Goodrich, E. (1930). Studies on the Structure and Development of Vertebrates.
Gostling, N.J., and Shimeld, S.M. (2003). Protochordate Zic genes define primitive somite compartments and highlight molecular changes underlying neural crest evolution. Evol Dev *5*, 136-144.
Grifone, R., Demignon, J., Giordani, J., Niro, C., Souil, E., Bertin, F., Laclef, C., Xu, P.X., and Maire, P. (2007). Eya1 and Eya2 proteins are required for hypaxial somitic myogenesis in the mouse embryo. Developmental biology *302*, 602-616.
Grifone, R., Demignon, J., Houbron, C., Souil, E., Niro, C., Seller, M.J., Hamard, G., and Maire, P. (2005). Six1 and Six4 homeoproteins are required for Pax3 and Mrf expression during myogenesis in the mouse embryo. Development *132*, 2235-2249.
Himeda, C.L., Barro, M.V., and Emerson, C.P., Jr. (2013). Pax3 synergizes with Gli2 and Zic1 in transactivating the Myf5 epaxial somite enhancer. Dev Biol *383*, 7-14.
Holland, L.Z., Holland, N.D., and Gilland, E. (2008). Amphioxus and the evolution of head segmentation. Integr Comp Biol *48*, 630-646.
Holland, L.Z., Holland, P.W.H., and Holland, N.D. (1996). Revealing homologies between body parts of distantly related animals by in situ hybridization to developmental genes: amphioxus versus vertebrates. In: Ferraris JD, Palumbi SR (eds) Molecular zoology: advances, strategies, and protocols *Wiley-Liss, New York, pp 267–282. (473–483)*.
Holland, L.Z., Schubert, M., Kozmik, Z., and Holland, N.D. (1999). AmphiPax3/7, an amphioxus paired box gene: insights into chordate myogenesis, neurogenesis, and the possible evolutionary precursor of definitive vertebrate neural crest. Evolution & development *1*, 153-165.

Holland, P.W., Koschorz, B., Holland, L.Z., and Herrmann, B.G. (1995). Conservation of Brachyury (T) genes in amphioxus and vertebrates: developmental and evolutionary implications. Development *121*, 4283-4291.
Houtmeyers, R., Souopgui, J., Tejpar, S., and Arkell, R. (2013). The ZIC gene family encodes multi-functional proteins essential for patterning and morphogenesis. Cellular and molecular life sciences : CMLS *70*, 3791-3811.
Hubaud, A., and Pourquie, O. (2014). Signalling dynamics in vertebrate segmentation. Nat Rev Mol Cell Biol *15*, 709-721.
Imai, K.S., Hino, K., Yagi, K., Satoh, N., and Satou, Y. (2004). Gene expression profiles of transcription factors and signaling molecules in the ascidian embryo: towards a comprehensive understanding of gene networks. Development *131*, 4047-4058.
Inoue, T., Ota, M., Mikoshiba, K., and Aruga, J. (2007). Zic2 and Zic3 synergistically control neurulation and segmentation of paraxial mesoderm in mouse embryo. Developmental biology *306*, 669-684.
Jaynes, J.B., and O'Farrell, P.H. (1991). Active repression of transcription by the engrailed homeodomain protein. The EMBO journal *10*, 1427-1433.
Jouve, C., Iimura, T., and Pourquie, O. (2002). Onset of the segmentation clock in the chick embryo: evidence for oscillations in the somite precursors in the primitive streak. Development *129*, 1107-1117.
Kenny, D.A., Jurata, L.W., Saga, Y., and Gill, G.N. (1998). Identification and characterization of LMO4, an LMO gene with a novel pattern of expression during embryogenesis. Proceedings of the National Academy of Sciences of the United States of America *95*, 11257-11262.
Kok, F.O., Shepherd, I.T., and Sirotkin, H.I. (2010). Churchill and Sip1a repress fibroblast growth factor signaling during zebrafish somitogenesis. Developmental dynamics : an official publication of the American Association of Anatomists *239*, 548-558.
Koltzoff, N. (1902). Entwicklungsgeschichte des Kopfes von Petromyzon Planeri. Ein Beitrag zur Lehreüber Metamerie desWirbelthierkopfes. . Bull Soc Imp Nat Moscou *16:259–589*.
Koop, D., Chen, J., Theodosiou, M., Carvalho, J.E., Alvarez, S., de Lera, A.R., Holland, L.Z., and Schubert, M. (2014). Roles of retinoic acid and Tbx1/10 in pharyngeal segmentation: amphioxus and the ancestral chordate condition. Evodevo *5*, 36.
Kozmik, Z., Holland, N.D., Kreslova, J., Oliveri, D., Schubert, M., Jonasova, K., Holland, L.Z., Pestarino, M., Benes, V., and Candiani, S. (2007). Pax-Six-Eya-Dach network during amphioxus development: conservation in vitro but context specificity in vivo. Developmental biology *306*, 143-159.
Kuratani, S. (2008). Is the vertebrate head segmented?-evolutionary and developmental considerations. Integr Comp Biol *48*, 647-657.
Kuratani, S., Horigome, N., and Hirano, S. (1999). Developmental morphology of the head mesoderm and reevaluation of segmental theories of the vertebrate head: evidence from embryos of an agnathan vertebrate, Lampetra japonica. Developmental biology *210*, 381-400.
Lane, M.E., Runko, A.P., Roy, N.M., and Sagerstrom, C.G. (2002). Dynamic expression and regulation by Fgf8 and Pou2 of the zebrafish LIM-only gene, lmo4. Mechanisms of development *119 Suppl 1*, S185-189.
Le Douarin, N.M., and Dupin, E. (2012). The neural crest in vertebrate evolution. Current opinion in genetics & development *22*, 381-389.
Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics *26*, 589-595.
Mahadevan, N.R., Horton, A.C., and Gibson-Brown, J.J. (2004). Developmental expression of the amphioxus Tbx1/ 10 gene illuminates the evolution of vertebrate branchial arches and sclerotome. Dev Genes Evol *214*, 559-566.
Mansfield, J.H., Haller, E., Holland, N.D., and Brent, A.E. (2015). Development of somites and their derivatives in amphioxus, and implications for the evolution of vertebrate somites. EvoDevo *6*, 21.

Maruhashi, M., Van De Putte, T., Huylebroeck, D., Kondoh, H., and Higashi, Y. (2005). Involvement of SIP1 in positioning of somite boundaries in the mouse embryo. Developmental dynamics : an official publication of the American Association of Anatomists *234*, 332-338.
Minguillon, C., Jimenez-Delgado, S., Panopoulou, G., and Garcia-Fernandez, J. (2003). The amphioxus Hairy family: differential fate after duplication. Development *130*, 5903-5914.
Oulion, S., Bertrand, S., Belgacem, M.R., Le Petillon, Y., and Escriva, H. (2012). Sequencing and analysis of the Mediterranean amphioxus (Branchiostoma lanceolatum) transcriptome. PLoS One *7*, e36554.
Pan, H., Gustafsson, M.K., Aruga, J., Tiedken, J.J., Chen, J.C., and Emerson, C.P., Jr. (2011). A role for Zic1 and Zic2 in Myf5 regulation and somite myogenesis. Developmental biology *351*, 120-127.
Raible, F., and Brand, M. (2001). Tight transcriptional control of the ETS domain factors Erm and Pea3 by Fgf signaling during early zebrafish development. Mechanisms of development *107*, 105-117.
Rasmussen, S.L., Holland, L.Z., Schubert, M., Beaster-Jones, L., and Holland, N.D. (2007). Amphioxus AmphiDelta: evolution of Delta protein structure, segmentation, and neurogenesis. Genesis *45*, 113-122.
Relaix, F., Demignon, J., Laclef, C., Pujol, J., Santolini, M., Niro, C., Lagha, M., Rocancourt, D., Buckingham, M., and Maire, P. (2013). Six homeoproteins directly activate Myod expression in the gene regulatory networks that control early myogenesis. PLoS genetics *9*, e1003425.
Sambasivan, R., Kuratani, S., and Tajbakhsh, S. (2011). An eye on the head: the development and evolution of craniofacial muscles. Development *138*, 2401-2415.
Sawada, K., Fukushima, Y., and Nishida, H. (2005). Macho-1 functions as transcriptional activator for muscle formation in embryos of the ascidian Halocynthia roretzi. Gene expression patterns : GEP *5*, 429-437.
Schlosser, G. (2008). Do vertebrate neural crest and cranial placodes have a common evolutionary origin? BioEssays : news and reviews in molecular, cellular and developmental biology *30*, 659-672.
Schlosser, G. (2015). Vertebrate cranial placodes as evolutionary innovations--the ancestor's tale. Current topics in developmental biology *111*, 235-300.
Schubert, F.R., Tremblay, P., Mansouri, A., Faisst, A.M., Kammandel, B., Lumsden, A., Gruss, P., and Dietrich, S. (2001). Early mesodermal phenotypes in splotch suggest a role for Pax3 in the formation of epithelial somites. Developmental dynamics : an official publication of the American Association of Anatomists *222*, 506-521.
Schubert, M., Meulemans, D., Bronner-Fraser, M., Holland, L.Z., and Holland, N.D. (2003). Differential mesodermal expression of two amphioxus MyoD family members (AmphiMRF1 and AmphiMRF2). Gene expression patterns : GEP *3*, 199-202.
Somorjai, I., Bertrand, S., Camasses, A., Haguenauer, A., and Escriva, H. (2008). Evidence for stasis and not genetic piracy in developmental expression patterns of Branchiostoma lanceolatum and Branchiostoma floridae, two amphioxus species that have evolved independently over the course of 200 Myr. Development genes and evolution *218*, 703-713.
Tremblay, P., Dietrich, S., Mericskay, M., Schubert, F.R., Li, Z., and Paulin, D. (1998). A crucial role for Pax3 in the development of the hypaxial musculature and the long-range migration of muscle precursors. Developmental biology *203*, 49-61.
Wasylyk, B., Hagman, J., and Gutierrez-Hartmann, A. (1998). Ets transcription factors: nuclear effectors of the Ras-MAP-kinase signaling pathway. Trends in biochemical sciences *23*, 213-216.
Wray, G.A., Hahn, M.W., Abouheif, E., Balhoff, J.P., Pizer, M., Rockman, M.V., and Romano, L.A. (2003). The evolution of transcriptional regulation in eukaryotes. Molecular biology and evolution *20*, 1377-1419.
Wu, W., Huang, R., Wu, Q., Li, P., Chen, J., Li, B., and Liu, H. (2014). The role of Six1 in the genesis of muscle cell and skeletal muscle development. International journal of biological sciences *10*, 983-989.

Yu, J.K., Holland, N.D., and Holland, L.Z. (2004). Tissue-specific expression of FoxD reporter constructs in amphioxus embryos. Developmental biology *274*, 452-461.

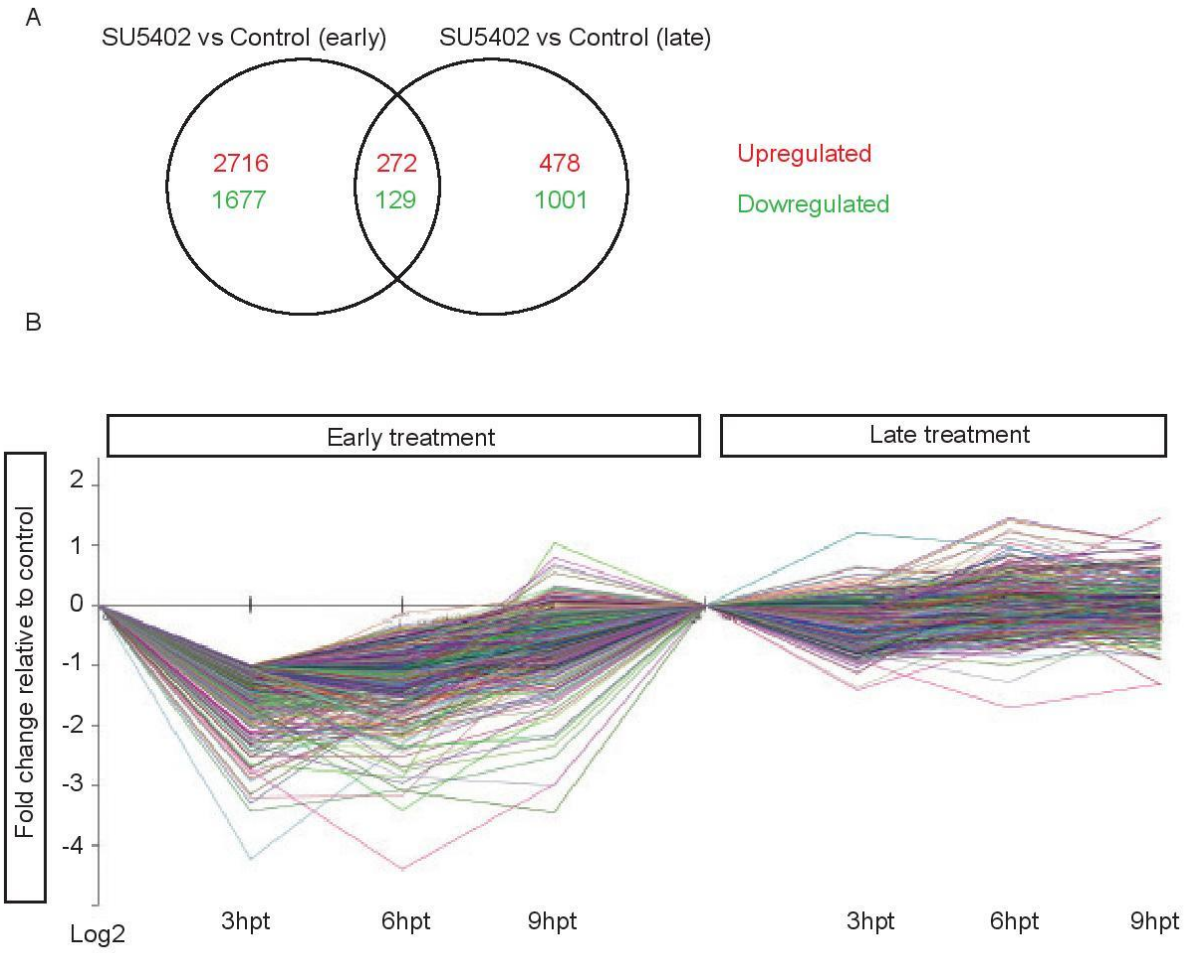
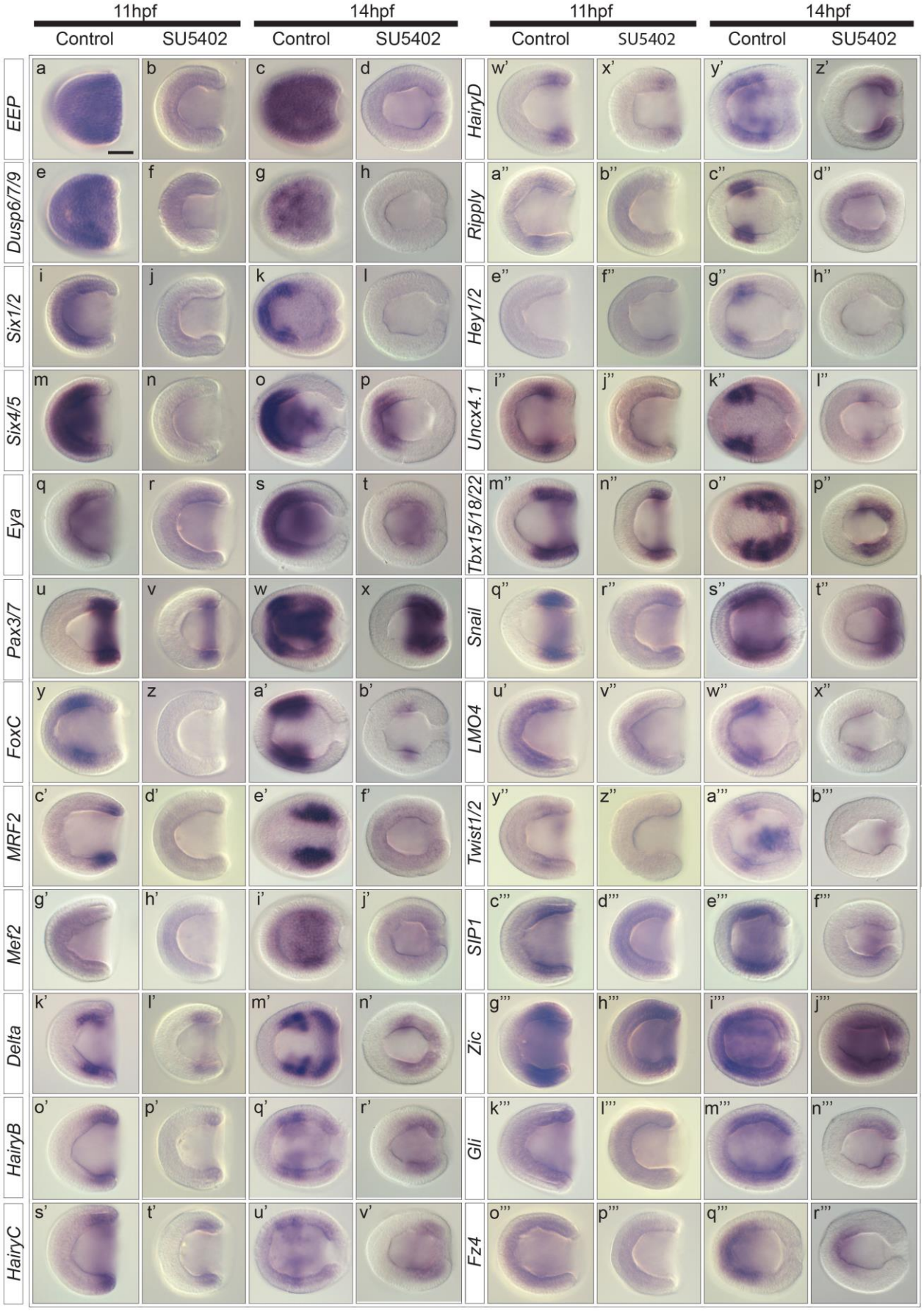

A
SU5402 vs Control (early)
SU5402 vs Control (late)
2716
1677
272
129
478
1001
Upregulated
Dowregulated
B
Early treatment
Late treatment
Fold change relative to control
2
1
0
-1
-2
-3
-4
Log2
3hpt
6hpt
9hpt
3hpt
6hpt
9hpt

Figure 1


11hpf
14hpf
11hpf
14hpf
Control
SU5402
Control
SU5402
Control
SU5402
Control
SU5402
EEP
Dusp6/7/9
Six1/2
Six4/5
Eya
Pax3/7
FoxC
MRF2
Mef2
Delta
HairyB
HairyC
HairyD
Ripply
Hey1/2
Uncx4.1
Tbx15/18/22
Snail
LMO4
Twist1/2
SIP1
Zic
Gli
Fz4
a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
p
q
r
s
t
u
v
w
x
y
z
a'
b'
c'
d'
e'
f'
g'
h'
i'
j'
k'
l'
m'
n'
o'
p'
q'
r'
s'
t'
u'
v'
w'
x'
y'
z'
a''
b''
c''
d''
e''
f''
g''
h''
i''
j''
k''
l''
m''
n''
o''
p''
q''
r''
s''
t''
u'
v''
w''
x''
y''
z''
a'''
b'''
c'''
d'''
e'''
f'''
g'''
h'''
i'''
j'''
k'''
l'''
m'''
n'''
o'''
p'''
q'''
r'''

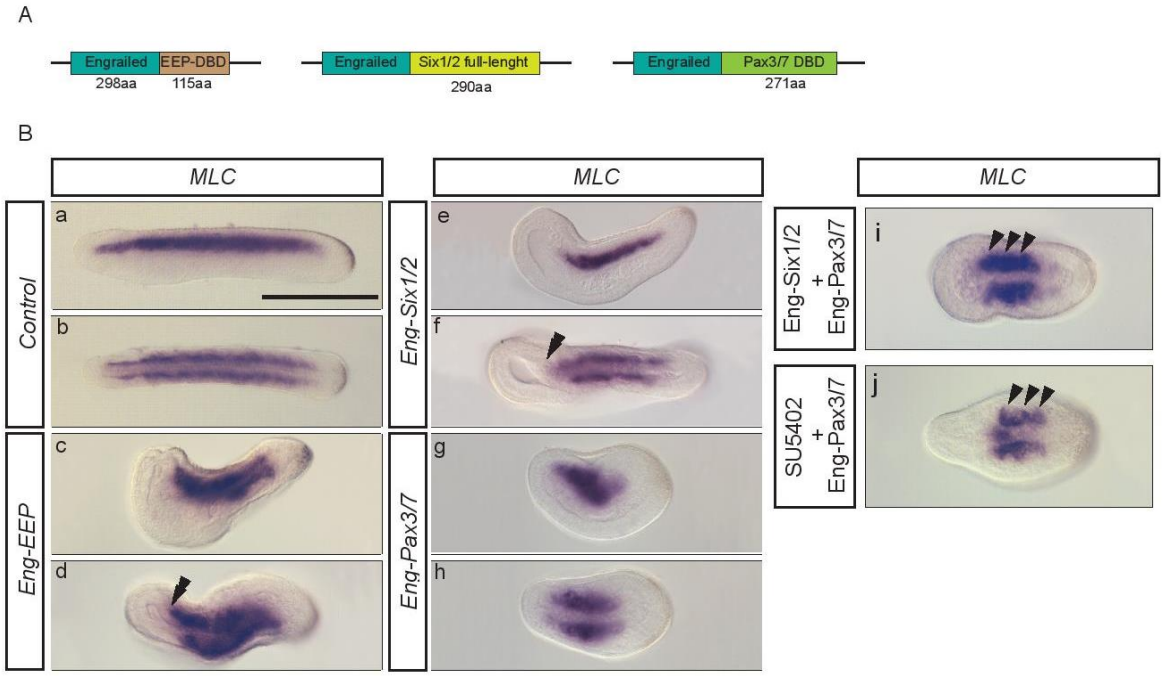

A
Engrailed
EEP-DBD
298aa
115aa
Engrailed
Six1/2 full-lenght
290aa
Engrailed
Pax3/7 DBD
271aa
B
MLC
MLC
MLC
Control
Eng-EEP
Eng-Six1/2
Eng-Pax3/7
Eng-Six1/2 + Eng-Pax3/7
SU5402 + Eng-Pax3/7
a
b
c
d
e
f
g
h
i
j

Figure 3

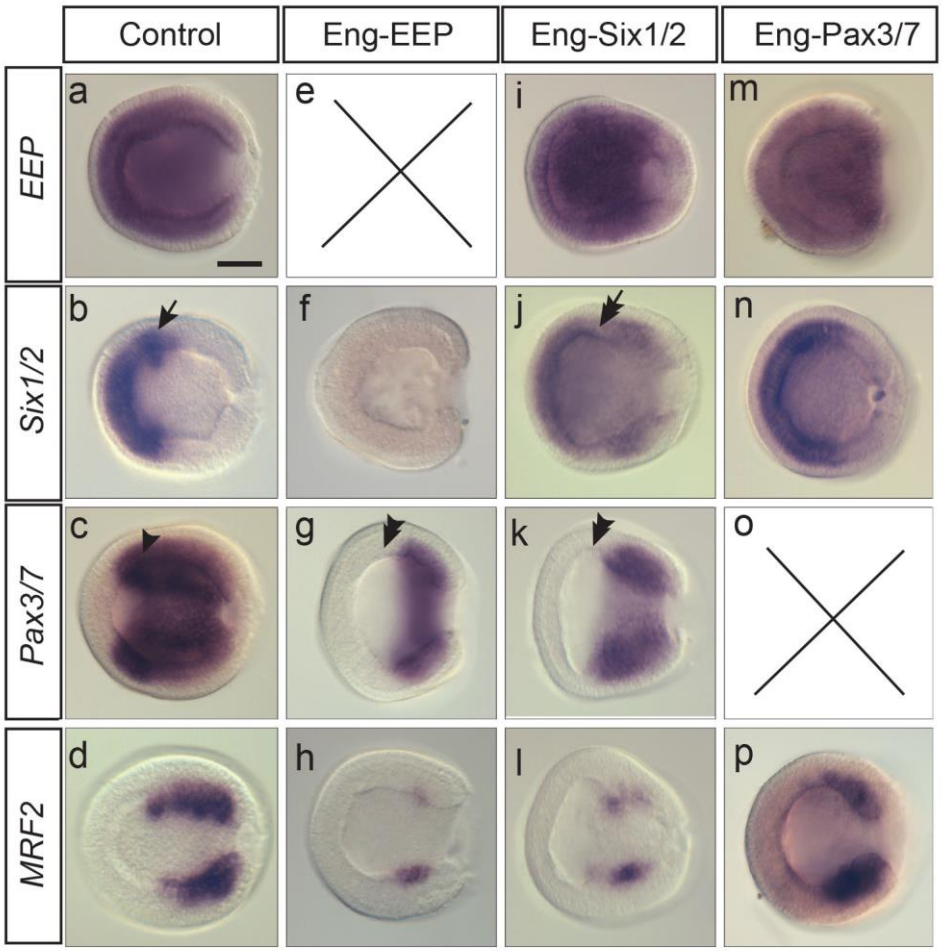

Control
Eng-EEP
Eng-Six1/2
Eng-Pax3/7
EEP
Six1/2
Pax3/7
MRF2
a
e
i
m
b
f
j
n
c
g
k
o
d
h
l
p

Figure 4

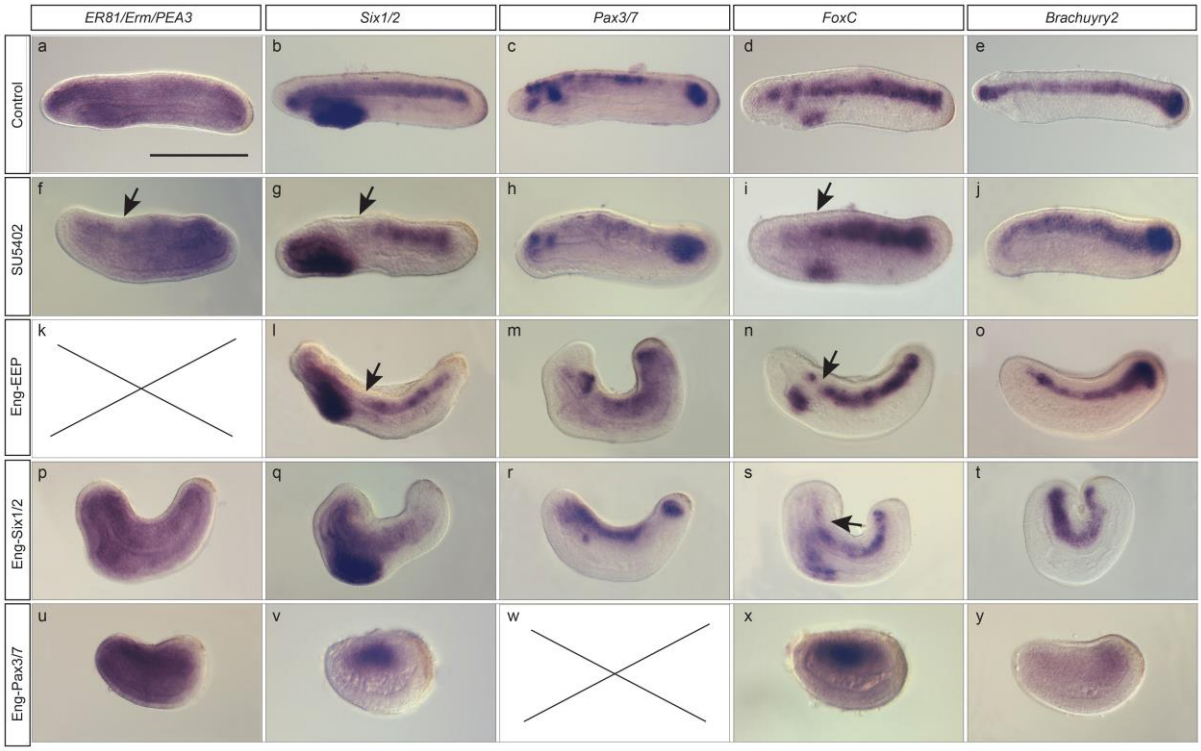

ER81/Erm/PEA3
Six1/2
Pax3/7
FoxC
Brachuyry2
Control
SU5402
Eng-EEP
Eng-Six1/2
Eng-Pax3/7
a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
p
q
r
s
t
u
v
w
x
y

Figure 5

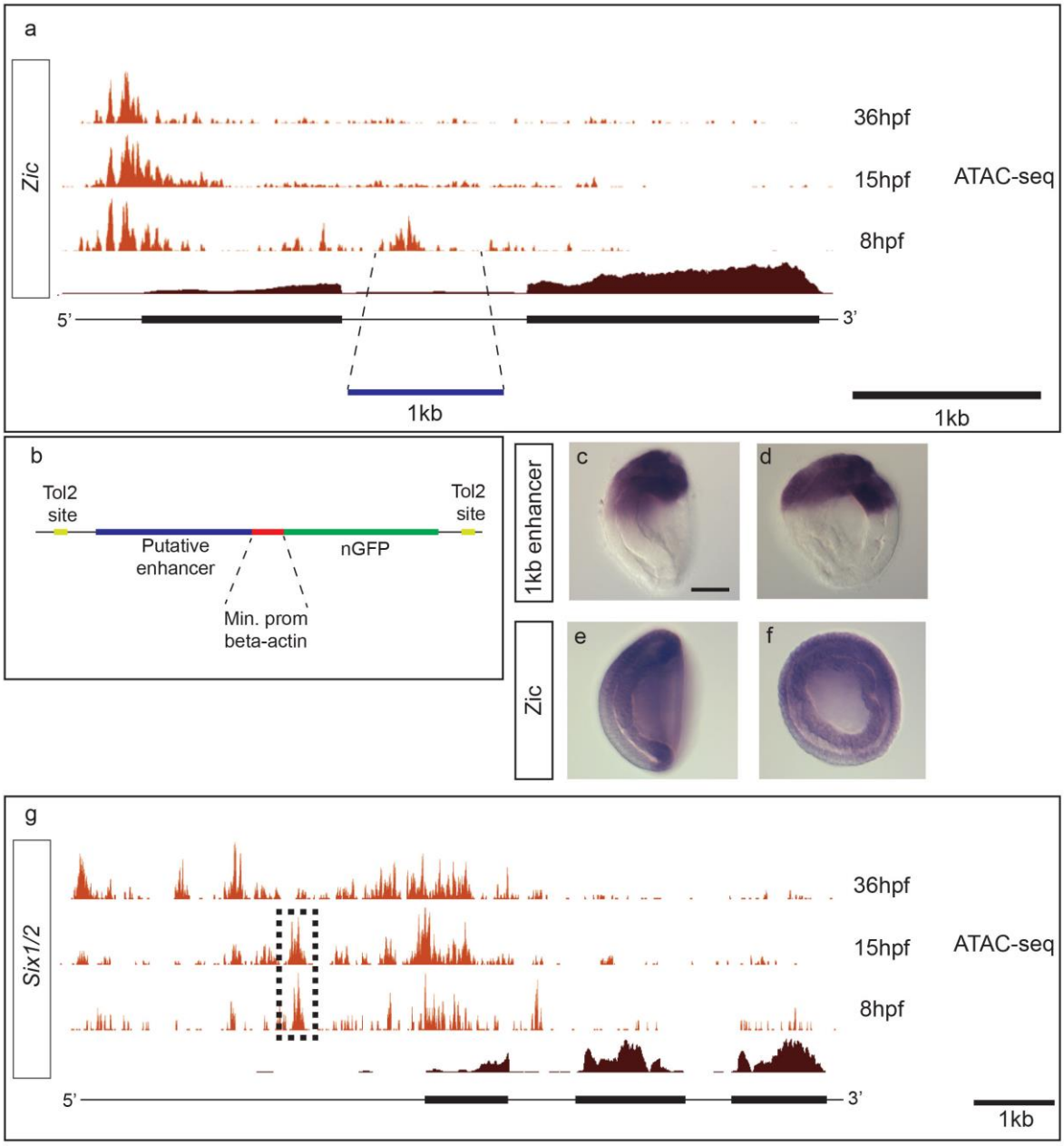

a
Zic
36hpf
15hpf
8hpf
ATAC-seq
5'
3'
1kb
1kb
b
Tol2 site
Tol2 site
Putative enhancer
nGFP
Min. prom beta-actin
1kb enhancer
Zic
c
d
e
f
g
Six1/2
36hpf
15hpf
8hpf
ATAC-seq
5'
3'
1kb

Figure 6

EARLY TREATMENT		log2(SU5402_early_3h/control_early_3h)	log2(SU5402_early_6h/control_early_6h)	log2(SU5402_early_9h/control_early_9h)
Gene	**Contig**			
FGFRL	ContigAmph185	-0,75022	-1,01257	-0,47420
ER81/Erm/Pea3	ContigAmph721	-1,56461	-1,79781	-1,40944
Dusp6/7/9	ContigAmph532	-0,99657	-1,31484	-1,50884
Six1/2	ContigAmph22225	0,05092	-3,31653	-2,15307
Six4/5	ContigAmph11898	-1,22021	-0,66773	-0,09167
Eya	ContigAmph11987	-0,61321	-0,82563	-0,45499
Pax3/7	ContigAmph722	-1,73175	-0,73943	0,00035
MRF1 (MyoD)	ContigAmph20	-2,57887	-5,02267	-4,03844
MRF2	ContigAmph48667	-1,58165	-2,27308	-4,66284
MRF2+	ContigAmph48666	Inf	-Inf	-3,70710
nMRF	ContigAmph23548	-1,83499	-1,56040	-3,17398
Mef2	ContigAmph68834	-1,26712	-0,45172	-0,74420
MyosinIX	ContigAmph25887	2,21487	-1,71201	0,55038
Tropomyosin	ContigAmph13883	-1,06364	-1,43108	-1,52932
Myosin M	ContigAmph3928	-1,20789	1,40176	0,08484
Megf10 (multi-epidermal growth factor)	ContigAmph5426	-0,84379	-1,4073	0,53060
Delta	ContigAmph545	-0,78555	-1,33571	-1,30155
HairyA	ContigAmph69139	-0,32616	-1,32118	-1,47409
HairyB	ContigAmph23802	0,24193	-0,81341	-1,13287
HairyC	ContigAmph13984	-0,29795	-0,94975	-1,08168
HairyD	ContigAmph26026	0,42427	-0,25837	-1,10155
Tbx15/18/22	ContigAmph10784	-1,79183	-1,11924	0,21992
Tbx2/3	ContigAmph25992	-1,25959	-0,23618	-0,19686
Ripply	ContigAmph12500	-1,70922	-3,10483	-3,46246
Uncx4.1	ContigAmph16193	0,37002	-0,86587	-2,55917
Hey1/2	ContigAmph10502	1,00286	-1,83256	-3,77714
FoxAa	ContigAmph13681	0,06526	-1,90987	0,00446
FoxAb	ContigAmph605	-1,09346	-0,47433	0,39692
FoxD	ContigAmph16311	-1,21079	-1,05958	-1,44851
FoxK	ContigAmph8300	-1,03081	-0,71772	-0,32742
FoxM	ContigAmph8796	-1,44343	0,42414	0,64578
FoxN1/4a	ContigAmph96	-1,28591	-1,39333	2,64616
FoxJ1	ContigAmph7689	-1,21796	-0,30612	0,74905
FoxN2/3	ContigAmph7898	-1,07121	-0,10349	0,20278
Iroquois A	ContigAmph3136	-0,72778	-1,39670	-1,67986
Iroquois B	ContigAmph190	-1,50283	0,37380	-0,12389

Iroquois C	ContigAmph20930	-3,43329	-3,08488	-2,53932
Snail	ContigAmph197	-1,16500	-2,71027	-2,18080
LMO4	ContigAmph6103	-1,12804	-0,38101	-0,38186
Twist1/2	ContigAmph6309	-0,46100	-1,06741	-1,24777
Nk2	ContigAmph16821	0,00838	-4,84895	-0,95520
Nkx6	ContigAmph12366	-1,75002	-1,06167	-0,02687
MYLK (myosin light chain kinase, smooth muscle)	ContigAmph6257	-1,17099	-1,16299	-0,02689
Mnx	ContigAmph26310	-0,22533	-1,39795	-0,06487
Dmbx	ContigAmph81202	1,94723	-4,04451	-2,19549
Lhx2	ContigAmph13034	0,36787	-3,42387	-0,54688
Lim1/5	ContigAmph27865	0,29275	-2,12117	1,29458
Gli	ContigAmph6177	-1,72810	-0,91774	-0,10497
Zic	ContigAmph13605	-0,53027	-0,96839	-0,92221
SIP1 (smad interacting protein 1)	ContigAmph10806	-0,94433	-2,31491	-1,17379
Hedgehog protein	ContigAmph85576	-1,50136	-1,32088	-1,08093
Neurogenin	ContigAmph372	-0,65756	-1,67807	-0,88917
BMP3/3b	ContigAmph6959	0,03403	-1,66404	-0,24959
Gremlim	ContigAmph14171	-0,05255	-1,63172	-1,52285
Gbx	ContigAmph272	-0,89803	-1,51858	-0,61909
Orthopedia	ContigAmph7952	-2,13908	-1,15440	-1,04832
CBFA2T1	ContigAmph18897	-1,61053	0,60386	0,70078
FezF	ContigAmph11	0,94353	-3,51383	-2,38567
Ash	ContigAmph2086	1,66188	-2,66552	-1,73907
SoxE	ContigAmph589	-0,45224	-1,14821	-1,88670
Frizzled4	ContigAmph6161	-1,09390	-1,39493	-0,56375
Frizzled5/8	ContigAmph30949	-0,52694	-0,79982	-0,28533
FoxC	ContigAmph25978	0,52571	-2,54197	-2,79534
Jagged1	ContigAmph5697	-0,90972	-1,56966	0,72447
Jagged2	ContigAmph5698	-0,74360	-1,58064	0,73330
Hox2	ContigAmph34651	-0,13737	-1,45370	-0,82365
Hox1	ContigAmph13529	-1,13991	-1,29861	0,08239
Cdx	ContigAmph13490	-1,17321	-0,60296	0,13784
EphrinA	ContigAmph852	0,45725	-1,23355	-0,06460
zinc finger FYVE domain	ContigAmph19894	-1,31487	-1,19708	-0,63517
DDR2 (discoidin domain-containing receptor)	ContigAmph8142	-1,96146	-0,53249	-0,46506
recombining binding protein suppressor of hairless	ContigAmph4612	-1,19756	-0,49787	-0,17080
SFMBT2 (Scm-like with four MBT domains protein 1)	ContigAmph33445	-1,01108	-0,31292	-0,15054
TGF-B (transforming growth factor beta receptor type 3 precursor)	ContigAmph19500	-1,15100	-0,68941	0,19205

Patched protein like	ContigAmph5682	-1,64617	-0,30394	-0,27210
Plecktrin5/7/9	ContigAmph17242	-2,43322	-1,15326	-0,65923
RS-like (rolling stone-like)	ContigAmph21895	-1,35113	-0,20784	-1,35544
Uncharacterized protein (Trauco)	ContigAmph71939	-1,75270	0,18728	0,77270
Zinc finger protein (Caleuche)	ContigAmph6363	-1,76936	-2,85918	-3,01188
F-box only protein 32 (Pincoya)	ContigAmph32885	-1,92100	-0,90971	0,43214
WD repeat 49	ContigAmph1857	-1,66174	-1,68762	-0,34796
AnkirinBR1	ContigAmph24453	-1,70193	-1,45096	-0,40491
Ankirin-notch	ContigAmph8400	-1,40145	-0,46392	-0,34361
Ataxin	ContigAmph3989	-1,44675	-0,91575	-0,95139
Coe	ContigAmph1527	-0,97345	-1,04549	0,68815
Mdp	ContigAmph11791	-0,36609	-0,57034	-0,09256
Arx	ContigAmph7953	-2,43752	-2,18903	-0,79826
KREMEN	ContigAmph1046	-0,56076	-1,04851	-1,18232
4andhalfLIM domain protein 2	ContigAmph17569	-1,22760	-0,89704	-0,26086
4andhalfLIM domain protein 5	ContigAmph15489	-1,02799	-0,83672	-0,46931
BIGH (transforming growth factor-beta-induced protein ig-3)	ContigAmph44472	-0,79400	-1,41611	-1,14225
C2orf81 homolog	ContigAmph12430	-1,51484	-1,26742	-0,30243
Pitx	ContigAmph618	-0,29838	1,37420	0,43097
Tbx1/10	ContigAmph12871	0,98260	1,69886	1,40295
FoxEa	ContigAmph10961	5,10491	5,60377	1,25743
Scratch	ContigAmph19102	3,91948	5,06805	5,02164

Table 1 (early treatment)

LATE TREATMENT		log2(SU5402_late_3h/control_late_3h)	log2(SU5402_late_6h/control_late_6h)	log2(SU5402_late_9h/control_late_9h)
Gene	**Contig**			
FGFRL	ContigAmph185	-0,55531	-0,25536	0,13081
ER81/Erm/Pea3	ContigAmph721	-0,13433	-0,79362	-1,41594
Dusp6/7/9	ContigAmph532	-0,47493	-2,09725	-2,73403
Six1/2	ContigAmph22225	-0,60962	-0,71552	-0,49800
Six4/5	ContigAmph11898	-0,26487	-0,20571	0,06941
Eya	ContigAmph11987	-0,55037	-0,74721	-0,37106
Pax3/7	ContigAmph722	0,07824	-0,06322	-0,12114
MRF1 (MyoD)	ContigAmph20	-1,22014	-1,12843	-0,60921
MRF2	ContigAmph48667	-1,90232	-1,42102	-1,15898
MRF2+	ContigAmph48666	-1,50011	-0,26131	-0,81189
nMRF	ContigAmph23548	-1,71384	-0,91422	-0,99641
Mef2	ContigAmph68834	-0,17179	-0,45941	0,16885
MyosinIX	ContigAmph25887	-0,24751	-0,03713	0,10879
Tropomyosin	ContigAmph13883	-0,48761	-1,02599	-1,08874
Myosin M	ContigAmph3928	-1,97685	-1,03739	-0,44782
Megf10 (multi-epidermal growth factor)	ContigAmph5426	1,56646	0,64270	0,26821
Delta	ContigAmph545	-0,44709	-0,71373	-0,25259
HairyA	ContigAmph69139	-0,52144	-0,67774	-0,44358
HairyB	ContigAmph23802	-0,78799	-0,88475	-0,45955
HairyC	ContigAmph13984	-0,68412	-0,41929	-0,63313
HairyD	ContigAmph26026	-0,68602	-0,68989	-0,75641
Tbx15/18/22	ContigAmph10784	-0,71614	-0,84508	0,10486
Tbx2/3	ContigAmph25992	-0,17460	-1,03965	-0,24722
Ripply	ContigAmph12500	-0,55466	-0,60423	-0,34517
Uncx4.1	ContigAmph16193	-0,56774	-0,30056	-0,38005
Hey1/2	ContigAmph10502	-0,29938	-0,44955	-0,36623
FoxAa	ContigAmph13681	-0,68988	-0,52425	-0,04030
FoxAb	ContigAmph605	-0,43547	-0,19450	0,045605
FoxD	ContigAmph16311	-1,36171	-1,12131	-0,77631
FoxK	ContigAmph8300	-0,44459	-0,39697	-0,36803
FoxM	ContigAmph8796	-0,43632	-0,48020	-0,23873
FoxN1/4a	ContigAmph96	0,01241	-0,23955	0,20088
FoxJ1	ContigAmph7689	0,90408	-0,00036	-0,27740
FoxN2/3	ContigAmph7898	-0,22115	-0,02295	0,11497
Iroquois A	ContigAmph3136	-0,63167	-0,99048	-0,34592
Iroquois B	ContigAmph190	-0,15264	-0,33287	0,01154

Iroquois C	ContigAmph20930	-0,49631	-0,79893	-0,16339
Snail	ContigAmph197	-0,74459	-0,71977	-0,43198
LMO4	ContigAmph6103	-0,57331	-0,65750	-0,67496
Twist1/2	ContigAmph6309	-0,83667	-0,93456	-0,54575
Nk2	ContigAmph16821	-0,70484	-1,74418	-1,26544
Nkx6	ContigAmph12366	-0,66370	-0,17798	-0,10810
MYLK (myosin light chain kinase, smooth muscle)	ContigAmph6257	-0,72112	0,35196	1,45288
Mnx	ContigAmph26310	-0,43697	-0,41574	-0,20361
Dmbx	ContigAmph81202	-0,60131	-0,37863	0,36055
Lhx2	ContigAmph13034	-1,39335	-2,91994	-1,17088
Lim1/5	ContigAmph27865	-0,54496	-1,15735	-0,20104
Gli	ContigAmph6177	-0,86631	-0,60416	-0,17213
Zic	ContigAmph13605	-0,62629	-0,54165	-0,31076
SIP1 (smad interacting protein 1)	ContigAmph10806	-1,00105	-0,59389	0,05705
Hedgehog protein	ContigAmph85576	-0,09886	-0,46778	-0,30100
Neurogenin	ContigAmph372	-0,58766	-1,17254	-0,17740
BMP3/3b	ContigAmph6959	-0,07497	-0,22288	-0,48451
Gremlim	ContigAmph14171	-0,16347	0,13195	0,96367
Gbx	ContigAmph272	-0,62605	-0,55233	0,07332
Orthopedia	ContigAmph7952	-0,11142	-0,43097	-0,49430
CBFA2T1	ContigAmph18897	-0,27458	-0,58582	-0,03539
FezF	ContigAmph11	-0,74738	-0,85548	-0,42916
Ash	ContigAmph2086	0,33137	-0,58692	-0,02093
SoxE	ContigAmph589	-0,47233	-0,56939	-0,32775
Frizzled4	ContigAmph6161	-0,28670	-0,22734	-0,13469
Frizzled5/8	ContigAmph30949	0,89175	0,15306	-0,20587
FoxC	ContigAmph25978	-0,45362	-0,40519	-0,21133
Jagged1	ContigAmph5697	0,87788	-0,20664	-0,73096
Jagged2	ContigAmph5698	0,97711	-0,16317	-0,71654
Hox2	ContigAmph34651	-1,37347	-1,71724	-1,00728
Hox1	ContigAmph13529	-0,28290	-0,32049	0,07783
Cdx	ContigAmph13490	-0,43771	-0,47840	0,00503
EphrinA	ContigAmph852	-0,36585	-0,20883	0,13685
zinc finger FYVE domain	ContigAmph19894	0,12568	-0,08577	-0,07575
DDR2 (discoidin domain-containing receptor)	ContigAmph8142	-1,04730	-0,39516	0,06935
recombining binding protein suppressor of hairless	ContigAmph4612	-0,54059	-0,24352	0,17370
SFMBT2 (Scm-like with four MBT domains protein 1)	ContigAmph33445	-0,39933	-0,32113	-0,37411
TGF-B (transforming growth factor beta receptor type 3 precursor)	ContigAmph19500	-0,41427	-0,26066	0,11028

Patched protein like	ContigAmph5682	-0,43257	-0,10653	0,34510
Plecktrin5/7/9	ContigAmph17242	-0,24745	-0,06351	0,16132
RS-like (rolling stone-like)	ContigAmph21895	-0,34099	0,06819	-0,00649
Uncharacterized protein (Trauco)	ContigAmph71939	-0,14778	-0,50688	1,00230
Zinc finger protein (Caleuche)	ContigAmph6363	-0,79437	-1,28931	-0,20981
F-box only protein 32 (Pincoya)	ContigAmph32885	-0,78725	-0,40273	-0,61728
WD repeat 49	ContigAmph1857	0,196526	0,01251	0,00767
AnkirinBR1	ContigAmph24453	-0,46296	-0,13931	0,15490
Ankirin-notch	ContigAmph8400	-0,67667	0,02464	0,28382
Ataxin	ContigAmph3989	-0,32558	-0,34553	0,17824
Coe	ContigAmph1527	-1,02104	-0,57769	-0,20524
Mdp	ContigAmph11791	-0,37071	-0,42159	-0,10674
Arx	ContigAmph7953	-0,41841	-0,08032	-0,17481
KREMEN	ContigAmph1046	-0,17690	-0,53029	-0,35302
4andhalfLIM domain protein 2	ContigAmph17569	0,584313	0,13634	0,22598
4andhalfLIM domain protein 5	ContigAmph15489	-0,72203	-1,06985	-0,59169
BIGH (transforming growth factor-beta-induced protein ig-3)	ContigAmph44472	0,12063	-0,40541	0,35041
C2orf81 homolog	ContigAmph12430	0,05603	-0,03099	-0,29486
Pitx	ContigAmph618	-0,67810	-0,84858	0,02582
Tbx1/10	ContigAmph12871	-0,86406	-1,18742	-1,32082
FoxEa	ContigAmph10961	-0,71653	-0,94467	-0,83185
Scratch	ContigAmph19102	-0,17393	-1,03846	-0,14243

Table 1 (late treatment)

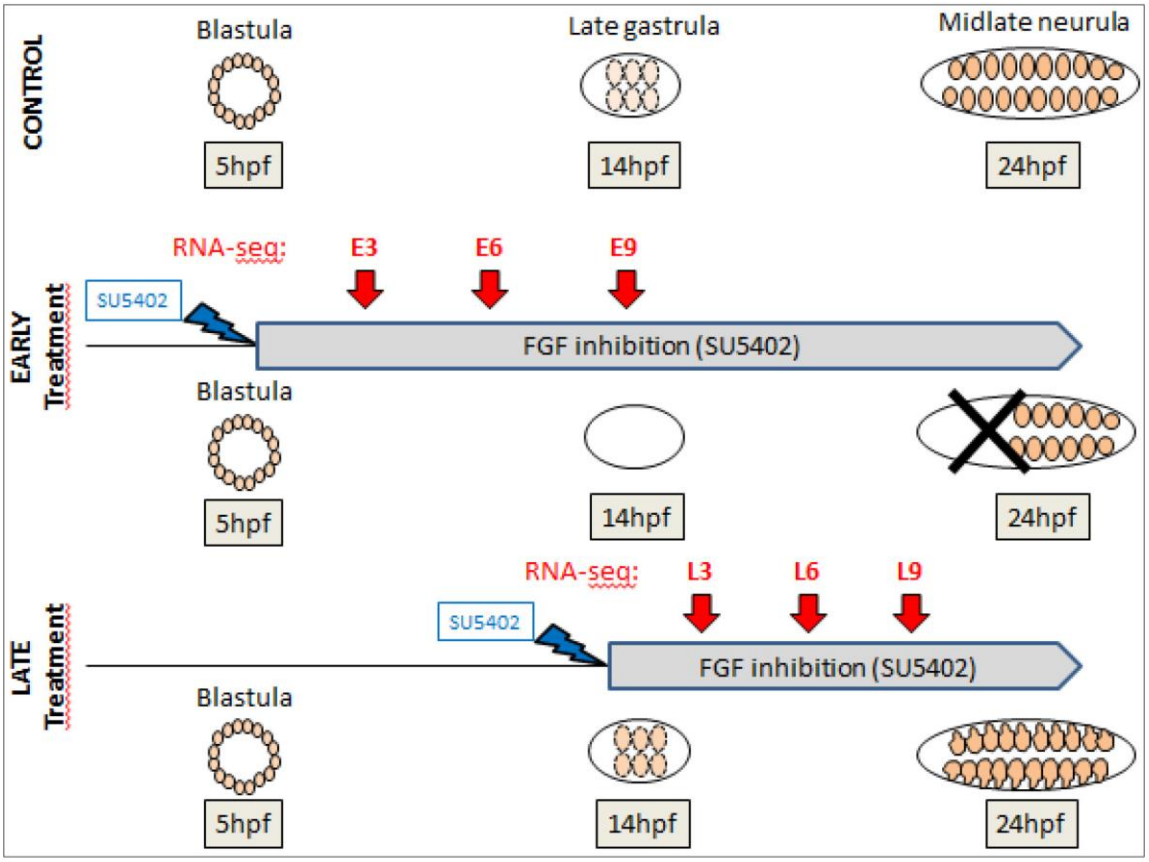
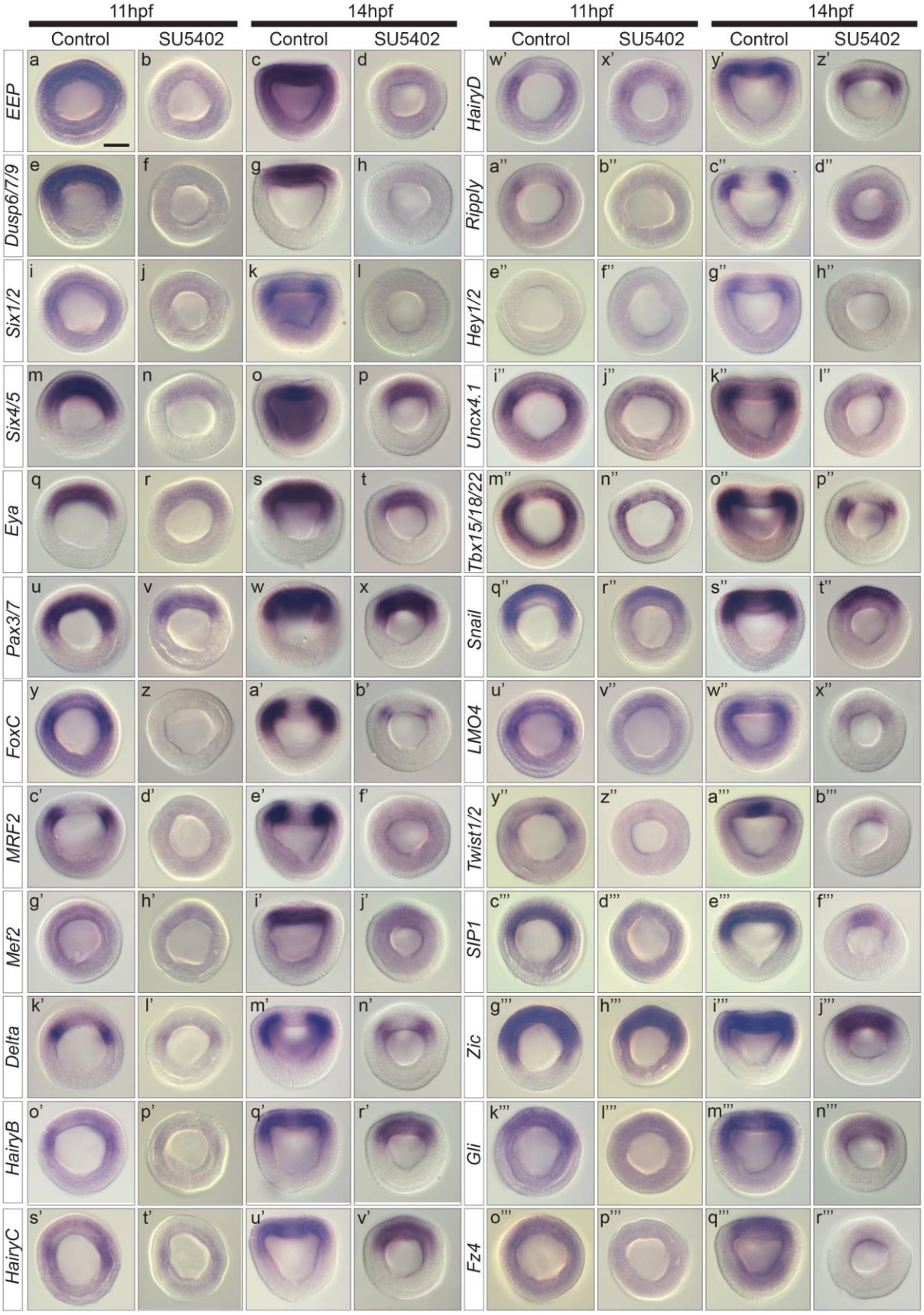


Figure supplementary 1


11hpf
14hpf
11hpf
14hpf
Control
SU5402
Control
SU5402
Control
SU5402
Control
SU5402
EEP
Dusp6/7/9
Six1/2
Six4/5
Eya
Pax3/7
FoxC
MRF2
Mef2
Delta
HairyB
HairyC
HairyD
Ripply
Hey1/2
Uncx4.1
Tbx15/18/22
Snail
LMO4
Twist1/2
SIP1
Zic
Gli
Fz4
a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
p
q
r
s
t
u
v
w
x
y
z
a'
b'
c'
d'
e'
f'
g'
h'
i'
j'
k'
l'
m'
n'
o'
p'
q'
r'
s'
t'
u'
v'
w'
x'
y'
z'
a''
b''
c''
d''
e''
f''
g''
h''
i''
j''
k''
l''
m''
n''
o''
p''
q''
r''
s''
t''
u'
v''
w''
x''
y''
z''
a'''
b'''
c'''
d'''
e'''
f'''
g'''
h'''
i'''
j'''
k'''
l'''
m'''
n'''
o'''
p'''
q'''
r'''

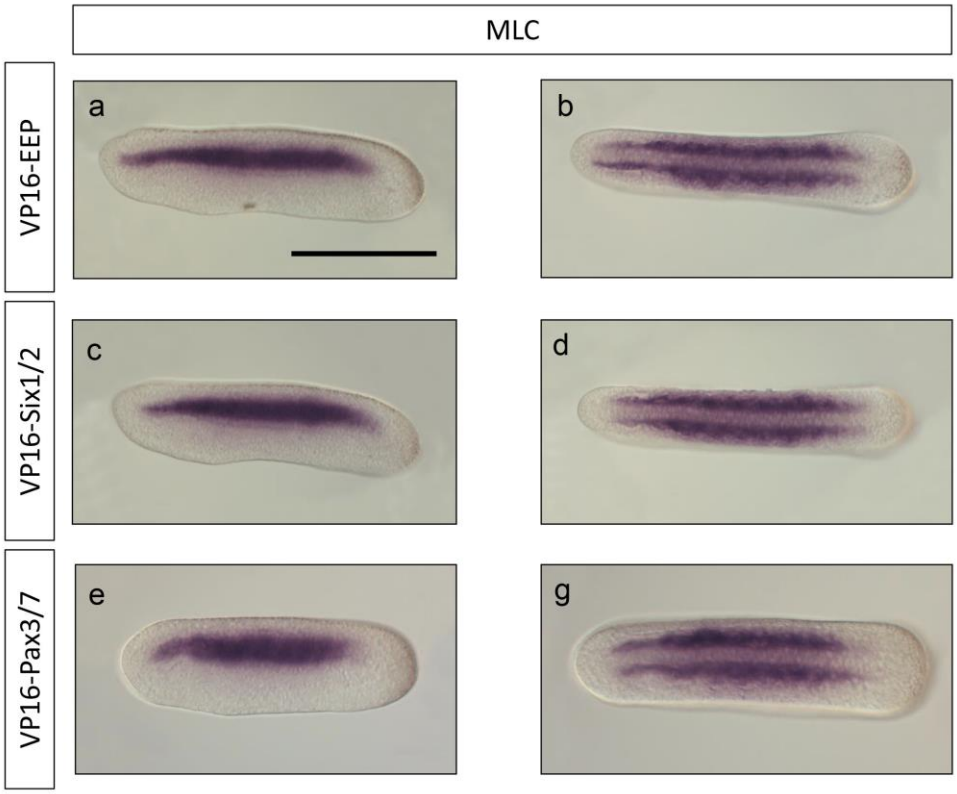


Figure supplementary 3

4. General discussion and additional data

4.1 Hox genes, FGF signal and the anterior somitogenesis in amphioxus

In the first part of this work, I was interested in investigating the relationship between *Hox* genes and the FGF signal in the formation of the anterior somites in amphioxus. This because the most anterior limit of expression of *Hox1* surprisingly matches with the posterior limit of the FGF-sensitive somites (Bertrand et al., 2011; Wada et al., 1999). This leads us to think that a functional relationship between both signals could exist, with the anterior limit of *Hox* genes expression acting as a boundary between the FGF-sensitive and FGF-insensitive somites. Moreover, it was previously proposed that nested expression of *Hox* genes and *Gbx*, a gene belonging to the *Homeobox* family (Castro et al., 2006), at early gastrula stage, establishes the position of the anterior somites (Holland et al., 2008b). Certainly, anterior somites are established at early gastrula, when boundaries of the future somites are indicated by the expression of genes such as *Delta*, their effectors *Hairy*, *Ripply*, *Tbx15/18/22*, *Uncx4.1*, *Hey1/2* or *Six1/2* (Article 2) (Beaster-Jones et al., 2008). Nevertheless, our finding shows that *Hox* genes do not define this boundary (Article 1). Indeed, anterior somitogenesis is not affected by the inhibition or activation of the level of RA signal, a known molecule that controls *Hox* expression in vertebrates as in amphioxus (Escriva et al., 2002). Thus, anterior somites are formed independently of the *Hox* code.

Interestingly, our RNA-seq data reveals that *Hox1* and *Gbx* were downregulated at early stages of development in amphioxus when embryos were treated with SU5402 (Article 2, Table 1), suggesting that they are controlled by the FGF signal. *Hox1* and *Gbx* are not only expressed in the mesoderm, but also in neuroectodermal and endodermal tissues (Castro et al., 2006; Wada et al., 1999). Therefore, it might be possible that at early stages of development this downregulation corresponds to a specific control of the FGF signal over the *Hox* genes for neural development. Indeed, *Hox* genes play major roles in the patterning of the central nervous system (CNS) in vertebrates as in amphioxus (Maden, 2002; Rhinn and Dolle, 2012; Schubert et al., 2006). Moreover, these observations are supported by the fact that genes involved in neural developmental processes (i.e. *Neurogenin*, *FezF*, *CBFA2T1* among other genes with expression domains in the neuroectoderm at early stages) were also downregulated in SU5402 treated embryos at the blastula stage (5hpf). In addition, recent works in our laboratory show the inductive role of the FGF signal for the formation of the anterior neural plate. This work shows how the anterior neural tissue is not maintained when the FGF signal is inhibited. Altogether, these observations are supported by our data from the RNA-seq analysis and corroborate the recent work performed in our laboratory about the neural induction process in amphioxus (unpublished data).

4.2 RNA-seq analysis and the role of the FGF signal in anterior somitogenesis

Our comparative RNA-seq analysis allows us to suggest many conclusions. First of all, we observed how the expression of the different genes fluctuates between the different points at early treatment (3hpt, 6hpt and 9hpt) and late treatment (3hpt, 6hpt and 9hpt) (Table 1). We used STEM software to cluster genes given their expression profiles. From these profiles, we concentrated on genes showing a downregulation at early stages after early treatment and for which no expression modification was observed after late treatment (Fig.1b, Article 2). We believed this profile should be observed for genes that were putatively under the control of the FGF signal at early stages of development but that were not controlled by this signaling pathway afterwards. We confirmed this by *in situ* hybridization. Indeed, genes that showed a downregulation in the paraxial mesoderm after the treatment at the blastula stage (early treatment) (Fig. 2, Article 2), did not show any change when embryos were treated at 15,5hpf (late treatment) (**Figure 31**). Thus, we detected genes highly downregulated (Log2 Fold Change lower than -1,5 and p-value < 0,05) just after early treatment (3hpt) as *ER81/Erm/PEA3*, *Pax3/7*, *Tbx15/18/22*, *Ripply*, *MRF1*, *Iroquois B*, *Iroquois C*, *Orthopedia*, *Nkx6*, *Gli*, *Plecktrin5/7/9*, *CBFA2T1*, *DDR2*, *Patched protein-like*, *Trauco*, *Caleuche*, *Pincoya*, *WDrepeat49*, *AnkirinBR1*, *C2orf81 homolog* and *Arx*. To validate the RNA-seq results, we performed *in situ* hybridization for all of them and deciphering their precise expression patterns. In addition, we were really interested in determining the expression pattern of unknown genes in amphioxus as *Trauco* (Contig 71939), *Caleuche* (Contig 6363) or *Pincoya* (Contig 32885) among others, because they showed a high response to the FGF signal inhibition just after treatment (3hpt) (Table 1, Article 2). However, we failed to observe any signal by *in situ* hybridization for these last genes. Later, we selected genes highly downregulated at 6hpt from the early treatment (Log2 Fold Change lower than -1,5 and p-value < 0,05) such as *Six1/2*, *MyosinIX*, *Hey1/2*, *FoxAa*, *Snail*, *Nk2*, *Dmbx*, *Lhx2*, *Lim1/5*, *SIP1*, *Neurogenin*, *BMP3/3b*, *Gremlin*, *Gbx*, *FezF*, *Ash*, *Jagged1*, and *Jagged2*. Finally, other genes were chosen manually to perform *in situ* hybridization based on their Log2 Fold Change, or their known role in myogenesis or segmental processes (those genes were *Dusp6/7/9*, *Six4/5*, *Eya*, *Twist1/2*, *MYLK*, *Mnx*, *SoxE*, *Fyve domain*, *RS-like*, and *BIGH*).

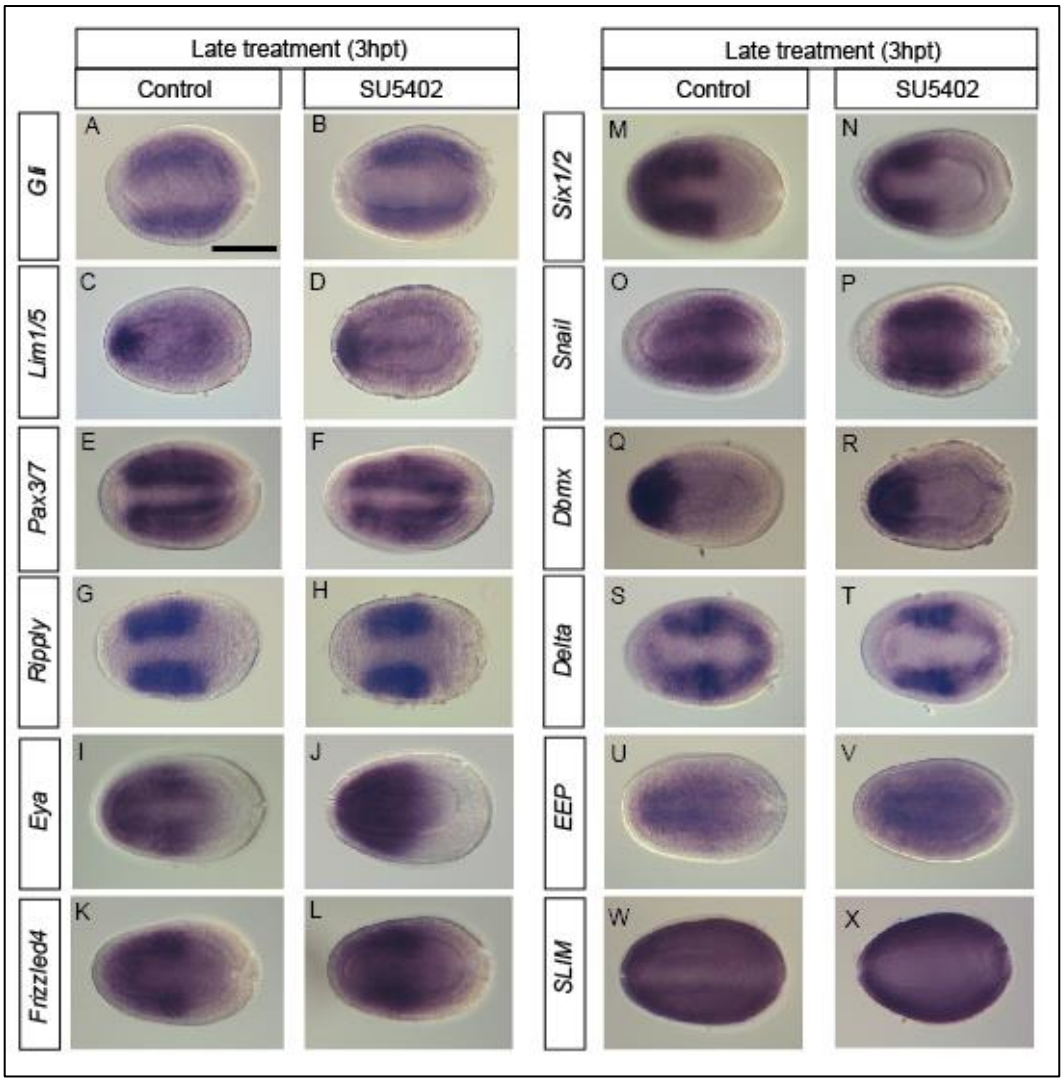


Figure 31. Expression pattern of some early downregulated genes in embryos treated with SU5402 at late stage. In all the panels we observe that expression of the genes that did not change between control and SU5402-treated embryos at late stage 15,5hpf. Embryos were fixed 3 hours post treatment (hpt). **(A-B)** *Gli*; **(C-D)** *Lim1/5*; **(E-F)** *Pax3/7*; **(G-H)** *Ripply*; **(I-J)** *Eya*; **(K-L)** *Frizzled4*; **(M-N)** *Six1/2*; **(O-P)** *Snail*; **(Q-R)** *Dbmx*; **(S-T)** *Delta*; **(U-V)** *ER81/Erm/PEA3*; **(W-X)** *SLIM*. Dorsal views for all the panels. Anterior part of the embryo to the left. Scale bar = 100 µm.

In the second article (Article 2) we highlighted the genes putatively under the control of the FGF signal and with known roles in vertebrate myogenesis (*Six1/2*, *Six4/5*, *Eya*, *Pax3/7*, *FoxC*, *MRFs* genes, *Mef2*), vertebrate segmentation process (*Delta*, theirs cofactors *Hairy*, *Ripply*, *Hey1/2*, *Uncx4.1*, *Tbx15/18/22*) and numerous other genes expressed in the paraxial mesoderm (*Snail*, its hypothetical cofactor *LMO4*, *Twist1/2*, *SIP1*, *Zic*, *Gli*, *Frizzled4*). Thus, RNA-seq together with *in situ* hybridization data show clearly how the expression of genes involved in anterior somitogenesis in amphioxus is abolished at early stages (notably observed in the Fig. 2, Paper 2). Indeed stripes of gene expression indicating the boundaries of the future somites are no longer visible after the treatment at blastula stage (early treatment) (Fig. 2, Article 2). Remarkably, genes with multiple expression domains as

paraxial mesoderm, neural plate or axial mesoderm (i.e. *Snail*, *Twist1/2*, *SIP1* among others) presented a specific downregulation of the expression in the paraxial mesoderm but not in the other territories.

Additionally, embryos treated with SU5402 develop a normal notochord as we observed at the late neurula stage, where the notochord extends all along the antero-posterior axis (Fig. 5j, Paper 2). Thus, FGF signal acts exclusively in the specification of the anterior paraxial mesoderm that gives rise to the most anterior somites in amphioxus. Similarly, in *Xenopus*, the FGF signal maintains a positive feed-back with *Brachyury*, being necessary for the initiation of its transcription and also plays a pivotal role in the establishment of the paraxial mesoderm but is not needed for the establishment of the axial mesoderm (Fletcher and Harland, 2008). In ascidians, the sister group of vertebrates, FGF signal is required for mesenchyme, notochord and secondary muscle development (Kim and Nishida, 2001; Kim et al., 2000). Hemichordates, together with echinoderms, represent the sister group of chordates called ambulacraria. In the hemichordate *Saccoglossus kowalevskii*, the ligand FGF8/17/18 induces mesoderm from endomesoderm (Green et al., 2013). The authors suggested that FGF signal could have been required for the formation of the whole mesoderm in the ancestor of all deuterostomes, followed by a secondary loss of FGF-dependency in different cell populations and within the different deuterostome groups (Green et al., 2013). Thus, in amphioxus, the role of the FGF signal for mesoderm induction was conserved in the anterior paraxial mesoderm and lost in the posterior paraxial mesoderm that is formed independently of the FGF signal.

In amphioxus, even if there are no functional studies demonstrating the role of FGF8/17/18, its expression pattern suggests that it could act as the ligand of FGFR to trigger and induce the anterior paraxial mesoderm in amphioxus (Bertrand et al., 2011). Indeed, *FGF8/17/18* is expressed at the gastrula stage in the dorsal posterior mesendoderm. Later on, at early neurula, transcripts are detected in the posterior part of the dorsal mesendoderm. Then *FGF8/17/18* is no longer expressed in mesodermal tissues. Remarkably, in embryos treated with SU5402 no downregulation of *FGF8/17/18* expression could be detected (data not shown). Additionally, other FGF genes (i.e. *FGF9/16/20*, *FGFA*, *FGFE*, and *FGFC*) are not expressed in mesodermal tissues at early stages when anterior somites are specified, supporting the fact that FGF8/17/18 could be the ligand that binds the FGF receptor to control this process.

Interestingly, the expression pattern of *ER81/Erm/PEA3*, one of the effectors of the FGF signal that is activated by the MAPK pathway (Munchberg and Steinbeisser, 1999; Roussigne and Blader, 2006) overlaps with the early expression of *FGF8/17/18* (Bertrand et al., 2011). Thus, *ER81/Erm/PEA3* starts to be expressed at the gastrula stage in the dorsal mesendoderm. Later on, at the late neurula stage, transcripts are highly expressed in the mesoderm and, thereafter, by the premouth stage, expression is conspicuous in the pharyngeal endoderm, in the anterior tip of the embryo and in the neural tube along the antero-posterior axis of the embryo (**Figure 32**). Moreover, our RNA-seq data showed a strong downregulation of this gene expression suggesting its direct role as an effector of the FGF signal in amphioxus for the formation of anterior somites.

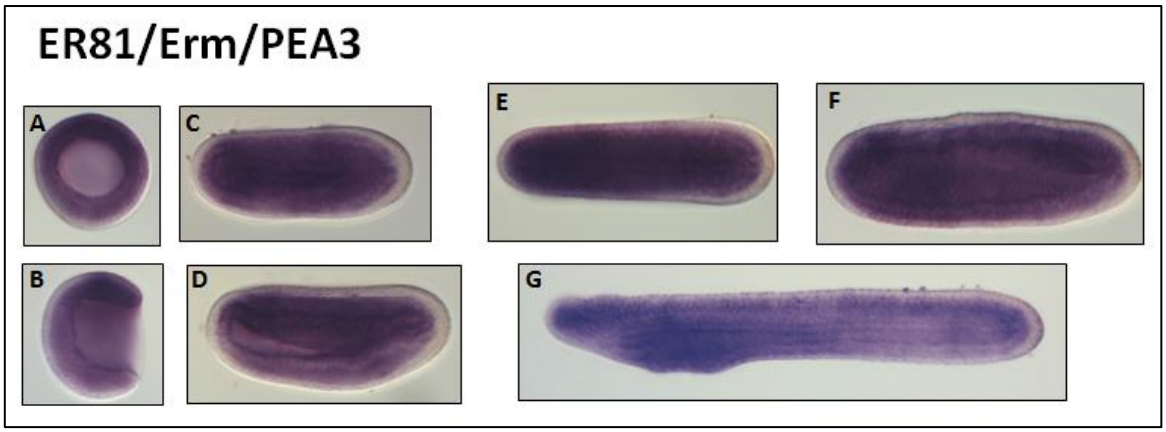


Figure 32. Expression pattern of ER81/Erm/PEA3. (A-B) Gastrula stage, **(C-D)** Early neurula stage, **(E-F)** Late neurula stage, **(G)** Premouth stage. Blastoporal view in A, lateral view in B, D, F and G and dorsal view in C and E. Anterior part of the embryo to the left.

To test the role of specific key transcription factors in the formation of anterior somites in amphioxus we constructed protein chimeras, as described in Article 2. Thus, most of the embryos injected with the constitutive repressor form of *ER81/Erm/PEA3* showed a phenotype similar to embryos treated with SU5402 (Fig 3c-d, Article2). Moreover, embryos injected with the constitutive activator form of *ER81/Erm/PEA3* and then treated with the inhibitor of the FGF signal at blastula stage, showed a mild rescue of the anterior somites formation (data not shown). However, for embryos injected with Eng-EEP, we also observed a range of different phenotypes, with embryos lacking *MLC* expression (**Figure 33**) or *Brachyury* expression in the notochord (**Figure 33**) at late stages. Since *ER81/Erm/PEA3* responds to the FGF signal, we propose that, during early development this transcription factor acts as the effector of the FGF signal for the formation of the anterior somites. This result is supported by our observations at late gastrula stage in embryos injected with Eng-ER81/Erm/PEA3 in which the expression of the genes *Six1/2*, *MRF2* and *Pax3/7* was similar to the expression in embryos treated with SU5402 (Fig. 4, Article 2). Then, at later stages *ER81/Erm/PEA3* could be acting through another pathway than FGF signal, which can explain the strong phenotype observed in some injected embryos (**Figure 33**). Altogether, these data demonstrate the link between the FGF signal and its effector (*ER81/Erm/PEA3*) for the formation of anterior somites in amphioxus at early stages and suggest that *ER81/Erm/PEA3* could also be necessary at later stages for other embryonic processes.

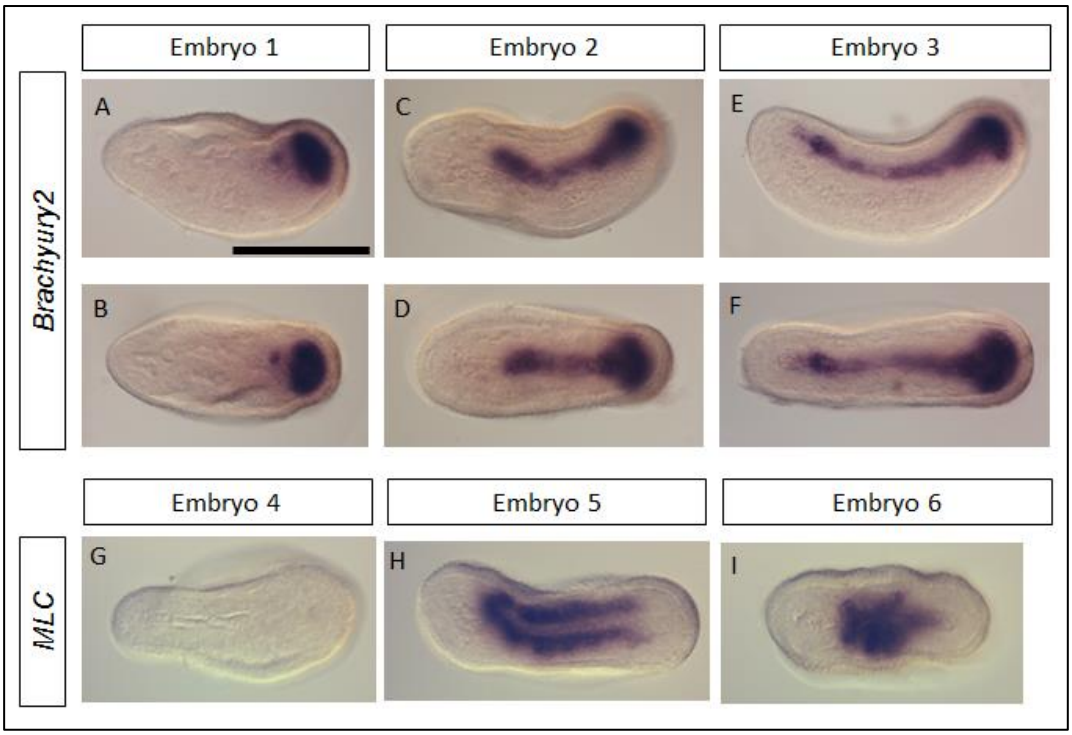


Figure 33. Embryos injected with the constative repressor form of *ER81/Erm/PEA3* (EEP) and fixed at late neurula stage. The different phenotypes obtained when we injected Eng-EEP are showed in this panel. We used *Brachyury* and *MLC* as markers. We observed several phenotypes, from mild to strong. A strong phenotype in some embryos that completely lost the mesoderm (Embryo 1 and 4), having only undifferentiated mesendermal cells at the posterior tip of the embryo **(A-B)**. We also observed a majority of embryos that lost the anterior somites (embryos 5 and 6). Embryo 2 shows a partial formation of the notochord and in embryo 3 the notochord is observed all along the antero-posterior body axis. Lateral views A, C, E. Dorsal views B, D, F, G, H, and I. Anterior part of the embryo is to the left. Scale bar = 250 µm.

4.3 Possible role of upregulated genes in anterior somitogenesis

Interestingly, our comparative RNA-seq analysis reveals that at early treatment 2716 contigs were upregulated compared to the 1677 downregulated contigs (Fig. 1, Article 2). During my thesis I was mainly focused on the study of downregulated genes, because they are genes that putatively control expression of other genes implicated in anterior somitogenesis. Thus, we did not undertake a deep study of upregulated genes. However, two genes drawn our attention because they were highly upregulated (Table 1, Article 2). *FoxEa*, which in amphioxus is expressed in the pharyngeal endoderm (Figure 3, Annex Article 1) and which function could have played an important role in thyroid gland evolution (Mazet, 2002), and *Scratch*, which is expressed in the lateral sides of the amphioxus neural tube (**Figure 34**) and which in vertebrates has a neural-specific role (Dam et al., 2011; Nakakura et al., 2001). When we performed *in situ* hybridization for these genes in wild type embryos, they were not expressed at early stages when anterior somites start to be formed. In addition, *in situ* hybridization in treated embryos did not show any clear upregulation of their expression (data not shown). Thus, no conclusion can be drawn from these data and further analysis will be necessary to define the role of the upregulated genes in the

formation of the anterior somites in amphioxus. Are these genes implicated in the formation of the endoderm or the axial mesoderm? One important question remaining is the fact that the axial mesoderm does not disappear after FGF signal inhibition and there is no indication on what do the paraxial mesoderm cells of the gastrula become. Do they become axial mesoderm, do they become endoderm? Further analyses using cell lineage approaches will be necessary to answer this question.

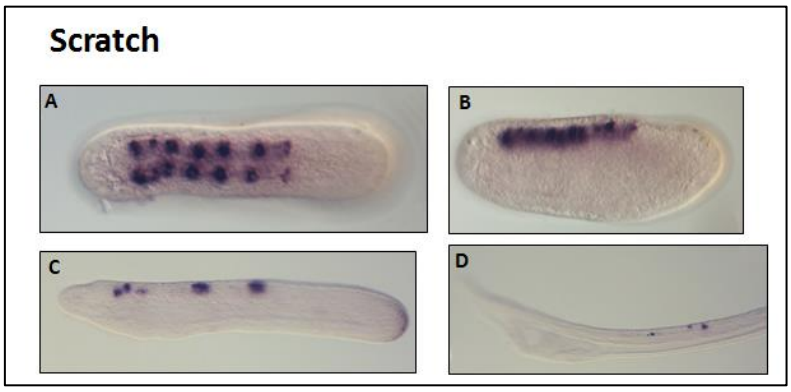


Figure 34. Expression pattern of *Scratch* during amphioxus development. (A-B) Late-neurula stage (N3), **(C)** Premounth stage, **(D)** Larva stage. Dorsal view in A, lateral view in B, D and F. Anterior part of the embryo is to the left.

4.4 Six1/2 and Pax3/7 control somitogenesis in amphioxus

Our results suggest that *ER81/Erm/PEA3* is the effector of the FGF signal for the formation of the anterior somites. So, during my work, the next step was to investigate the downstream target genes of this transcription factor for the formation of the anterior somites in amphioxus. As we showed in Article 2, we observed that *Six1/2* plays a crucial role in the formation of the FGF-sensitive somites and that *Pax3/7* is necessary for the formation of the posterior somites that form by schizocoely in amphioxus (Fig. 3, Article 2). The SIX family includes the genes *Six1/2*, *Six4/5* and *Six3/6* and they act together with their cofactor *Eya* in different developmental processes (Kozmik et al., 2007). Our RNA-seq and *in situ* hybridization data show that *Six1/2*, *Six4/5* and *Eya* are under the control of the FGF signal at earlies stages. In vertebrates, *Six1/2* acts in conjunction with *Six4/5* to activate the myogenic program (Daubas and Buckingham, 2013; Santolini et al., 2016). In amphioxus, coincidently at gastrula stage the expression pattern of *Six1/2* and *Six4/5* overlap, but then from early neurula stage *Six4/5* start to be expressed only in the most posterior forming somites, whereas *Six1/2* is expressed in all the somites (Kozmik et al., 2007). In mice, it has been shown that *Six4* knockout (KO) do not show developmental defects, while *Six1* KO (Grifone et al., 2005) show developmental defects. However, the double KO *Six4*:*Six1* exhibits stronger developmental defects than *Six1* KO. Differential DNA-binding specificity explains this difference in vertebrates (Ando et al., 2005). Therefore, it could be interesting to investigate the role of *Six4/5* in the development of amphioxus and test whether *Six4/5* plays an essential role in the formation of anterior somites in amphioxus or not.

The development is finely controlled by different signaling pathways, which in turn control gene expression at the right time and place. During the last years, studies of *cis*-regulatory elements (CREs) have demonstrated their important role in this control. Thus, CREs are able to drive the expression of a specific gene in different tissues and at different time. In vertebrates, it has been shown that different CREs located upstream of the starting site of *Six1* are able to drive the differential expression pattern of this gene during development (Sato et al., 2012). Thus, they observed specific enhancer activity in somites, cranial mesoderm, endoderm, notochord or cranial placodes covering almost all the tissues in which *Six1* is expressed (Sato et al., 2012). We performed an ATAC-seq analysis to investigate the open chromatin regions in the genome of amphioxus at three different developmental stages. When we looked at the genetic landscape surrounding the gene *Six1/2*, we observed the presence of open chromatin territories in the region located upstream of this gene (Fig. 6, Paper2). Thus, it could be interesting to dissect and test the role of these putative enhancers *in vivo*. In the Article 2, we showed how a specific genomic region was able to drive the expression of *Zic*, validating this kind of approaches.

Certainly, an interesting experiment to perform in the future could be a ChIP-seq using a specific antibody for ER81/Erm/PEA3 to corroborate our findings and search for new candidate genes involved in anterior somitogenesis. Moreover, this could help us better understand our observations that some embryos injected with Eng-EEP completely loss the expression of *MLC*, and therefore both axial and paraxial mesoderm.

4.5 FGF signal and vertebrates head mesoderm

In chick embryos, patterning of the head mesoderm depends on the antagonistic roles of the FGF and BMP signaling pathways together with RA as a suppressor signal (Bothe et al., 2011). These signals in turn control the expression of master genes such as *Pitx2*, *Tbx1*, *Alx4*, and *MyoR* (Bothe and Dietrich, 2006). Moreover, it has been shown that the head mesoderm in chick embryos is regionalized into two territories. An anterior territory marked by the expression of *Pitx2* and a posterior territory, adjacent to the first one, where muscles start to be formed following the expression of *Tbx1*. Interestingly, FGF signal is essential to activate the posterior head mesoderm expression of *Tbx1* (Abu-Issa et al., 2002; Bothe et al., 2011). There is no study showing which signaling pathway triggers the expression of *Pitx2* in the anterior territory of the head mesoderm, nevertheless it has been shown that low levels of the FGF signal induce the expression of *Pitx2* in the anterior territory of the head mesoderm (Bothe et al., 2011). In amphioxus, *Tbx1/10* is expressed starting from the late neurula stage in the ventral part of the first 8-10 somites (Mahadevan et al., 2004). Besides, *Pitx* is involved in the left-right asymmetry and is expressed only in the left side of the ectoderm, mesoderm and endoderm starting from early neurula stage (Boorman and Shimeld, 2002). From our RNA-seq data of the early treatment experiment *Pitx* is upregulated at 6hpt and *Tbx1/10* expression is not modified after FGF signal inhibition (Table 1, Article 2). Recent studies of *Tbx1/10* knock-down in amphioxus were performed through morpholino injection (Koop et al., 2014). The authors observed a smaller pharynx and fused gill slit, however they did not notice any difference in the 8-10 anterior somites (Koop et al., 2014). Altogether, these data allow us to propose that *Tbx1,* together with *Pitx2* were co-opted in vertebrates to control cranial myogenesis when the head mesoderm appeared secondarily.

4.6 Our results under the context of the evolution of the vertebrates' head

Functional changes in signaling pathways have been instrumental in the appearance of morphological novelties (Pires-daSilva and Sommer, 2003; Wagner and Lynch, 2010). For instance, in vertebrates, ectopic activation of the RA signal leads to the truncation of the rostral part of the embryo (Holder and Hill, 1991; Kuratani et al., 1998; Morriss-Kay et al., 1991; Papalopulu et al., 1991). Moreover, Kuratani and colleagues observed, in lampreys embryos, severe phenotypes after treatment with RA which induce extension of anterior segments from the most anterior to the most posterior part of the embryo (Kuratani et al., 1998). Furthermore, simple inhibition of the FGF signal leads to the loss of anterior segments in amphioxus (Bertrand et al., 2011). Comparative gene expression patterns between lampreys, gnathostomes and amphioxus, led segmentalists to propose that anterior somites of amphioxus evolved into the existent head mesoderm of vertebrates (Holland et al., 2008b). However, their ideas are only based on observations and no functional analyses were undertaken to test their hypothesis. This work, together with previous publications of our laboratory, has deciphered the role of the FGF signal in the formation of the most anterior somites (this work) (Bertrand et al., 2011). Remarkably, we demonstrated that *Pax3/7* is necessary for the formation of the posterior somites of amphioxus (those formed by schizocoely). In lamprey, *Pax3/7* is not expressed in the head mesoderm (that is located anterior to the otic vesicle) but is expressed in somite-derived skeletal muscles (Kusakabe et al., 2011). Likewise, in gnathostomes, *Pax3* is expressed in all the somites but not in the head mesoderm (Schubert et al., 2001a). Additionally, *Six1/2* is essential for the formation of the FGF-sensitive somites in amphioxus (Fig 3, Article 2). In vertebrates, *Six1* plays a pivotal role in myogenesis (Buckingham and Rigby, 2014). Moreover, *Six1* is expressed in the posterior head mesoderm and pharyngeal pouches in contrary to *Pax3/7* that is not expressed in the head mesoderm (Sato et al., 2012; Zou et al., 2006). Thus, is possible to imagine an ancient compartmentalization, where *Pax3/7* sensitives-somites conserved its role in the trunk of the vertebrates, whereas at the place of the anterior *Six1/2* sensitives-somites the unsegmented head mesoderm emerged but conserving the expression of *Six1/2* as is still observed in anterior structures of extant vertebrates (Guo et al., 2011; Lin et al., 2009; Relaix et al., 2013).

Altogether, these data allow us to propose that if the body of the ancestor of all chordates was completely segmented as it is widely accepted, then changes in the FGF signal led to the loss of the anterior paraxial mesoderm, relaxing developmental constraints in the most anterior part of the embryo and allowing the acquisition of the unsegmented head mesoderm. This new head mesoderm co-opted *Pitx* as the master gene for the formation of head muscles in the anterior part, and *Tbx1* in the posterior part adjacent to the forming somites. FGF signal activates *Tbx1* and suppresses *Pitx2* expression in head mesoderm of chick embryos (Bothe et al., 2011), then we could imagine that the loss of anterior FGF signal facilitated the co-option of *Pitx* in the anterior head mesoderm. We observed in our RNA-seq analysis that *Tbx1/10* is downregulated after late treatment with SU5402 (Table 1, Article 2), suggesting a putative control of *Tbx1/10* by FGF in amphioxus at late stages. All these changes were accompanied by the two rounds of whole genome duplication (2RWGD) (Ohno, 1970) that brought not only new genes to this scenario, but also new cis-regulatory elements to control gene expression in specific tissues (**Figure 35**).

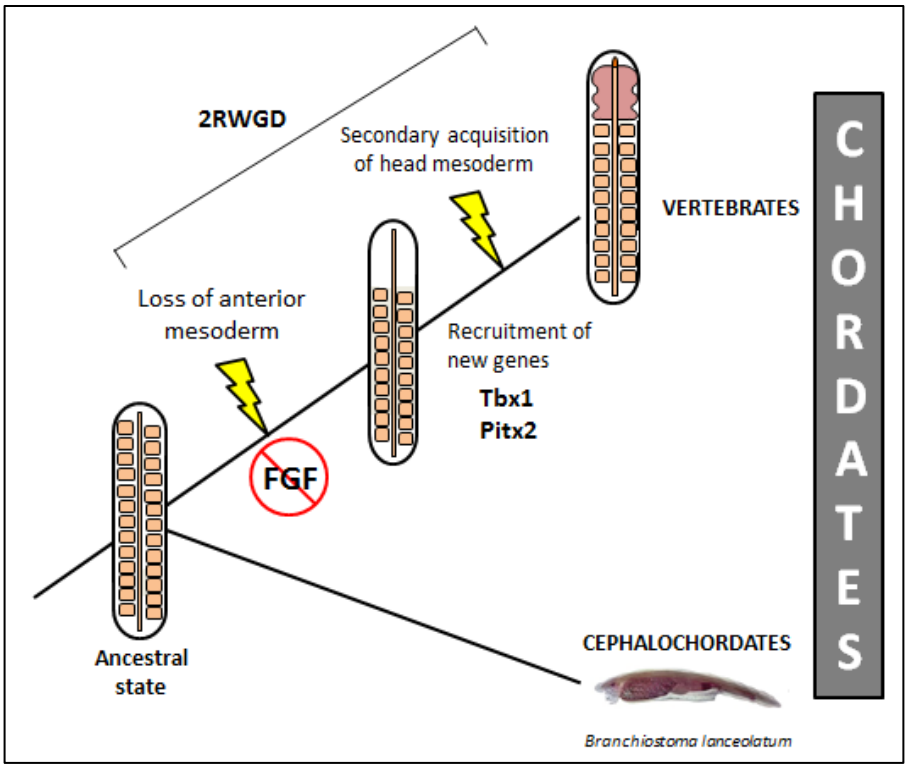


Figure 35. Proposed evolutionary scenario. It is extensively accepted that the last common ancestor of chordates possessed a completely segmented body (ancestral state). Then, changes in the FGF signal led to the loss of anterior somites releasing anterior developmental constraints and therefore allowing the appearance of a new head mesoderm through the recruitment of *Tbx1* and *Pitx2* as master genes.

5. Annex Articles

In this section I have included two articles that were part of my work during my thesis, although they are not completely related with the central topic of my research, which is the origin of the vertebrates' head. In the first article "Expression of Fox genes in the cephalochordate *Branchiostoma lanceolatum*", we analyzed the expression pattern of several Fox genes during amphioxus development, as well as the phylogenetic relationships of Fox amphioxus genes within the Fox family that is composed by 24 classes (ranging from FoxA to FoxS). Importantly, Fox genes are involved in pivotal developmental processes, and they are present in fungi as in metazoans. Finally, a comparative analysis within the chordates supports a well-conserved expression domain for some Fox genes but also some divergent expression domains. This suggests that functional evolution of some Fox genes was essential for the evolution of new characters.

The second article entitled "A single three-dimensional chromatin compartment in amphioxus indicates a stepwise evolution of vertebrate Hox bimodal regulation" shows that the amphioxus How cluster is contained within a topologically associating domain (TAD) while in vertebrates, the Hox cluster is surrounded by two TADs located at the 5' and 3' regions of the Hox genes. These TADs in vertebrates play essential roles in the control of Hox gene expression in the during limb development. Our results in amphioxus suggest a stepwise evolution in the bimodal control of the Hox genes in vertebrates in which no TADs existed in pre-chordate metazoans, a single TAD appeared in amphioxus in which long range regulatory elements acquired specific functions in the control of Hox gene expression and then a second TAD appeared in vertebrates together with the bimodal control of Hox gene expression in the limbs.

5.1 Expression of Fox genes in the cephalochordate *Branchiostoma lanceolatum*

frontiers in Ecology and Evolution

ORIGINAL RESEARCH
published: 28 July 2015
doi: 10.3389/fevo.2015.00080

Expression of Fox genes in the cephalochordate *Branchiostoma lanceolatum*

Daniel Aldea, Anthony Leon, Stephanie Bertrand* and Hector Escriva*

Centre National De La Recherche Scientifique, UPMC université Paris 06, UMR 7232, BIOM, Observatoire Océanologique de Banyuls sur Mer, Banyuls sur Mer, France

Forkhead box (Fox) genes code for transcription factors that play important roles in different biological processes. They are found in a wide variety of organisms and appeared in unicellular eukaryotes. In metazoans, the gene family includes many members that can be subdivided into 24 classes. Cephalochordates are key organisms to understand the functional evolution of gene families in the chordate lineage due to their phylogenetic position as an early divergent chordate, their simple anatomy and genome structure. In the genome of the cephalochordate amphioxus *Branchiostoma floridae,* 32 Fox genes were identified, with at least one member for each of the classes that were present in the ancestor of bilaterians. In this work we describe the expression pattern of 13 of these genes during the embryonic development of the Mediterranean amphioxus, *Branchiostoma lanceolatum*. We found that *FoxK* and *FoxM* genes present an ubiquitous expression while all the others show specific expression patterns restricted to diverse embryonic territories. Many of these expression patterns are conserved with vertebrates, suggesting that the main functions of Fox genes in chordates were present in their common ancestor.

Keywords: Fox genes, amphioxus, Evo-Devo, chordates, embryonic development

OPEN ACCESS

Edited by:
Naoki Osada,
Hokkaido University, Japan

Reviewed by:
Jr-Kai Yu,
Academia Sinica, Taiwan
Haruki Ochi,
Yamagata University, Japan

****Correspondence:***
Stephanie Bertrand and Hector Escriva,
Laboratoire Arago, Avenue du Fontaulé, 66650 Banyuls-sur-Mer, France
stephanie.bertrand@obs-banyuls.fr; hescriva@obs-banyuls.fr

Specialty section:
This article was submitted to Evolutionary and Population Genetics, a section of the journal Frontiers in Ecology and Evolution

Received: *29 May 2015*
Accepted: *07 July 2015*
Published: *28 July 2015*

Citation:
Aldea D, Leon A, Bertrand S and Escriva H (2015) Expression of Fox genes in the cephalochordate Branchiostoma lanceolatum. Front. Ecol. Evol. 3:80. doi: 10.3389/fevo.2015.00080

Introduction

Forkhead box (Fox) transcription factors originated early during evolution and are specific to opisthokonts. They are present in fungi as well as in metazoans (Mazet et al., 2006; Larroux et al., 2008; Shimeld et al., 2010a) in which they play essential roles during embryonic development (Carlsson and Mahlapuu, 2002; Tuteja and Kaestner, 2007a,b; Benayoun et al., 2011). Fox proteins possess a helix-turn-helix DNA-binding domain called the forkhead domain which corresponds to a conserved region of approximately 110 amino acids (Weigel and Jackle, 1990; Clark et al., 1993). A molecular phylogeny-based classification of the Fox gene family allowed to propose its subdivision into 24 classes (ranged from FoxA to FoxS and including subfamilies that were recently subdivided: FoxJ (FoxJ1 and FoxJ2), FoxL (FoxL1 and FoxL2), and FoxN (FoxN1/4 and FoxN2/3) (Mazet et al., 2003). Many Fox gene losses or duplications occurred in different bilaterian clades, affecting different Fox classes. For example, FoxAB is found in cephalochordates and in the sea urchin but not in tunicates or vertebrates (Tu et al., 2006; Yu et al., 2008a), and families R and S are vertebrate-specific (Wotton and Shimeld, 2006; Shimeld et al., 2010b). Using phylogenetic analyses, it has been proposed that 22 Fox gene families were already present in the bilaterian ancestor (Shimeld et al., 2010b).

Cephalochordates (i.e., amphioxus) belong to the chordate phylum together with tunicates and their sister group, the vertebrates. They present morphological, developmental, and genomic characteristics that are proposed to be very similar to the ancestral state in the chordate clade, making amphioxus a key model system to understand chordate evolution (Bertrand and Escriva, 2011, 2014). Interestingly, it has been shown that amphioxus is the only living bilaterian possessing at least one member of each of the 22 Fox gene families proposed to have been present in Urbilateria (Yu et al., 2008a). Thus, the study of Fox genes in this cephalochordate may shed light on the functional evolutionary history of this transcription factor gene family. Past studies using genomic data from the Caribbean cephalochordate *Branchiostoma floridae* described the presence of 32 Fox genes in this species (Yu et al., 2008a) and the expression pattern of 11 of these genes was previously described: *FoxAa* and *FoxAb* (formerly named AmHNF3-1 and AmHNF3-2, respectively) (Shimeld, 1997), *FoxB* (Mazet and Shimeld, 2002), *FoxC* (Mazet et al., 2006), *FoxD* (Yu et al., 2002b), *FoxE4* (Yu et al., 2002a), *FoxF* (Mazet et al., 2006; Onimaru et al., 2011), *FoxG* (Toresson et al., 1998), *FoxL1* (Mazet et al., 2006), *FoxN1/4a* (Bajoghli et al., 2009), *FoxQ1* and *FoxQ2* (Yu et al., 2003; Mazet et al., 2006). In this work we searched for Fox sequences in the transcriptome of the Mediterranean amphioxus *Branchiostoma lanceolatum*. We found 28 Fox sequences and we describe here the spatiotemporal expression pattern of 13 Fox genes during embryonic development, including seven previously described in *B. floridae* and six for which expression was not known. We show that in *B. lanceolatum* some Fox genes exhibit ubiquitous expression as *FoxK* and *FoxM*, while the others show specific and dynamic expression patterns restricted to diverse embryonic territories. These expression patterns suggest that Fox genes are performing both general and specific functions during amphioxus embryonic development, most of them being probably ancestral in the chordate clade.

Materials and Methods

Phylogenetic Analysis

All reference sequences, except for *B. lanceolatum*, were obtained from Genbank or from Fritzenwanker et al. (2014) The multiple alignment was performed only for the conserved Forkhead amino acid domain sequences using the MUSCLE module implemented in MEGA 6 and manually refined in its interface (Tamura et al., 2013). The best fit substitution model for phylogenetic reconstruction was estimated using MEGA 6 (Tamura et al., 2011). Bayesian inference (BI) tree was inferred using MrBayes 3.2 (Ronquist et al., 2012), with the model recommended by MEGA 6 under the Akaike information criterion (RtRev+Γ), at the CIPRES Science Gateway V. 3.1 (Miller et al., 2015). Two independent runs were performed, each with four chains and 1 million generations. A burn-in of 25% was used and a 50 majority-rule consensus tree was calculated for the remaining trees.

Cloning and Expression Study

B. lanceolatum Fox sequences were recovered from its reference transcriptome (Oulion et al., 2012) by TBLASTN using sequences from *B. floridae* as queries. Specific primers were then designed for RT-PCR amplification from total RNA. Primer sequences are as follow:

FoxA_a_5′ AAGTCGCCGGTGTACGAGATG
FoxA_a_3′ GTATTATAGAGACGAAGGTTG
FoxA_b_5′ CATTTCCTCAGAACAGACATG
FoxA_b_3′ TCCTAAAGACTCCCAACAACA
FoxAB_5′ CAGTGTGAGGTGAACATCATG
FoxAB_3′ CGATTGACAGGTTGATAGAAC
FoxB_5′ ACAACAGGACCCTGACTCGT
FoxB_3′ GCATTCCCTGACGTCTTGA
FoxC_5′ AACCGTCCCGTTTTCCTCATG
FoxC_3′ CAGTTTTGATTCGTAAGGACT
FoxD_5′ ACAGCTGTGGAGTGGACACTT
FoxD_3′ CACGAGACATGTAAGTCTCCG
FoxEa_5′ AACCAACCCCGTACCAGCATG
FoxEa_3′ ATATGACACGGACACTGAACT
FoxG_5′ ACGCACATTAGCACAGTTCG
FoxG_3′ ACTTGACCCTGGCTTGACAC
FoxJ1_5′ TACAGACAACTGTAAACCATG
FoxJ1_3′ TTGTAATGCAGGGTGGGGCCT
FoxK_5′ GGAAGGCGGAGTTGGACAATG
FoxK_3′ CCGGACACGTCCTGCACCTGT
FoxM_5′ AGGAGAGTGTGACAAACCATG
FoxM_3′ TTCTCAGCTATTCAGTAATAC
FoxN1/4a_5′ GCGCACCGAGTATCGTTCTGA
FoxN1/4a_3′ ACATAGGTAGGACTATGTACT
FoxN2/3_5′ CAGTAAACACGAGCAGACATG
FoxN2/3_3′ AGCTGAAGACAATGATGATCC

A mix of total mRNA of *B. lanceolatum* extracted from embryos at different developmental stages was used as a template for retro-transcription. Amplification was performed using Advantage 2 Polymerase kit (Clontech) and a touch-down PCR program with annealing temperature ranging from 65 to 40°C. Amplified fragments were cloned using the pGEM-T Easy system (Promega) and sub-cloned in pBluescript II KS+ for probe synthesis.

Whole Mount *In situ* Hybridization

Probes were synthesized using the DIG labeling system (Roche) after plasmid linearization with the appropriate enzymes. Ripe animals of *B. lanceolatum* were collected in Argelès-sur-Mer (France), and gametes were obtained by heat stimulation (Fuentes et al., 2004, 2007). *In vitro* fertilization was undertaken in Petri dishes filled with filtered sea water. Fixation and whole mount *in situ* hybridization were performed as described in Somorjai et al. (2008).

Results

Molecular Phylogenetic Analysis of *B. lanceolatum* Fox Gene Sequences

We looked for Fox gene sequences in the reference transcriptome of *B. lanceolatum* (Oulion et al., 2012). The sequences that were recovered were used to conduct a phylogenetic

tree reconstruction presented in **Figure 1**. We showed that *B. lanceolatum* possesses at least 28 Fox genes, each of them being orthologous to one of the 32 genes described in *B. floridae* and corresponding to at least one member of each of the 22 families present in the bilaterian ancestor (Yu et al., 2008a). Specific duplications, that occurred in the cephalochordate clade at least in the ancestor of *B. floridae* and *B. lanceolatum*, gave rise to three members in the FoxQ2 group (*FoxQ2a*, *FoxQ2b*, *FoxQ2c*), two members in the FoxN1/4 group (*FoxN1/4a* and *FoxN1/4b*), and two genes in the FoxE group (*FoxEa* and *FoxEc*). We then analyzed the expression pattern during *B. lanceolatum* embryonic development of 13 of these 28 Fox genes corresponding to those showing a higher expression level in the transcriptome (Oulion et al., 2012).

FoxAa and *FoxAb*

FoxAa (formerly named *AmHNF3-1*) (Shimeld, 1997) was first expressed at the gastrula stage in the anterior ventral endoderm and in the mesendodermal layer of the dorsal blastoporal lip (**Figures 2A,B**). At the late gastrula stage, we detected transcripts in the axial dorsal mesendoderm corresponding to the presumptive notochord territory, as well as in mesendoderm cells of the archenteron floor (**Figures 2C,D**). Expression in the axial mesoderm and endoderm persisted through mid-late neurula stage (**Figures 2E,F**). Later on, at late neurula stage before the mouth opens, the expression in the notochord was restricted to the most anterior and posterior tips of the embryo, while the endodermal expression was restricted to the middle region of the gut (**Figure 2G**). At the larva stage, the expression at the anterior tip of the notochord and in the tailbud was still observed and we detected a diffuse expression in the gut (**Figure 2H**).

FoxAb (formerly named *AmHNF3-2*) (Shimeld, 1997) expression was first detected at the gastrula stage as a weak signal in the mesendodermal part of the dorsal blastoporal lip (**Figures 2I,J**). At the late gastrula stage, we detected expression in the central paraxial mesoderm on both sides of the notochord anlagen (**Figures 2K,L**). At the mid-late neurula stage transcripts were detected in the neural tube, including the cerebral vesicle, and in the dorsal part of the endoderm (**Figures 2M,N**). At the late neurula stage, before the mouth opens, *FoxAb* was expressed in the neural tube and in the most anterior part of the pharynx. In the posterior region, expression was detected in the tailbud and in the dorsal midline of the gut (**Figure 2O**). At the larva stage, we observed expression in the pharynx, in the preoral pit, in the club-shaped gland and in the tailbud. At this stage, the expression in the neural tube gets restricted to some neurons and to the posterior part of the cerebral vesicle (**Figure 2P** and Figure S1A).

FoxAB

FoxAB transcripts were detected as a weak and ubiquitous signal from the eight-cell stage to the blastula stage (**Figures 2Q,R**). This ubiquitous expression was confirmed by the presence of reads in transcriptome analyses (data not shown). At the gastrula stage we observed a strong specific expression in the dorsal blastoporal lip, the amphioxus putative organizer (**Figures 2S,T**). At the late gastrula stage, expression gets restricted to the presumptive notochord territory (**Figures 2U,V**). No expression could be detected by *in situ* hybridization in later stages.

FoxB

FoxB expression was first detected dorsally, both in the ectoderm and in the mesendoderm, as a weak signal in mid gastrula stage embryos (**Figures 2W,X**). Later on, in early neurula stage embryos, a signal could be observed in the neural plate on either side of the midline, as well as in two patches in the posterior paraxial mesendoderm (**Figures 2Y,Z**). During the late neurula stage, expression was detected in the most posterior paraxial mesoderm that give rise to the newly formed somites and in the neural tube posterior to the cerebral vesicle (**Figures 2A',B'**). Then, *FoxB* expression in the mesoderm faded away in late neurulae (**Figure 2C'**) and get later restricted to the cerebral vesicle and to some neurons along the neural tube in larvae (**Figure 2D'** and Figure S1B).

FoxC

FoxC was expressed at the gastrula stage in the dorsal paraxial mesendoderm (**Figures 3A,B**). Later on, at the late gastrula stage, expression was detected in the region that gives rise to the three most anterior somites (**Figures 3C,D**). In mid-late neurulae, the transcripts remained all along the body in the somites and a new expression domain appeared in the anterior endoderm at the level where the first gill slit opens (**Figure 3E**). At the late neurula stage, the expression persisted in the pharynx and somites and was also detected in the club-shaped gland anlagen (**Figures 3F,G**). At the larva stage a diffuse expression was observed in the somites as well as in the preoral pit, in the club-shaped gland and in the first gill slit (**Figure 3H** and Figure S1C).

FoxD

FoxD transcripts were first detected at the gastrula stage in the dorsal blastoporal lip (**Figures 3I,J**). Then, at the late gastrula stage, *FoxD* was expressed in the dorsal axial mesendoderm, in part of the dorsal paraxial mesendoderm as two patches on both sides of the midline and in the anterior region of the neural plate (**Figures 3K,L**). At the mid-late neurula stage, the notochord and the somites, as well as the cerebral vesicle, were labeled (**Figure 3M**). At the late neurula stage, before the mouth opens, transcripts were detected in the paraxial somitic mesoderm, in the notochord, in the cerebral vesicle and in the posterior endoderm (**Figures 3N,O**). A faint labeling was also detected at this stage in the first gill slit and in the club-shaped gland anlagens. At the larva stage, we observed a low expression level in the cerebral vesicle, in the preoral pit, in the club-shaped gland, in the first gill slit, in the notochord and in the posterior part of the gut. We also observed an anterior to posterior gradient of expression in the somites (**Figure 3P** and Figure S1D).

FoxEa

FoxEa (formerly named *FoxE4* in *B. floridae*) expression was first detected at early neurula stage in the antero-ventral mesendoderm (**Figures 3Q,R**). Later on, at the mid-late neurula

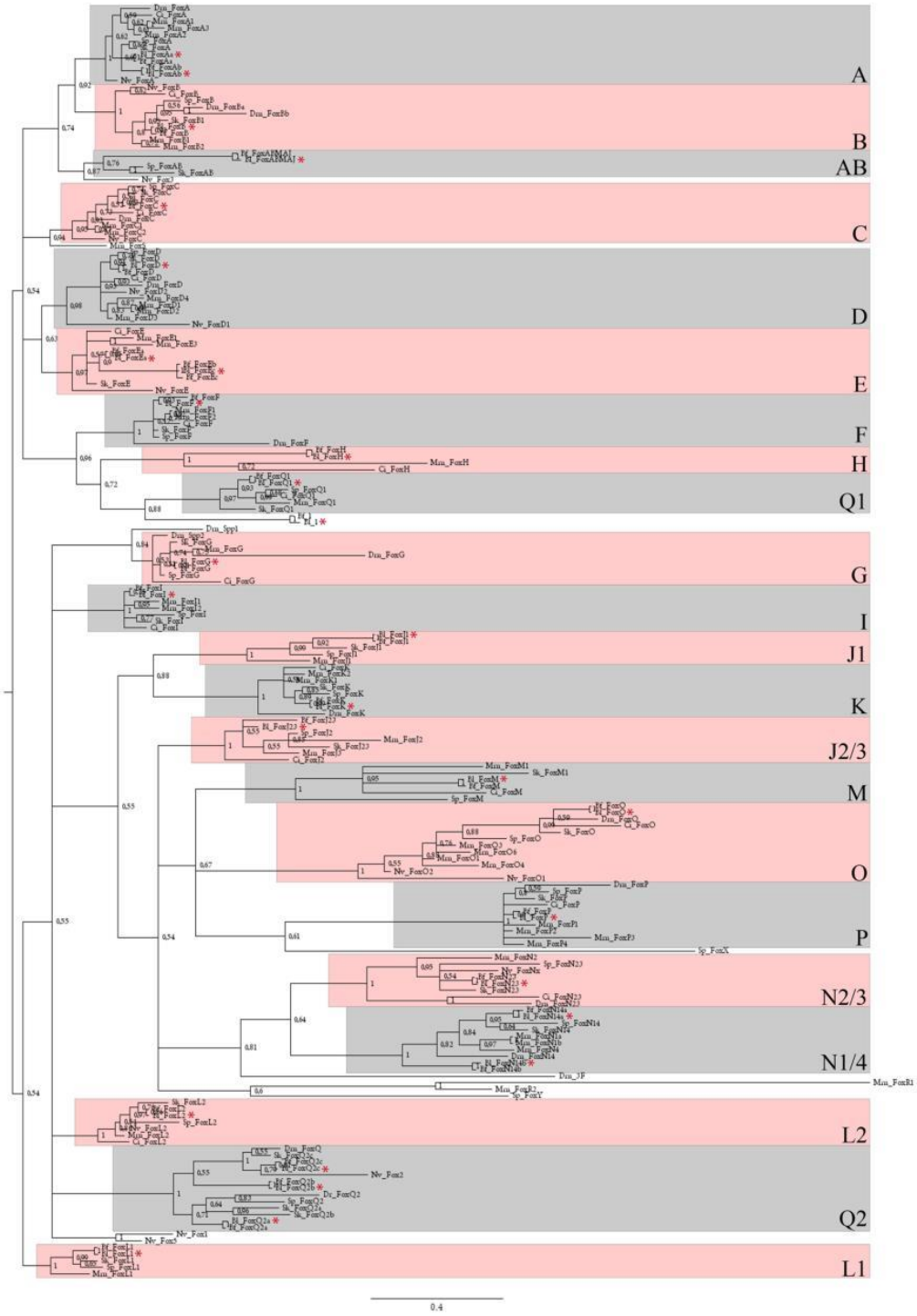


FIGURE 1 | Phylogenetic analysis of *B. lanceolatum* Fox genes. Unrooted 50 majority-rule consensus Bayesian inference tree based on the amino acid sequences of the forkhead domain. Posterior probablilities are shown at each node. The different paralogy groups are colored in pink or light blue boxes. Divergent sequences appeared outside these boxes. Only one amphioxus Fox gene, named Fox1 (Yu et al., 2008a), that probably originated by a specific duplication and fast evolutionary rate in cephalochordates,

(Continued)

FIGURE 1 | Continued
localizes outside these paralogy groups. Abbreviations: Dm, *Drosophila melanogaster*; Mm, *Mus musculus;* Dr, *Danio rerio;* Ci, *Ciona intestinalis;* Sp, *Strongylocentrotus purpuratus;* Sk, *Saccoglossus kowalevskii;* Nv, *Nematostella vectensis;* Bf, *Branchiostoma floridae;* Bl, *Branchiostoma lanceolatum.* Red stars indicate Bl sequences. Scale bar represents 0.4 amino acid substitution per site.

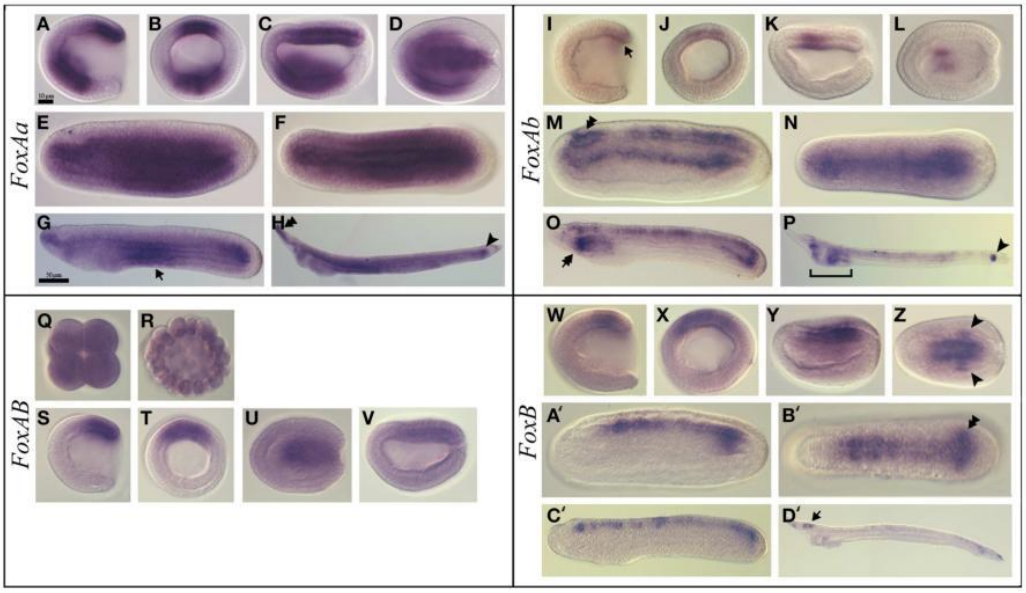


FIGURE 2 | Expression of *B. lanceolatum FoxAa, FoxAb, FoxAB, and FoxB*. In all the panels except **(B, J, Q, R, T, X)** anterior is to the left. In lateral and blastoporal views dorsal is to the top. *FoxAa* expression pattern **(A–H)**. Gastrula lateral **(A)** and blasporal **(B)** views. Late gastrula lateral **(C)** and dorsal **(D)** views. Mid-late neurula lateral **(E)** and dorsal **(F)** views. In the late neurula lateral view **(G)** arrow marks the endodermal expression in the middle region. In the larva stage lateral view **(H)**, the double arrowhead indicates the expression in the anterior tip of the notochord and the arrowhead marks the expression in the tailbud. *FoxAb* expression pattern **(I–P)**. In the gastrula lateral **(I)** and blastoporal **(J)** views the arrow indicates the expression in the mesendodermal part of the dorsal blastoporal lip. Late gastrula lateral **(K)** and dorsal **(L)** views. In the mid-late neurula lateral **(M)** and dorsal **(N)** views the double arrowhead marks the expression in the cerebral vesicle. In the late neurula lateral view **(O)**, the double arrow marks the expression in the most anterior part of the pharynx. In larva lateral view **(P)** the arrowhead indicates the expression in the tailbud. *FoxAB* expression pattern **(Q–V)**. Eight-cell stage **(Q)**. Blastula stage **(R)**. Gastrula lateral **(S)** and blasporal **(T)** views. Late gastrula lateral **(U)** and dorsal **(V)** views. *FoxB* expression pattern **(W–D')**. Gastrula lateral **(W)** and blastoporal **(X)** views. Early neurula lateral view **(Y)**. In the early neurula dorsal **(Z)** view the arrowhead indicates the two expression patches in the posterior paraxial mesendoderm. Mid-late neurula lateral **(A')** and dorsal **(B')** views. The double arrowhead marks the expression in the newly formed somites. Late neurula lateral view **(C')**. In larva lateral view **(D')** the arrow indicates the expression in the cerebral vesicle. Scale bar: 10 μm **(A–F), (I–N)**, **(Q-V)**, **(W-B')**, and 50 μm **(G,H)**, **(O,P)**, **(C',D')**.

stage, *FoxEa* transcripts were detected ventrally in the endoderm with a higher expression level on the right side of the pharynx (**Figures 3S,T**), and a slight expression domain in the posterior gut was also visible. At the late neurula stage, *FoxEa* transcripts remained ventrally in the pharyngeal endoderm on the right side (**Figure 3U**). Finally, at the larva stage, transcripts were detected in the club-shaped gland (**Figure 3V** and Figure S1E).

FoxG

FoxG expression was first observed at the neurula stage in the anterior region of the first somites (**Figures 3W,Y**). At the late neurula stage, *FoxG* was expressed in the anterior ventral region of the three most anterior somites (**Figures 3X,Z**). Later on, in late neurula before the mouth opens, a neural expression appeared in some individual neurons within the neural tube, while the expression observed in the first somites disappeared (**Figure 3A'**). This expression persisted in the larva stage embryos in which *FoxG* was also detected in some neurons of the cerebral vesicle (**Figure 3B'** and Figure S1F).

FoxJ1

FoxJ1 showed a dynamic expression pattern. Expression began during gastrulation and was detected in the ectoderm except the ectoderm around the blastopore (**Figures 4A,B**). Later on, at the late gastrula stage, this expression pattern persisted in the ectoderm that give rise to the epidermis (**Figures 4C,D**). At the mid-late neurula stage, we detected transcripts in the neural tube while the expression in the epidermis was completely lost (**Figures 4E,F**). This neural tube expression was no more observed in late neurula stage embryos before the mouth

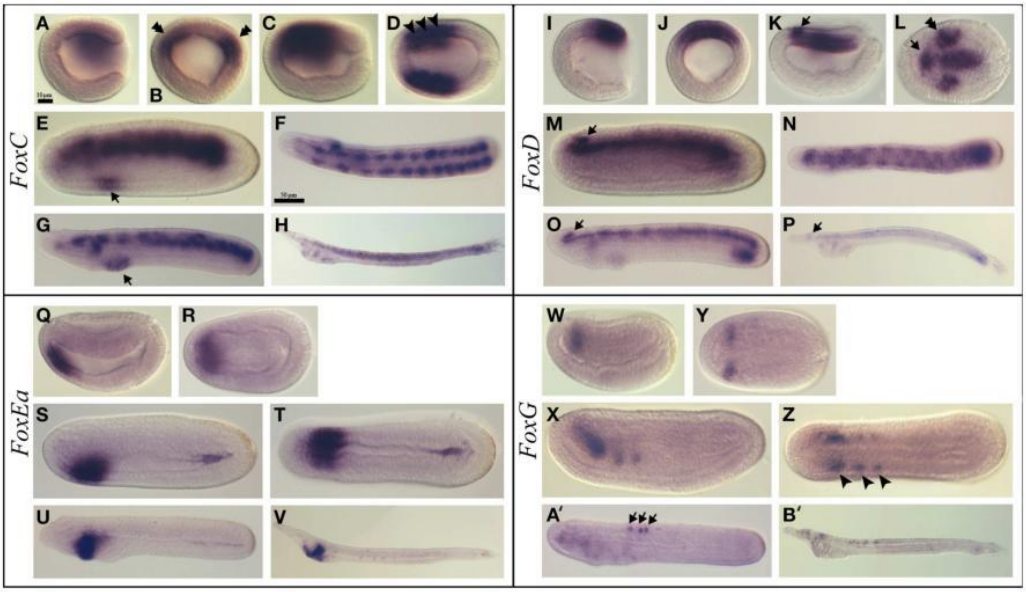


FIGURE 3 | Expression of *B. lanceolatum FoxC*, *FoxD*, *FoxEa*, and *FoxG*. In all the panels except **(B,J)**, anterior is to the left. In lateral and blastoporal views dorsal is to the top. *FoxC* expression pattern **(A–H)**. Gastrula lateral **(A)** and blastoporal **(B)** views. The double arrowhead indicates the expression in the paraxial mesoderm. Late gastrula lateral **(C)** and dorsal **(D)** views. The arrowheads marks the region that will give rise to the three most anterior somites. In mid-late neurula lateral view **(E)** the arrow indicates a new expression domain in the anterior endoderm. Late neurula dorsal **(F)** and lateral **(G)** views. The arrow marks the expression domain in the pharynx. Larva lateral view **(H)**. *FoxD* expression pattern **(I–P)**. Gastrula lateral **(I)** and blasporal **(J)** views. Late gastrula lateral **(K)** and dorsal **(L)** views. The arrow indicates the expression in the anterior region of the neural plate and the double arrowhead marks the expression in the paraxial dorsal mesendoderm. Mid-late neurula lateral view **(M)**. Late neurula dorsal **(N)** and lateral **(O)** views. Larva lateral view **(P)**. In **(M, O, P)** the arrows indicate the expression domain in the cerebral vesicle. *FoxE* expression pattern **(Q–V)**. Early neurula lateral **(Q)** and dorsal **(R)** views. Mid-late neurula lateral **(S)** and dorsal **(T)** views. Late neurula lateral view **(U)**. Larva lateral view **(V)**. *FoxG* expression pattern **(W–B')**. Early neurula lateral **(W)** and dorsal **(Y)** views. Mid-late neurula lateral **(X)** and dorsal **(Z)** views. The arrowhead indicates the expression in the three most anterior somites. In the late neurula stage lateral view **(A')** the arrows mark the neurons within the neural tube. Larva stage lateral view **(B')**. Scale bar: 10 μm **(A–E), (I–L), (Q–T), (W–Z),** and 50 μm **(F–H), (N–P), (U,V), (A',B')**.

opens (data not show), however at the larva stage we observed expression at the anterior tip of the embryo and in the pharynx at the level of the preoral pit and of the first gill slit (**Figure 4G** and Figure S1G).

FoxK

FoxK was ubiquitously expressed from the eight-cell stage to the blastula stage (Figures S2A,B). At the gastrula stage, the expression became restricted to the mesendoderm (Figures S2C,D), and by the late gastrula stage transcripts were detected mostly in the dorsal mesoderm (Figures S2E,F). At the mid-late neurula stage, we detected a stronger expression in the most anterior region of the embryo (Figures S2G,H). Transcripts were then detected in the whole embryo at the late neurula stage with a stronger expression in the anterior tip (Figures S2I,J). Finally, at the larva stage, we observed a ubiquitous expression with a higher level at the anterior tip and in the pharynx (Figure S2K).

FoxM

FoxM transcripts were detected ubiquitously during the whole embryonic development, from the eight-cell stage until the mid-late neurula stage except in the epidermis (Figures S2L–S). Later on, at late neurula stage, *FoxM* expression could not be detected anymore by *in situ* hybridization (Figure S2T).

FoxN1/4a

Ubiquitous *FoxN1/4a* expression was detected from the eight-cell stage until the blastula stage (**Figures 4H,I**). At the gastrula stage, a signal was detected in the anterior ectoderm (**Figures 4J,K**). Later on, at the early neurula stage, we observed transcripts in the anterior endoderm as well as in the axial central mesoderm (**Figures 4L,M**). At the mid-late neurula stage, we detected three major expression domains: one anterior, at the level of the cerebral vesicle, a second one in the anterior ventral endoderm and a third one in the posterior mesoderm (**Figure 4N**). At the late neurula stage before the mouth opens, we observed expression in the anterior and posterior endoderm (**Figure 4O**). Finally, at the larva stage, we detected expression in the posterior region of the gut and in the anus (**Figure 4P**).

FoxN2/3

Ubiquitous expression of *FoxN2/3* was observed from the eight-cell stage (**Figure 4Q**) to the blastula stage (**Figure 4R**). Then, at the gastrula stage, the expression was restricted to the

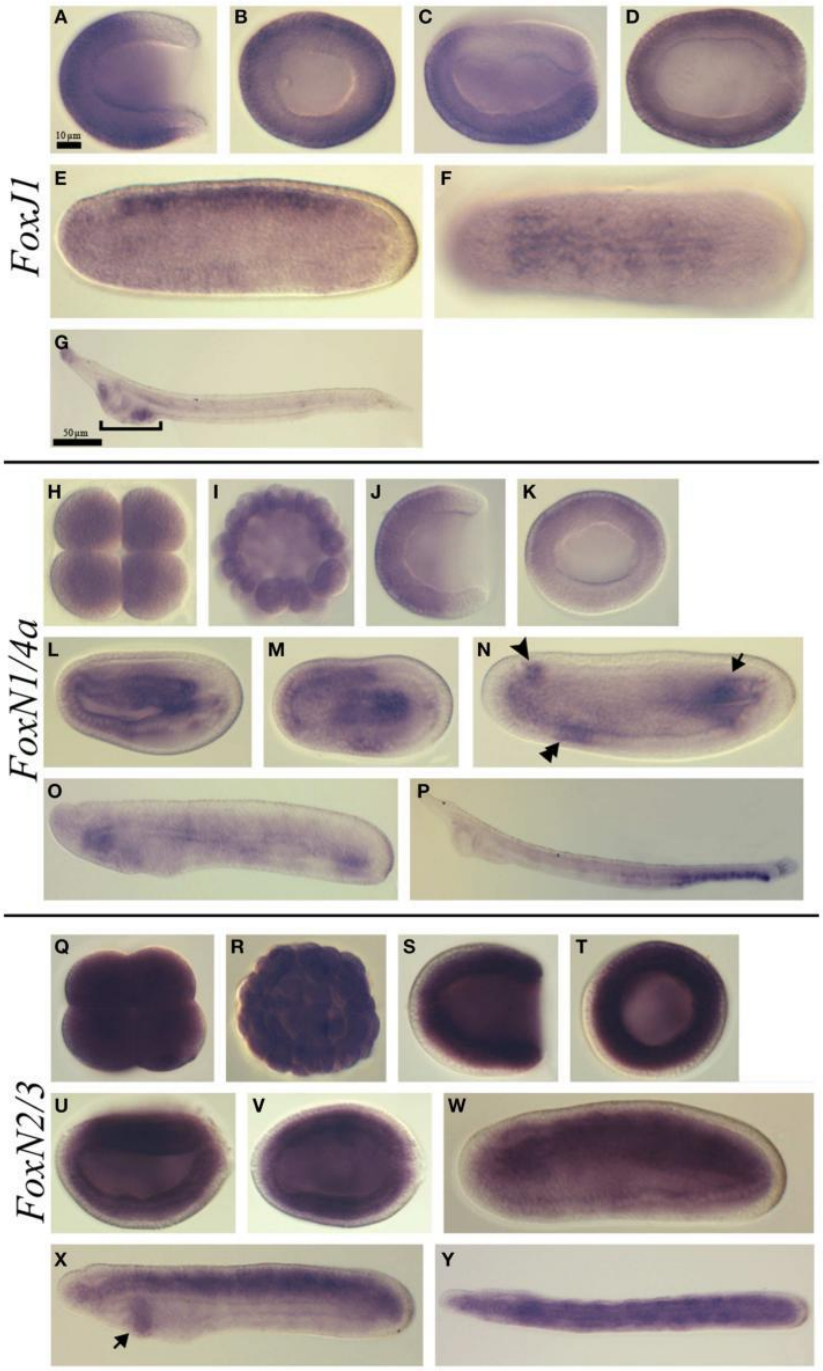


FIGURE 4 | Expression of *B. lanceolatum FoxJ1*, *FoxN1/4a*, and *FoxN2/3*. In all the panels except **(B, H, I, K, Q, R, T)** anterior is to the left. In lateral and blastoporal views dorsal is to the top. *FoxJ1* expression pattern **(A–G)**. Gastrula lateral **(A)** and blasporal **(B)** views. Late gastrula lateral **(C)** and dorsal **(D)** views. Mid-late neurula lateral **(E)** and dorsal **(F)** views. In the larva lateral view **(G)** the bracket indicates the pharyngeal region. *FoxN1/4a* expression pattern **(H–P)**. Eight-cell stage **(H)**. Blastula stage **(I)**. Gastrula lateral **(J)** and blastoporal views **(K)**. Early neurula lateral **(L)** and dorsal **(M)** views. In the mid-late neurula lateral view **(N)**, the arrowhead, double arrowhead and arrow mark the three main expression domains: at the level of the cerebral vesicle, in the anterior ventral endoderm and in the posterior

(Continued)

FIGURE 4 | Continued
mesoderm, respectively. Late neurula stage lateral view **(O)**. Larva stage lateral view **(P)**. *FoxN2/3* expression pattern **(Q–Y)**. Eight-cell stage **(Q)**. Blastula stage **(R)**. Gastrula lateral **(S)** and blasporal **(T)** views. Late gastrula lateral **(U)** and dorsal **(V)** views. Mid-late neurula lateral view **(W)**. Late neurula lateral **(X)** and dorsal **(Y)** views. The arrow in **(X)** indicates the expression domain in the pharyngeal endoderm. Scale bar: 10 μm **(A–F), (H–N), (Q–W),** and 50 μm **(G), (O–P), (X,Y)**.

mesendoderm (**Figures 4S,T**). At the late gastrula stage, the expression remained strong in the mesendoderm but started to become lower in the ventral part (**Figures 4U,V**). By the mid-late neurula stage, *FoxN2/3* transcripts were detected in the mesoderm and in the neural tube (**Figure 4W**). At the late neurula stage, before the mouth opens, the expression was mainly detected in the paraxial mesoderm (somites) and in the notochord. A new expression domain also appeared at this stage in the pharyngeal endoderm (**Figures 4X,Y**). At the larva stage, we did not detect any specific signal using *in situ* hybridization.

Discussion

Fox Genes Expression in Cephalochordate Species

The complete or partial embryonic expression patterns of *FoxAa*, *FoxAb*, *FoxB*, *FoxC*, *FoxD*, *FoxEa, FoxG,* and *FoxN1/4a* were previously described in *B. floridae* and/or *B. belcheri* (Shimeld, 1997; Terazawa and Satoh, 1997; Toresson et al., 1998; Mazet and Shimeld, 2002; Yu et al., 2002a,b; Mazet et al., 2006; Bajoghli et al., 2009). These genes overwhelmingly show a similar embryonic expression to what we observed in *B. lanceolatum*, as we have previously noticed for other important developmental genes (Somorjai et al., 2008). However, our work brings some new information.

First, in contrast to what has been described in *B. floridae*, we showed that *FoxAa* and *FoxAb* have different expression patterns. Indeed, in *B. floridae*, *FoxAb in situ* hybridization data showed that it has a similar expression to *FoxAa* at early stages whereas expression was no more detected after the eight somites stage (Shimeld, 1997). Here we showed that although both genes were expressed in the mesendodermal part of the dorsal blastoporal lip at the gastrula stage, the overall expression patterns are consistently different between the two genes and we observed a restricted expression of *FoxAb* from the gastrula to the larva stage. These discrepancies might be explained by the fact that the level of expression of *FoxAb* is very low. Indeed, staining of embryos hybridized to *FoxAb* took very long suggesting a low expression level. Thus, the staining time used in *B. floridae* might have been too short to detect expression in late stage embryos. Moreover, the expression we observed for *FoxAa* in *B. lanceolatum* is different from what was observed in *B. floridae* but similar to what has been described in *B. belcheri* (Terazawa and Satoh, 1997). Indeed, as in *B. belcheri*, *FoxAa* was not expressed in the central nervous system of *B. lanceolatum*. On the other hand, *FoxAb* showed a very specific expression in the ventral part of the neural tube in neurula stage embryos, which has been proposed to be homologous to the vertebrate floor plate. Vertebrates have three FoxA group paralogous genes that are expressed in the organizer, the notochord, the floor plate and the endoderm (Friedman and Kaestner, 2006). In *Ciona* (Di Gregorio et al., 2001), *Ci-fkh* is also expressed in the notochord, the floor plate and the endoderm. The data we obtained in *B. lanceolatum* suggest that the expression of *FoxA* in the chordate ancestor was similar to what is observed in tunicates and that independent sub-functionalizations occurred in cephalochordates after specific gene duplication and in vertebrates after the two rounds of whole genome duplications.

Concerning *FoxB*, expression in *B. floridae* was first detected in neurulae with five somites (Mazet and Shimeld, 2002). Here we showed that in *B. lanceolatum FoxB* expression could be observed in gastrula embryos in the dorsal posterior mesendoderm and ectoderm. Then, in neurulae, we detected expression in the neural plate similar to *B. floridae*, as well as an expression in the most posterior somites that was not previously described. This expression in the neural plate/neural tube and in the lastly formed somites persisted until the late neurula stage. Interestingly, in amphioxus three different somitic populations have been described (Bertrand et al., 2011). The first, most anterior, population forms under the control of the FGF signal and the two posterior populations forms independently of the FGF signal. Several genes are expressed specifically in these three somitic populations but only one gene, *Mox,*(Minguillon and Garcia-Fernandez, 2002) is expressed in the second and third populations. The present data suggest that *FoxB* also plays a role in the formation of these somitic population since it is also expressed in the two most-posterior somitic populations.

In *B. floridae, FoxC* has been described as being firstly expressed in the mesoderm of neurulae but its expression was described only in one developmental stage (Mazet et al., 2006). Here we showed that expression starts much earlier, at the gastrula stage, in the dorsal paraxial mesendoderm, the presumptive somitic mesoderm territory. Expression persisted in the paraxial mesoderm/somites until the larva stage, and at the late neurula stage we started to observe expression in the club-shaped gland anlagen and at the place where the first gill slit opens. These data suggest a major ancestral role of *FoxC* during somitogenesis which would have been conserved in vertebrates (Kume et al., 2001; Wilm et al., 2004; Wotton et al., 2008) and lost in tunicates in which *FoxC* is expressed in neural and palp cells (Imai et al., 2006).

FoxD and *FoxEa* expression in *B. lanceolatum* was very similar to previous descriptions in *B. floridae* (Yu et al., 2002a,b). However we noticed expression in some specific regions of the pharynx in late neurulae and larvae for *FoxD*, and a transient expression in mid-late and late neurula stage embryos in the posterior endoderm for *FoxEa* that were not described in the Caribbean species.

FoxG, previously known as *Brain Factor 1* (*BF-1*), was described in *B. floridae* as a gene that is ventrally expressed in the cerebral vesicle and in the anterior-most portion of the first somite pair (Toresson et al., 1998). Our results showed

a conserved expression pattern in the cerebral vesicle area in *B. lanceolatum*. However, mesoderm expression is not only limited to the first somite pair but the first three somite pairs exhibit the same pattern at the neurula stage suggesting that this gene might play a role during anterior somitogenesis. This result highlights the functional differences between the formation of the anterior somites which is under the control of the FGF signaling pathway and the formation of the most posterior somites which is not FGF-dependent (Bertrand et al., 2011). Moreover, expression is localized in the ventral part of these three most anterior somites which will give rise to the perivisceral coelom, suggesting a function of *FoxG* in the establishment of the somitic compartments.

FoxJ1 and the Formation of Motile Cilia

FoxJ1 orthologs were identified in many eumetazoans as well as in sponges (Larroux et al., 2006) and choanoflagellates (King et al., 2008). In vertebrates, *FoxJ1* plays an essential role in the generation of motile cilia and in mediating Left/Right asymmetry (Chen et al., 1998; Brody et al., 2000; Yu et al., 2008b). It has also recently been shown that misexpression of *FoxJ1* from placozoans, echinoderms and platyhelminthes in zebrafish embryos induces the expression of ciliary genes, whereas the inactivation of *FoxJ1* in the flatworm *Schmidtea mediterranea* impairs the normal differentiation of motile cilia, suggesting a conserved function in metazoans (Vij et al., 2012). This conserved function is also supported by the embryonic expression of *FoxJ1* in different phyla (Choi et al., 2006; Tu et al., 2006; Fritzenwanker et al., 2014). In *B. lanceolatum*, we showed that *FoxJ1* is first expressed in the ectoderm of the gastrulae, excluding the blastoporal region and the presumptive neural plate, at the time at which motile cilia start to grow. Then, in neurulae, expression was lost in the epidermis and appeared in the closed neural tube. At the larva stage, expression was restricted to the anterior tip of the animal and to the ciliated preoral pit and first gill slit. This expression pattern suggests that in amphioxus *FoxJ1* might also play a role in the formation of motile cilia. However, other cells, like the epithelial gut cells, also harbor motile cilia and do not express *FoxJ1*, suggesting that other genes might also be implicated in ciliogenesis in these embryonic structures.

FoxAB

In *B. lanceolatum*, *FoxAB* was transiently expressed in the organizer at the gastrula stage and in the presumptive notochord later on. No expression could be detected in mid-neurulae or larvae. FoxAB family genes were described in hemichordates (Fritzenwanker et al., 2014), sea urchin (Tu et al., 2006) and cnidarians and are absent in vertebrates and tunicates, the two other chordate clades (Yu et al., 2008a). In the hemichordate *Saccoglossus kowalevskii*, *FoxAB* is expressed in the ectoderm and the mouth perforates through the ring expressing this gene in the ventral side (Fritzenwanker et al., 2014). In bryozoans, *FoxAB* also shows an ectodermal expression (Fuchs et al., 2011). Therefore, it is still difficult to propose any scenario for the evolution of the function of FoxAB family genes in bilaterians. *FoxAB* could have been recruited for the patterning of the notochord field in the ancestor of chordates, but the absence of genes of this family in tunicates and vertebrates make this hypothesis unlikely.

FoxK and *FoxM* Ubiquitous Expression

We detected a ubiquitous expression of *FoxK* starting at the eight-cell stage until the larva stage. In other bilaterians data are scarce. In vertebrates, there are two paralogs in the FoxK family, *FoxK1* and *FoxK2*. In mouse, the study of the function of *FoxK1* during embryonic development was undertaken showing that the gene is involved in myogenic differentiation (Bassel-Duby et al., 1994). In *Ciona intestinalis* (Imai et al., 2004) as in the hemichordate *S. kowalevskii* (Fritzenwanker et al., 2014), the expression of *FoxK* is quite ubiquitous as observed for *B. lanceolatum*. Finally, studies in *Drosophila* have shown that *FoxK* is involved in the differentiation of midgut in the fly embryo (Casas-Tinto et al., 2008). Altogether these data do not allow us to infer any putative ancestral function for FoxK family genes and further studies are required in different animal phyla.

FoxM expression is also ubiquitous in *B. lanceolatum* and was first detected as early as the eight-cell stage. Then the expression level continuously decreased while development proceeds and became undetectable by *in situ* hybridization at the late neurula stage. In *Xenopus*, *FoxM1* is maternally expressed and transcripts are thereafter detected in the neuroectoderm (Pohl et al., 2005). Moreover this gene has been shown to be important for early neuronal differentiation (Ueno et al., 2008). In mouse, *FoxM1* is expressed in dividing cells and knock-out animals exhibit embryonic lethal phenotype due to many malformations affecting different organs such as the liver, the heart, the lung, or the vasculature (Kalin et al., 2011). As for FoxK, the data available up to now do not give us any indication on the putative ancestral function of genes belonging to the FoxM family.

FoxN1/4a and *FoxN2/3* Expression

In all vertebrates studied so far, *FoxN1* plays an essential role in thymus development (Ma et al., 2012; Neves et al., 2012; Lee et al., 2013; Romano et al., 2013). Moreover, in mammals, *FoxN1* is essential for hair formation whereas it is also expressed in chick during feather development (Darnell et al., 2014). Although mammal and fish FoxN1s are able to activate the expression of hair keratin genes, FoxN1/4 from amphioxus is not because its N-terminal region of the forkhead domain is different compared with vertebrates (Schlake et al., 2000). On the other hand, *FoxN4* is expressed in the nervous system, including retina, during vertebrate development (Danilova et al., 2004; Kelly et al., 2007; Boije et al., 2013). Outside vertebrates, embryonic expression has been described in *S. kowalevskii* (Fritzenwanker et al., 2014) and in a single developmetal stage of *B. floridae* (Bajoghli et al., 2009). In the hemichordate, expression of *FoxN1/4* is ubiquitous during early development and is thereafter observed in the ectoderm. In *B. lanceolatum*, the expression of *FoxN1/4a* was very dynamic with a maternal ubiquitous expression followed by restricted expression in the ectoderm at the gastrula stage, in the endoderm and axial mesoderm in neurulae, in the cerebral vesicle, the pharynx and the posterior somites later on, and, finally, in

the posterior gut of the larvae. These data suggest that *FoxN1* and *FoxN4* probably acquired new functions in vertebrates, and analysis of the expression of *FoxN1/4* family genes in tunicates will be needed to better understand this point. Interestingly, the gut of amphioxus larva and adult is considered as a major organ for immunity and *FoxN1/4a* might, as vertebrates *FoxN1*, play a role in the control of immune system function in amphioxus. However, further functional studies are required to test this hypothesis.

In vertebrates, *FoxN3* is important for craniofacial and eye development (Schuff et al., 2007; Samaan et al., 2010; Schmidt et al., 2011). In *Xenopus, FoxN3* is expressed in neural crest and eye field whereas *FoxN2* is expressed early in the eye field and then in branchial arches, retina and vagal ganglion (Schuff et al., 2006). In mouse, *FoxN2* is expressed in craniofacial, limb, nervous system and somitic tissues (Tribioli et al., 2002). In *Ciona intestinalis*, expression of *FoxN2/3* is quite ubiquitous during early development and becomes more intense in the sensory vesicle, the mesenchyme, the notochord and the palps after gastrulation (Imai et al., 2004). In sea urchin *FoxN2/3* is expressed in the non-skeletogenic mesoderm and, later on, in the endoderm and it has been shown that *FoxN2/3* function is important for ingression and for the expression of genes coding for proteins of the skeletal matrix (Rho and Mcclay, 2011). Here, we show that *FoxN2/3* in amphioxus was ubiquitously expressed at early stages. Then, at the gastrula stage, its expression was restricted to the endomesoderm and later on we observed a specific expression in the somites. Altogether, this suggests a conserved role of *FoxN2/3* in the development of mesoderm in deuterostomes, although genes of this family seem to have acquired specific functions in each chordate lineage.

Conclusions

Analyzing the expression of Fox genes in the Mediterranean amphioxus, *B. lanceolatum* showed us several points. First, as previously described for other gene families (Somorjai et al., 2008), the expression of orthologous genes in different amphioxus species shows a high degree of stasis. However, differences may be found that can easily be explained by variation in experimental sensitivity. And, second, the comparative analyzes of the expression of amphioxus Fox genes with other metazoans and particularly chordates have shown a high degree of conservation for some genes (e.g., *FoxC*, *FoxD*), but also divergent patterns in others (e.g., *FoxM*, *FoxN1/4a*). This indicates that Fox genes were necessary for essential functions in metazoans but they were also instrumental for the evolution of new functions. Further studies in amphioxus and other metazoans, and particularly functional studies, will be extremely important in the future to establish the complete picture of Fox genes expression and function and their role in the evolution of animals.

Acknowledgments

DA holds a fellowship from CONICYT "Becas Chile." Part of this study was supported by the EXOMOD grant from the CNRS.

Supplementary Material

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fevo.2015.00080

References

Bajoghli, B., Aghaallaei, N., Hess, I., Rode, I., Netuschil, N., Tay, B. H., et al. (2009). Evolution of genetic networks underlying the emergence of thymopoiesis in vertebrates. *Cell* 138, 186–197. doi: 10.1016/j.cell.2009.04.017

Bassel-Duby, R., Hernandez, M. D., Yang, Q., Rochelle, J. M., Seldin, M. F., and Williams, R. S. (1994). Myocyte nuclear factor, a novel winged-helix transcription factor under both developmental and neural regulation in striated myocytes. *Mol. Cell Biol.* 14, 4596–4605.

Benayoun, B. A., Caburet, S., and Veitia, R. A. (2011). Forkhead transcription factors: key players in health and disease. *Trends Genet.* 27, 224–232. doi: 10.1016/j.tig.2011.03.003

Bertrand, S., Camasses, A., Somorjai, I., Belgacem, M. R., Chabrol, O., Escande, M. L., et al. (2011). Amphioxus FGF signaling predicts the acquisition of vertebrate morphological traits. *Proc. Natl. Acad. Sci. U.S.A.* 108, 9160–9165. doi: 10.1073/pnas.1014235108

Bertrand, S., and Escriva, H. (2011). Evolutionary crossroads in developmental biology: amphioxus. *Development* 138, 4819–4830. doi: 10.1242/dev.066720

Bertrand, S., and Escriva, H. (2014). "Chordates: The acquisition of an axial backbone," *The Tree of Life*, eds P. Vargas and R. Zardoya (Sunderland, MA: Sinauer Associates, Inc), 460–468.

Boije, H., Shirazi Fard, S., Ring, H., and Hallbook, F. (2013). Forkheadbox N4 (FoxN4) triggers context-dependent differentiation in the developing chick retina and neural tube. *Differentiation* 85, 11–19. doi: 10.1016/j.diff.2012.12.002

Brody, S. L., Yan, X. H., Wuerffel, M. K., Song, S. K., and Shapiro, S. D. (2000). Ciliogenesis and left-right axis defects in forkhead factor HFH-4-null mice. *Am. J. Respir. Cell Mol. Biol.* 23, 45–51. doi: 10.1165/ajrcmb.23.1.4070

Carlsson, P., and Mahlapuu, M. (2002). Forkhead transcription factors: key players in development and metabolism. *Dev. Biol.* 250, 1–23. doi: 10.1006/dbio.2002.0780

Casas-Tinto, S., Gomez-Velazquez, M., Granadino, B., and Fernandez-Funez, P. (2008). FoxK mediates TGF-beta signalling during midgut differentiation in flies. *J. Cell Biol.* 183, 1049–1060. doi: 10.1083/jcb.200808149

Chen, J., Knowles, H. J., Hebert, J. L., and Hackett, B. P. (1998). Mutation of the mouse hepatocyte nuclear factor/forkhead homologue 4 gene results in an absence of cilia and random left-right asymmetry. *J. Clin. Invest.* 102, 1077–1082. doi: 10.1172/JCI4786

Choi, V. M., Harland, R. M., and Khokha, M. K. (2006). Developmental expression of FoxJ1.2, FoxJ2, and FoxQ1 in Xenopus tropicalis. *Gene Expr. Patterns* 6, 443–447. doi: 10.1016/j.modgep.2005.11.007

Clark, K. L., Halay, E. D., Lai, E., and Burley, S. K. (1993). Co-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone H5. *Nature* 364, 412–420. doi: 10.1038/364412a0

Danilova, N., Visel, A., Willett, C. E., and Steiner, L. A. (2004). Expression of the winged helix/forkhead gene, foxn4, during zebrafish development. *Brain Res. Dev. Brain Res.* 153, 115–119. doi: 10.1016/j.devbrainres.2004.05.014

Darnell, D. K., Zhang, L. S., Hannenhalli, S., and Yaklichkin, S. Y. (2014). Developmental expression of chicken FOXN1 and putative target genes during feather development. *Int. J. Dev. Biol.* 58, 57–64. doi: 10.1387/ijdb.130023sy

Di Gregorio, A., Corbo, J. C., and Levine, M. (2001). The regulation of forkhead/HNF-3beta expression in the Ciona embryo. *Dev. Biol.* 229, 31–43. doi: 10.1006/dbio.2000.9964

Friedman, J. R., and Kaestner, K. H. (2006). The Foxa family of transcription factors in development and metabolism. *Cell Mol. Life Sci.* 63, 2317–2328. doi: 10.1007/s00018-006-6095-6

Fritzenwanker, J. H., Gerhart, J., Freeman, R. M. Jr., and Lowe, C. J. (2014). The Fox/Forkhead transcription factor family of the hemichordate *Saccoglossus kowalevskii*. *Evodevo* 5:17. doi: 10.1186/2041-9139-5-17

Fuchs, J., Martindale, M. Q., and Hejnol, A. (2011). Gene expression in bryozoan larvae suggest a fundamental importance of pre-patterned blastemic cells in the bryozoan life-cycle. *Evodevo* 2:13. doi: 10.1186/2041-9139-2-13

Fuentes, M., Benito, E., Bertrand, S., Paris, M., Mignardot, A., Godoy, L., et al. (2007). Insights into spawning behavior and development of the European amphioxus (*Branchiostoma lanceolatum*). *J. Exp. Zool. B Mol. Dev. Evol.* 308, 484–493. doi: 10.1002/jez.b.21179

Fuentes, M., Schubert, M., Dalfo, D., Candiani, S., Benito, E., Gardenyes, J., et al. (2004). Preliminary observations on the spawning conditions of the European amphioxus (*Branchiostoma lanceolatum*) in captivity. *J. Exp. Zool. B Mol. Dev. Evol.* 302, 384–391. doi: 10.1002/jez.b.20025

Imai, K. S., Hino, K., Yagi, K., Satoh, N., and Satou, Y. (2004). Gene expression profiles of transcription factors and signaling molecules in the ascidian embryo: towards a comprehensive understanding of gene networks. *Development* 131, 4047–4058. doi: 10.1242/dev.01270

Imai, K. S., Levine, M., Satoh, N., and Satou, Y. (2006). Regulatory blueprint for a chordate embryo. *Science* 312, 1183–1187. doi: 10.1126/science.1123404

Kalin, T. V., Ustiyan, V., and Kalinichenko, V. V. (2011). Multiple faces of FoxM1 transcription factor: lessons from transgenic mouse models. *Cell Cycle* 10, 396–405. doi: 10.4161/cc.10.3.14709

Kelly, L. E., Nekkalapudi, S., and El-Hodiri, H. M. (2007). Expression of the forkhead transcription factor FoxN4 in progenitor cells in the developing *Xenopus laevis* retina and brain. *Gene Expr. Patterns* 7, 233–238. doi: 10.1016/j.modgep.2006.10.003

King, N., Westbrook, M. J., Young, S. L., Kuo, A., Abedin, M., Chapman, J., et al. (2008). The genome of the choanoflagellate Monosiga brevicollis and the origin of metazoans. *Nature* 451, 783–788. doi: 10.1038/nature06617

Kume, T., Jiang, H., Topczewska, J. M., and Hogan, B. L. (2001). The murine winged helix transcription factors, Foxc1 and Foxc2, are both required for cardiovascular development and somitogenesis. *Genes Dev.* 15, 2470–2482. doi: 10.1101/gad.907301

Larroux, C., Fahey, B., Liubicich, D., Hinman, V. F., Gauthier, M., Gongora, M., et al. (2006). Developmental expression of transcription factor genes in a demosponge: insights into the origin of metazoan multicellularity. *Evol. Dev.* 8, 150–173. doi: 10.1111/j.1525-142X.2006.00086.x

Larroux, C., Luke, G. N., Koopman, P., Rokhsar, D. S., Shimeld, S. M., and Degnan, B. M. (2008). Genesis and expansion of metazoan transcription factor gene classes. *Mol. Biol. Evol.* 25, 980–996. doi: 10.1093/molbev/msn047

Lee, Y. H., Williams, A., Hong, C. S., You, Y., Senoo, M., and Saint-Jeannet, J. P. (2013). Early development of the thymus in *Xenopus laevis*. *Dev. Dyn.* 242, 164–178. doi: 10.1002/dvdy.23905

Ma, D., Wang, L., Wang, S., Gao, Y., Wei, Y., and Liu, F. (2012). Foxn1 maintains thymic epithelial cells to support T-cell development via mcm2 in zebrafish. *Proc. Natl. Acad. Sci. U.S.A.* 109, 21040–21045. doi: 10.1073/pnas.1217021110

Mazet, F., Amemiya, C. T., and Shimeld, S. M. (2006). An ancient Fox gene cluster in bilaterian animals. *Curr. Biol.* 16, R314–R316. doi: 10.1016/j.cub.2006.03.088

Mazet, F., and Shimeld, S. M. (2002). The evolution of chordate neural segmentation. *Dev. Biol.* 251, 258–270. doi: 10.1006/dbio.2002.0831

Mazet, F., Yu, J. K., Liberles, D. A., Holland, L. Z., and Shimeld, S. M. (2003). Phylogenetic relationships of the Fox (Forkhead) gene family in the Bilateria. *Gene* 316, 79–89. doi: 10.1016/S0378-1119(03)00741-8

Miller, M. A., Schwartz, T., Pickett, B. E., He, S., Klem, E. B., Scheuermann, R. H., et al. (2015). A RESTful API for access to phylogenetic tools via the CIPRES science gateway. *Evol. Bioinform. Online* 11, 43–48. doi: 10.4137/EBO.S21501

Minguillon, C., and Garcia-Fernandez, J. (2002). The single amphioxus Mox gene: insights into the functional evolution of Mox genes, somites, and the asymmetry of amphioxus somitogenesis. *Dev. Biol.* 246, 455–465. doi: 10.1006/dbio.2002.0660

Neves, H., Dupin, E., Parreira, L., and Le Douarin, N. M. (2012). Modulation of Bmp4 signalling in the epithelial-mesenchymal interactions that take place in early thymus and parathyroid development in avian embryos. *Dev. Biol.* 361, 208–219. doi: 10.1016/j.ydbio.2011.10.022

Onimaru, K., Shoguchi, E., Kuratani, S., and Tanaka, M. (2011). Development and evolution of the lateral plate mesoderm: comparative analysis of amphioxus and lamprey with implications for the acquisition of paired fins. *Dev. Biol.* 359, 124–136. doi: 10.1016/j.ydbio.2011.08.003

Oulion, S., Bertrand, S., Belgacem, M. R., Le Petillon, Y., and Escriva, H. (2012). Sequencing and analysis of the Mediterranean amphioxus (*Branchiostoma lanceolatum*) transcriptome. *PLoS ONE* 7:e36554. doi: 10.1371/journal.pone.0036554

Pohl, B. S., Rossner, A., and Knochel, W. (2005). The Fox gene family in *Xenopus laevis*:FoxI2, FoxM1 and FoxP1 in early development. *Int. J. Dev. Biol.* 49, 53–58. doi: 10.1387/ijdb.051977bp

Rho, H. K., and Mcclay, D. R. (2011). The control of foxN2/3 expression in sea urchin embryos and its function in the skeletogenic gene regulatory network. *Development* 138, 937–945. doi: 10.1242/dev.058396

Romano, R., Palamaro, L., Fusco, A., Giardino, G., Gallo, V., Del Vecchio, L., et al. (2013). FOXN1: a master regulator gene of thymic epithelial development program. *Front. Immunol.* 4:187. doi: 10.3389/fimmu.2013.00187

Ronquist, F., Teslenko, M., Van Der Mark, P., Ayres, D. L., Darling, A., Hohna, S., et al. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* 61, 539–542. doi: 10.1093/sysbio/sys029

Samaan, G., Yugo, D., Rajagopalan, S., Wall, J., Donnell, R., Goldowitz, D., et al. (2010). Foxn3 is essential for craniofacial development in mice and a putative candidate involved in human congenital craniofacial defects. *Biochem. Biophys. Res. Commun.* 400, 60–65. doi: 10.1016/j.bbrc.2010.07.142

Schlake, T., Schorpp, M., Maul-Pavicic, A., Malashenko, A. M., and Boehm, T. (2000). Forkhead/winged-helix transcription factor Whn regulates hair keratin gene expression: molecular analysis of the nude skin phenotype. *Dev. Dyn.* 217, 368–376. doi: 10.1002/(SICI)1097-0177(200004)217:4<368::AID-DVDY4>3.0.CO;2-Z

Schmidt, J., Schuff, M., and Olsson, L. (2011). A role for FoxN3 in the development of cranial cartilages and muscles in *Xenopus laevis* (Amphibia: *Anura: Pipidae*) with special emphasis on the novel rostral cartilages. *J. Anat.* 218, 226–242. doi: 10.1111/j.1469-7580.2010.01315.x

Schuff, M., Rossner, A., Donow, C., and Knochel, W. (2006). Temporal and spatial expression patterns of FoxN genes in *Xenopus laevis* embryos. *Int. J. Dev. Biol.* 50, 429–434. doi: 10.1387/ijdb.052126ms

Schuff, M., Rossner, A., Wacker, S. A., Donow, C., Gessert, S., and Knochel, W. (2007). FoxN3 is required for craniofacial and eye development of *Xenopus laevis*. *Dev. Dyn.* 236, 226–239. doi: 10.1002/dvdy.21007

Shimeld, S. M., Boyle, M. J., Brunet, T., Luke, G. N., and Seaver, E. C. (2010a). Clustered Fox genes in lophotrochozoans and the evolution of the bilaterian Fox gene cluster. *Dev. Biol.* 340, 234–248. doi: 10.1016/j.ydbio.2010.01.015

Shimeld, S. M., Degnan, B., and Luke, G. N. (2010b). Evolutionary genomics of the Fox genes: origin of gene families and the ancestry of gene clusters. *Genomics* 95, 256–260. doi: 10.1016/j.ygeno.2009.08.002

Shimeld, S. M. (1997). Characterisation of amphioxus HNF-3 genes: conserved expression in the notochord and floor plate. *Dev. Biol.* 183, 74–85. doi: 10.1006/dbio.1996.8481

Somorjai, I., Bertrand, S., Camasses, A., Haguenauer, A., and Escriva, H. (2008). Evidence for stasis and not genetic piracy in developmental expression patterns of *Branchiostoma lanceolatum* and *Branchiostoma floridae*, two amphioxus species that have evolved independently over the course of 200 Myr. *Dev. Genes Evol.* 218, 703–713. doi: 10.1007/s00427-008-0256-6

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 28, 2731–2739. doi: 10.1093/molbev/msr121

Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* 30, 2725–2729. doi: 10.1093/molbev/mst197

Terazawa, K., and Satoh, N. (1997). Formation of the chordamesoderm in the amphioxus embryo: analysis with Brachyury and fork head/HNF-3 genes. *Dev. Genes Evol.* 207, 1–11. doi: 10.1007/s004270050086

Toresson, H., Martinez-Barbera, J. P., Bardsley, A., Caubit, X., and Krauss, S. (1998). Conservation of BF-1 expression in amphioxus and zebrafish suggests evolutionary ancestry of anterior cell types that contribute to the vertebrate telencephalon. *Dev. Genes Evol.* 208, 431–439. doi: 10.1007/s004270050200

Tribioli, C., Robledo, R. F., and Lufkin, T. (2002). The murine fork head gene Foxn2 is expressed in craniofacial, limb, CNS and somitic tissues during embryogenesis. *Mech. Dev.* 118, 161–163. doi: 10.1016/S0925-4773(02)00220-4

Tu, Q., Brown, C. T., Davidson, E. H., and Oliveri, P. (2006). Sea urchin Forkhead gene family: phylogeny and embryonic expression. *Dev. Biol.* 300, 49–62. doi: 10.1016/j.ydbio.2006.09.031

Tuteja, G., and Kaestner, K. H. (2007a). Forkhead transcription factors II. *Cell* 131, 192. doi: 10.1016/j.cell.2007.09.016

Tuteja, G., and Kaestner, K. H. (2007b). SnapShot: forkhead transcription factors I. *Cell* 130, 1160. doi: 10.1016/j.cell.2007.09.005

Ueno, H., Nakajo, N., Watanabe, M., Isoda, M., and Sagata, N. (2008). FoxM1-driven cell division is required for neuronal differentiation in early Xenopus embryos. *Development* 135, 2023–2030. doi: 10.1242/dev.019893

Vij, S., Rink, J. C., Ho, H. K., Babu, D., Eitel, M., Narasimhan, V., et al. (2012). Evolutionarily ancient association of the FoxJ1 transcription factor with the motile ciliogenic program. *PLoS Genet* 8:e1003019. doi: 10.1371/journal.pgen.1003019

Weigel, D., and Jackle, H. (1990). The fork head domain: a novel DNA binding motif of eukaryotic transcription factors? *Cell* 63, 455–456. doi: 10.1016/0092-8674(90)90439-L

Wilm, B., James, R. G., Schultheiss, T. M., and Hogan, B. L. (2004). The forkhead genes, Foxc1 and Foxc2, regulate paraxial versus intermediate mesoderm cell fate. *Dev. Biol.* 271, 176–189. doi: 10.1016/j.ydbio.2004.03.034

Wotton, K. R., Mazet, F., and Shimeld, S. M. (2008). Expression of FoxC, FoxF, FoxL1, and FoxQ1 genes in the dogfish *Scyliorhinus canicula* defines ancient and derived roles for Fox genes in vertebrate development. *Dev. Dyn.* 237, 1590–1603. doi: 10.1002/dvdy.21553

Wotton, K. R., and Shimeld, S. M. (2006). Comparative genomics of vertebrate Fox cluster loci. *BMC Genomics* 7:271. doi: 10.1186/1471-2164-7-271

Yu, J. K., Holland, L. Z., Jamrich, M., Blitz, I. L., and Hollan, N. D. (2002a). AmphiFoxE4, an amphioxus winged helix/forkhead gene encoding a protein closely related to vertebrate thyroid transcription factor-2: expression during pharyngeal development. *Evol. Dev.* 4, 9–15. doi: 10.1046/j.1525-142x.2002.01057.x

Yu, J. K., Holland, N. D., and Holland, L. Z. (2002b). An amphioxus winged helix/forkhead gene, AmphiFoxD: insights into vertebrate neural crest evolution. *Dev. Dyn.* 225, 289–297. doi: 10.1002/dvdy.10173

Yu, J. K., Holland, N. D., and Holland, L. Z. (2003). AmphiFoxQ2, a novel winged helix/forkhead gene, exclusively marks the anterior end of the amphioxus embryo. *Dev. Genes Evol.* 213, 102–105. doi: 10.1007/s00427-003-0302-3

Yu, J. K., Mazet, F., Chen, Y. T., Huang, S. W., Jung, K. C., and Shimeld, S. M. (2008a). The Fox genes of *Branchiostoma floridae*. *Dev. Genes Evol.* 218, 629–638. doi: 10.1007/s00427-008-0229-9

Yu, X., Ng, C. P., Habacher, H., and Roy, S. (2008b). Foxj1 transcription factors are master regulators of the motile ciliogenic program. *Nat. Genet.* 40, 1445–1453. doi: 10.1038/ng.263

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Aldea, Leon, Bertrand and Escriva. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

5.2 A single three-dimensional chromatin compartment in amphioxus indicates a stepwise evolution of vertebrate Hox bimodal regulation

nature genetics

A single three-dimensional chromatin compartment in amphioxus indicates a stepwise evolution of vertebrate Hox bimodal regulation

Rafael D Acemel[1,5], Juan J Tena[1,5], Ibai Irastorza-Azcarate[1,5], Ferdinand Marlétaz[2,5], Carlos Gómez-Marín[1], Elisa de la Calle-Mustienes[1], Stéphanie Bertrand[3], Sergio G Diaz[1], Daniel Aldea[3], Jean-Marc Aury[4], Sophie Mangenot[4], Peter W H Holland[2], Damien P Devos[1], Ignacio Maeso[1], Hector Escrivá[3] & José Luis Gómez-Skarmeta[1]

© 2016 Nature America, Inc. All rights reserved.

The HoxA and HoxD gene clusters of jawed vertebrates are organized into bipartite three-dimensional chromatin structures that separate long-range regulatory inputs coming from the anterior and posterior Hox-neighboring regions[1]. This architecture is instrumental in allowing vertebrate Hox genes to pattern disparate parts of the body, including limbs[2]. Almost nothing is known about how these three-dimensional topologies originated. Here we perform extensive 4C-seq profiling of the Hox cluster in embryos of amphioxus, an invertebrate chordate. We find that, in contrast to the architecture in vertebrates, the amphioxus Hox cluster is organized into a single chromatin interaction domain that includes long-range contacts mostly from the anterior side, bringing distant *cis*-regulatory elements into contact with Hox genes. We infer that the vertebrate Hox bipartite regulatory system is an evolutionary novelty generated by combining ancient long-range regulatory contacts from DNA in the anterior Hox neighborhood with new regulatory inputs from the posterior side.

How the three-dimensional organization of DNA in the nucleus influences regulation of gene expression is a topic of central importance in biology[3]. Despite recent progress in understanding chromatin organization, little is known about how such functional interactions evolve. Here we study the evolutionary pathway leading to the bipartite three-dimensional chromatin architecture regulating vertebrate Hox gene expression. In animals, chromatin is compartmentalized into topological associating domains (TADs)—megabase-scale chromatin regions containing DNA sequences that preferentially interact with one another[4,5]. A paradigmatic example of how TADs organize gene regulatory information is presented by the vertebrate Hox clusters, which contain genes of pivotal importance for animal development[6]. Different chromosome conformation capture techniques have shown that HoxA and HoxD genomic regions are each divided into two main adjacent TADs. These TADs compartmentalize long-range regulatory inputs coming from either side of the clusters into two major domains: enhancers distal to the 3′ flank preferentially contact 'anterior' Hox genes, whereas those beyond the 5′ flank mostly interact with 'posterior' genes (**Fig. 1a**; refs. 2,7–10). This bipartite regulatory topology provides gnathostomes with a versatile bimodal system, allowing Hox genes to pattern multiple structures, including an ancestral role in anteroposterior axis patterning and novel roles in morphological innovations such as paired limbs[1].

To investigate whether the TADs associated with HoxA and HoxD clusters arose independently or have a shared ancestry dating to before the two vertebrate-specific whole-genome duplications (2R WGDs; **Supplementary Fig. 1** and ref. 11), we first studied synteny conservation around Hox clusters between and within species. In mouse, the HoxA- and HoxD-neighboring regions are strikingly different, with many HoxA long-range *cis*-regulatory elements embedded in the introns of neighboring genes, whereas HoxD long-range *cis*-regulatory elements are located in gene deserts (large intergenic regions devoid of coding genes). Data from divergent vertebrates, including elephant shark, indicate the architecture in mouse represents a derived situation and that all vertebrate Hox cluster neighborhoods were originally very similar. What is now a HoxD gene desert in mammals contained copies of HoxA-neighboring genes[12], and the gene-free regions surrounding the other two Hox clusters have also resulted from differential loss of the coding exons of neighboring genes[13] (**Fig. 1b**, **Supplementary Fig. 2** and **Supplementary Note**). Thus, we conclude that differences in the genomic organization of mammalian HoxA and HoxD regulation are derived, not ancestral. This implies that the *cis*-regulatory elements currently engaged in Hox long-range bipartite contacts were primarily intronic and intergenic

[1]Centro Andaluz de Biología del Desarrollo (CABD), Consejo Superior de Investigaciones Científicas/Universidad Pablo de Olavide, Seville, Spain. [2]Department of Zoology, University of Oxford, Oxford, UK. [3]Université Pierre et Marie Curie Université Paris 6, CNRS, UMR 7232, Biologie Integrative des Organismes Marins (BIOM), Observatoire Océanologique de Banyuls-sur-Mer, Banyuls-sur-Mer, France. [4]Commissariat à l'Energie Atomique (CEA), Institut de Génomique (IG), Genoscope, Evry, France. [5]These authors contributed equally to this work. Correspondence should be addressed to J.L.G.-S. (jlgomska@upo.es), H.E. (hescriva@obs-banyuls.fr), I.M. (nacho.maeso@gmail.com) or D.P.D. (damienpdevos@gmail.com).

Received 30 October 2015; accepted 30 December 2015; published online 1 February 2016; doi:10.1038/ng.3497

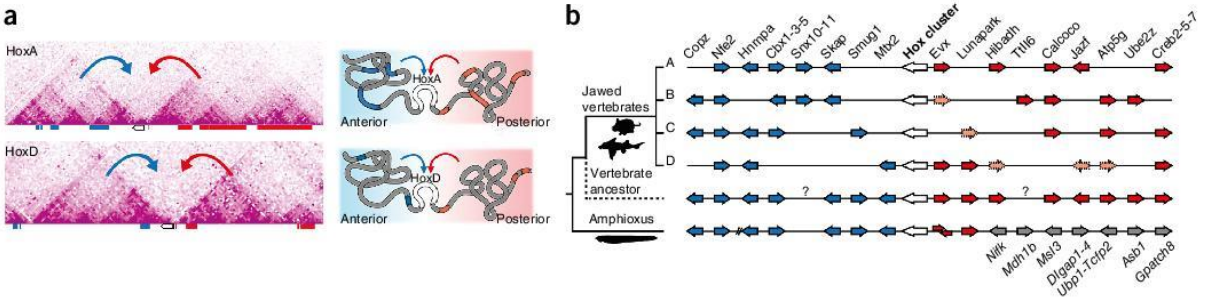


Figure 1 Genomic organization of vertebrate and amphioxus Hox clusters. (**a**) Distribution of TADs (obtained from human Hi-C data sets[20]) and schematics of the chromatin architecture of HoxA and HoxD clusters, showing their similar three-dimensional topologies. Colored bars represent Hox genes (white) and anterior (blue) and posterior (red) neighboring genes. Pink color intensity in the Hi-C plots corresponds to the number of interacting counts between bin pairs. (**b**) Microsynteny arrangements around the Hox clusters of gnathostomes, the pre-WGD vertebrate ancestor and amphioxus. Genes are represented by arrows showing transcriptional orientation (white, Hox clusters; blue, anterior genes; red, posterior genes; gray, genes with non-conserved linkages); those outlined by dashed lines correspond to vertebrate paralogs that have been lost in at least one species. Question marks indicate genes whose status in the vertebrate ancestor could not be inferred. Slashes correspond to the non-conserved amphioxus loci shown in **Supplementary Figure 3**.

within a conserved array of neighboring protein-coding loci before Hox cluster duplications (**Fig. 1b**).

We investigated the ancestry of this arrangement by examining the location of vertebrate Hox-neighboring genes in invertebrate genomes. We find that few of the invertebrate homologs are closely linked to Hox clusters outside chordates and that gene order and orientation are highly variable (for example, vertebrate anterior-linked genes are frequently found on the posterior side in invertebrates and vice versa; **Supplementary Fig. 3**). This shuffling of the Hox syntenic environment suggests that, in the bilaterian ancestor, long-range Hox *cis*-regulatory interactions were either absent or not important enough to constrain microsynteny. In contrast, in amphioxus (a non-vertebrate chordate that retains many ancestral genomic and morphological features; see refs. 14–16), synteny on the anterior side of the Hox cluster is strikingly conserved with vertebrates; gene order and orientation are almost identical to those inferred for the vertebrate ancestor (**Fig. 1b**). On the posterior side, most neighboring genes are different from those in vertebrates: only two immediately adjacent genes, *Evx* and *Lnp*, are conserved in position. The conservation of anterior flanking genes between vertebrates and amphioxus suggests that long-range regulatory interactions from the 3′ side of the cluster had become essential for Hox regulation at the base of the chordate lineage, imposing strong constraints on genomic rearrangements in this region. With regard to the posterior side, given the lack of synteny conservation in non-chordates, at present we cannot discern whether

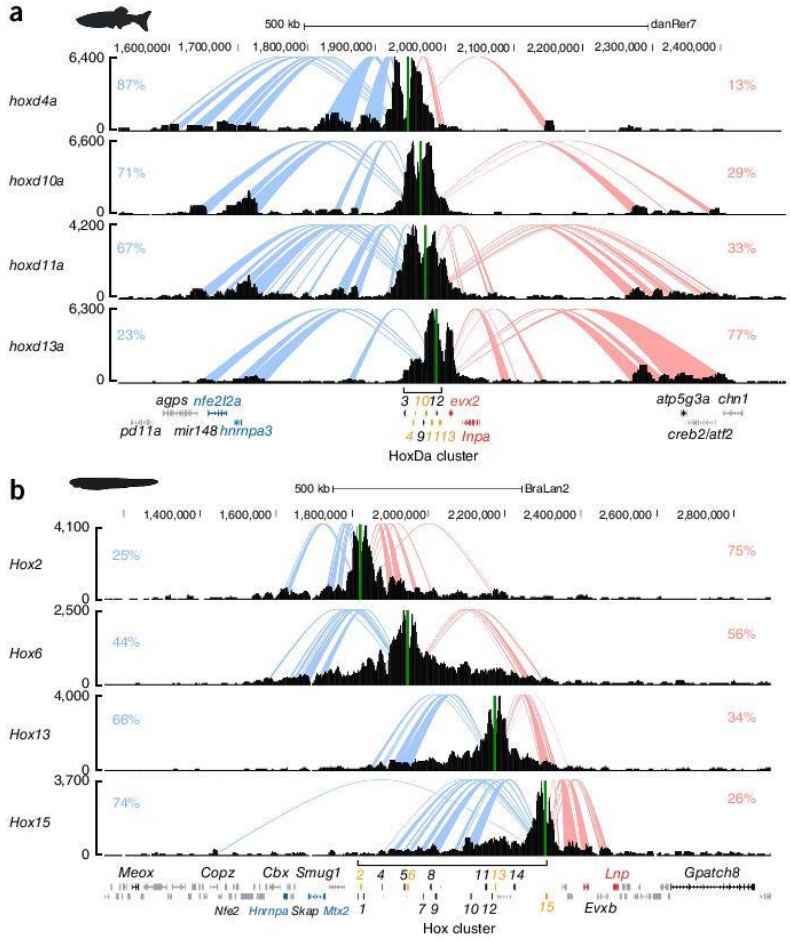


Figure 2 4C-seq interaction profiles of zebrafish and amphioxus Hox clusters. (**a**,**b**) Normalized 4C-seq profiles of several Hox gene promoters in the zebrafish HoxDa locus (**a**) and the amphioxus Hox locus (**b**). Spider plots show statistically significant contacts to the left (blue arcs) and right (red arcs) of each viewpoint. Percentages of reads aligned to statistically significant targets on each side of the viewpoints are indicated in blue (left contacts) and red (right contacts). Units on the *y* axes correspond to normalized interacting counts. Green bars indicate the positions of the viewpoints.

© 2016 Nature America, Inc. All rights reserved.

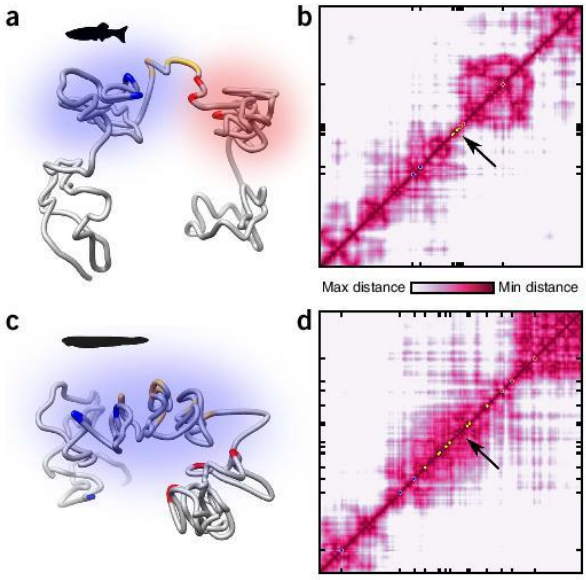


Figure 3 Three-dimensional chromatin architecture of amphioxus and zebrafish Hox clusters. (**a**,**c**) Three-dimensional models of the zebrafish HoxDa (**a**) and amphioxus Hox (**c**) regions. 4C-seq viewpoints are highlighted (blue, anterior genes; yellow, Hox genes; red, posterior genes). (**b**,**d**) Zebrafish and amphioxus virtual Hi-C consensus for all three-dimensional model solutions. 4C-seq viewpoints are represented by circles, using the same color scheme as in **a** and **c**. Arrows indicate the TAD border bisecting the zebrafish Hox cluster (**b**) and the absence of this border in the case of amphioxus (**d**).

amphioxus or vertebrates have diverged the most from the syntenic organization of the chordate ancestor. Whatever the case, beyond *Evx* and *Lnp*, gene synteny has followed different evolutionary routes in these two chordate groups, suggesting that, in stark contrast to the scenario in anterior territories, the regulatory contribution of distant posterior regions was less important or even absent in the last common chordate ancestor.

To evaluate this hypothesis experimentally, we compared Hox chromatin contacts between amphioxus and vertebrate embryos using circular chromosome conformation capture followed by high-throughput sequencing (4C-seq), a method that identifies distal chromatin contacts. Studies in mouse embryonic tissues and whole zebrafish embryos have demonstrated that 4C-seq efficiently resolves the organization of HoxA and HoxD long-range contacts into two adjacent TADs[2,7,8,10,17]. We generated 4C-seq data for 14 gene 'viewpoints' (eight Hox genes and six neighboring genes) in amphioxus embryos and compared these results with previously reported[8] and newly generated zebrafish data (four Hox genes and five neighboring genes). In total, 73 4C-seq data sets were generated, including replicates for all viewpoints and three amphioxus developmental stages (Online Methods).

With these data sets, we first defined target interacting regions for each of the 4C-seq viewpoints (genomic regions showing a statistically significant (P values $<1 \times 10^{-5}$) read enrichment against a randomized background) and quantified the number of reads corresponding to each of these targets (Online Methods). These analyses highlighted the characteristic bipartite distribution of anterior and posterior Hox long-range contacts previously reported in mouse and zebrafish[2,8,10,18] (**Fig. 2a** and **Supplementary Fig. 4**). The zebrafish *hoxd4a* and *hoxd13a* genes showed little contact overlap, with the majority of their interactions mapping to opposing sides of the cluster (83.3% anterior and 76.6% posterior, respectively). In contrast, in amphioxus, Hox genes located at the edges of the cluster showed the opposite trend: most *Hox2* and *Hox15* contacts converged in the same direction, with their interacting regions located primarily within the Hox complex (75.2% and 74.2%, respectively). In fact, regardless of their positions within the cluster, anterior, central and posterior Hox genes exhibited 4C-seq profiles that overlapped extensively, with no signs of a bipartite distribution (**Fig. 2b** and **Supplementary Fig. 5**). Notably, these Hox interaction profiles were developmentally stable, even though the number of active Hox genes in amphioxus changes dramatically from early gastrula to premouth embryo[19] (**Supplementary Fig. 6**). This temporal stability is in line with previous findings in mouse and *Drosophila melanogaster*, where most long-range three-dimensional chromatin interactions are organized similarly across tissues and developmental stages, with only some differences in the intensity of the contacts upon activation of different sets of distal enhancers[7,20,21]. However, despite this temporal uniformity, it is conceivable that the amphioxus TAD structures could be less similar across cell populations with different transcriptional activities than they are in vertebrates; thus, by using whole embryos, we may be missing cell type–specific chromatin interactions.

We then correlated 4C-seq results with synteny data. Consistent with the high level of conservation of anterior neighboring genes, in the majority of amphioxus Hox viewpoints, a significant fraction of contacts mapped to the conserved anterior region (ranging from 14 to 24.8% for the promoters of *Hox2*, *Hox5*, *Hox6*, *Hox7* and *Hox9*; **Supplementary Table 1**). Long-range interactions between Hox genes and anterior territories were even clearer when using 3′ neighboring genes as viewpoints (**Supplementary Fig. 5** and **Supplementary Table 1**). The amphioxus Hox cluster contained 25.5% of *Hnrnpa* interactions, a fraction in a similar range to that of its ortholog in zebrafish (33.4%), and, in the case of amphioxus *Mtx2*, the percentage of contacts corresponding to the Hox complex reached 42.7%. In contrast, on the posterior side, we found striking differences between amphioxus and vertebrates. Hox genes contacted posterior neighboring regions in both chordate lineages; however, the distribution of these 5′ interactions was very different (**Fig. 2a**,**b**). In zebrafish, *hoxd13a* interactions entered into far distant 5′ territories, well beyond the *evx2-lnpa* syntenic region, reaching vertebrate-specific posterior neighboring genes such as *atp5g3a* and *creb2*, consistent with previous reports on the location of zebrafish and mouse 5′ long-range Hox enhancers[7,22,23]. In amphioxus, by contrast, the target interacting regions of the most posterior Hox gene, *Hox15*, were circumscribed to the most proximal neighboring region, with no significant contacts crossing the *Lnp* promoter into the amphioxus-specific territory. Thus, even within the only 5′ region with synteny conservation, interaction profiles are different. In both cases, the Evx-Lnp region contacted Hox genes, but, whereas in amphioxus *Evxa* and *Lnp* showed a clear interaction preference for the Hox cluster (66.1% and 73% of contacts, respectively), zebrafish *evx2* and *lnpa* preferentially contacted vertebrate-specific genomic regions (with only 26.8% and 20.7% of the contacts interacting with the Hox cluster, respectively) (**Supplementary Figs. 4** and **5**, and **Supplementary Table 1**). Taken together, these results suggest that there is an inflexion point for long-range chromatin interactions around the *Evxa-Evxb-Lnp* region in amphioxus, with no significant Hox contacts with 5′ amphioxus-specific genes.

To better characterize vertebrate and amphioxus Hox topologies and identify interaction compartments, we generated virtual

© 2016 Nature America, Inc. All rights reserved.

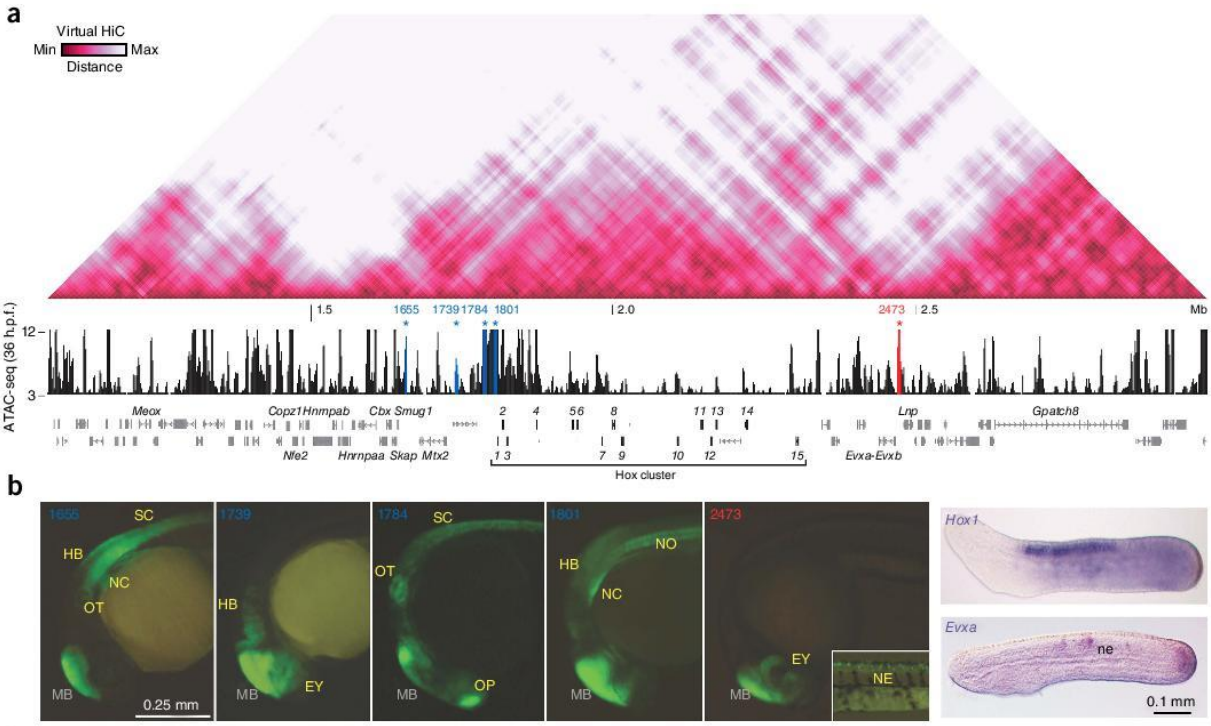


Figure 4 Regulatory compartments in the amphioxus Hox region and evolution of Hox cluster three-dimensional architecture. (**a**) Amphioxus virtual Hi-C heat map and ATAC-seq profile at 36 h.p.f. in the Hox region. Amphioxus ATAC-seq peaks tested in zebrafish are colored and highlighted by asterisks (blue for those in the anterior region, red for those on the posterior side of the cluster). (**b**) Lateral views of embryos from zebrafish transgenic lines at 24 h.p.f. and 48 h.p.f. (inset in 2473) showing GFP expression driven by the amphioxus ATAC-seq peaks (1655, 1739, 1784, 1801 and 2473) highlighted in **a**. Midbrain expression corresponds to the enhancer positive control included in the reporter constructs. Whole-mount *in situ* hybridizations for *Hox1* and *Evxa* in amphioxus embryos at 36 h.p.f. are shown for comparison. Anterior is to the left. EY, eye; HB, hindbrain; MB, midbrain; NC, neural crest cells; NE, neurons; NO, notochord; OT, otic vesicle, OP, olfactory placode, SC, spinal chord. (**c**) Three-dimensional architecture schematics showing an evolutionary scenario for the origin of the bimodal regulatory system of jawed vertebrates. The Hox-only chromatin domain of early bilaterians is first expanded on the anterior side in the chordate ancestor and then on the posterior side at the origin of vertebrates, allowing bipartition of the regulatory topologies of the HoxA and HoxD clusters. LCA, last common ancestor.

three-dimensional chromatin architecture models using the read counts of the 4C-seq signals as a proxy for the distance from each viewpoint (Online Methods and **Supplementary Fig. 7**). As the 4C-seq data correspond to pooled cells from whole embryos, our three-dimensional models provide an average view of chromatin topologies rather than a picture of the dynamic chromatin folding present in each individual cell. These integrative visualizations emphasized how strikingly different the vertebrate and amphioxus three-dimensional chromatin architectures are (**Fig. 3**). In zebrafish, the HoxDa cluster sits between the two separate anterior and posterior chromatin domains, like a hinge on which the two sets of long-range regulatory inputs can swing. In contrast, the amphioxus Hox cluster appears as a large single chromatin domain that contains distant anterior neighboring genes but not posterior ones. To visualize boundaries between chromatin domains, we developed a new approach to transform our three-dimensional modeling data into a heat map of distances (analogous to those obtained by Hi-C, hereafter termed virtual Hi-C; see the Online Methods and **Supplementary Figs. 8–10** for details

© 2016 Nature America, Inc. All rights reserved.

on virtual Hi-C validations). As expected, zebrafish virtual Hi-C recovered the bipartite architecture that divides vertebrate HoxD clusters into anterior and posterior TADs (**Fig. 3b**). In contrast, the amphioxus cluster was contained within a single TAD that included the conserved anterior neighboring genes but not the amphioxus-specific posterior genes (such as *Gpatch8*, which has its own interacting compartment) (**Fig. 3d**). Notably, no boundaries bisected the cluster or separated Hox genes from anterior neighboring territories. In the case of *Lnp* and the amphioxus Evx genes, the situation was less clear: although these loci seemed to be part of their own small interaction domain, this region was not completely isolated from its two adjacent compartments (the one containing the Hox cluster and the one including *Gpatch8*). This suggests that the single Hox three-dimensional chromatin domain present in amphioxus has a weaker contact border on its posterior side than in its anterior region and that the *EvxA-EvxB-Lnp* territory can be considered to be an extended boundary region (**Fig. 3d**).

To examine the functional relevance of amphioxus Hox chromatin organization, we searched for putative enhancers active in amphioxus embryos at 36 hours post-fertilization (h.p.f.) (immediately preceding what can be regarded as a pharyngula stage in amphioxus, equivalent to the zebrafish phylotypic stage at 24 h.p.f.) using ATAC-seq (assay for transposase-accessible chromatin using sequencing[24]; **Fig. 4a**). In agreement with the three-dimensional chromatin topologies inferred from the virtual Hi-C results, the distribution of ATAC-seq peaks on either side of the amphioxus Hox gene cluster suggested very different regulatory potentials for the two Hox-neighboring regions (**Supplementary Fig. 11**). Whereas anterior territories were rich in putative distal enhancer regions, the posterior side contained comparatively fewer ATAC-seq peaks. In fact, apart from the peaks tightly associated with the Evx genes or directly overlapping transcriptional start sites and repetitive elements, we only found a single candidate enhancer region, within the intergenic region between *Evxb* and *Lnp*. We then tested four putative enhancer elements from the anterior side of the TAD containing the amphioxus Hox cluster located at different distances from the closest Hox gene (elements 1655, 1739, 1784 and 1801, located 150 kb, 66 kb, 20 kb and 3 kb downstream of *Hox1*, respectively) and the element identified at the posterior side (element 2473, 165 kb upstream of *Hox15*) by generating zebrafish with stable GFP reporter transgenes. All the anterior enhancers promoted expression along the anteroposterior axis, consistent with the expression patterns of amphioxus Hox genes but not with those of neighboring loci (**Fig. 4b** and **Supplementary Fig. 12**; see also ref. 19), suggesting that these regions are amphioxus Hox *cis*-regulatory elements. In contrast, the 2473 posterior element activated GFP expression in isolated neurons in the spinal cord, in a pattern reminiscent of the amphioxus *Evxa* gene (**Fig. 4b**; ref. 25) rather than a Hox gene. These experiments suggest that the three-dimensional organization identified, using 4C-seq and modeling, brings long-range regulatory elements into proximity with amphioxus Hox genes mostly on the anterior side of the cluster (**Fig. 4**).

In summary, our results support a stepwise evolution of the bimodal regulatory machineries of the HoxA and HoxD clusters of jawed vertebrates (**Fig. 4c**). The relatively simple Hox cluster three-dimensional topology of early bilaterian animals, where external, long-range regulation was probably absent, changed profoundly in early chordate evolution, with newly incorporated distal regulatory inputs from anterior neighboring loci becoming a fundamental part of the Hox regulatory architecture. This unipolar topology was further developed in the vertebrate lineage. The acquisition of interactions with distal *cis*-regulatory elements on the posterior side introduced the possibility of a switch between two separate sets of long-range regulatory inputs, allowing an unprecedented plasticity in the developmental usage of the Hox patterning system in vertebrates.

URLs. Sickle (v1.290), https://github.com/najoshi/sickle; RepeatModeler, http://www.repeatmasker.org/RepeatModeler.html; UCSC Genome Browser, http://genome.ucsc.edu/; Ensembl Metazoa, http://metazoa.ensembl.org/.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. Data sets presented in this study are available under Gene Expression Omnibus (GEO) accession GSE68737.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We specially thank J. Pascual-Anaya for helping with some figures and helpful discussions. We would also like to thank F. Casares, I. Almudí and J.R. Martínez-Morales for fruitful discussions. Work was funded by grants from the Ministerio de Economía y Competitividad (BFU2013-41322-P to J.L.G.-S.; Juan de la Cierva postdoctoral contract to I.M.; BFU2014-58449-JIN to J.J.T.); the Andalusian government (BIO-396 to J.L.G.-S.; C2A (EE: 2013/2506) to D.P.D. and I.I.-A.); the European Research Council (ERC; grant 268513) to P.W.H.H. and F.M.; a European Molecular Biology Organization (EMBO) short fellowship to I.M.; the Universidad Pablo de Olavide to J.J.T.; and Conicyt 'Becas Chile' to D.A.

AUTHOR CONTRIBUTIONS

R.D.A. carried out the 4C-seq experiments with the help of C.G.-M. and S.G.D. J.J.T. performed the bioinformatic analysis of all the 4C-seq and ATAC-seq data sets. I.I.-A. developed and applied the three-dimensional modeling and virtual Hi-C procedures. F.M. generated the assembly and annotation of the Hox locus in the European amphioxus. J.-M.A. and S.M. ensured sequencing project management at Genoscope. E.d.l.C.-M., R.D.A., J.J.T. and I.M. carried out the zebrafish reporter assays. S.B. and D.A. collected and processed the amphioxus embryonic material and performed *in situ* hybridizations. I.M. completed the amphioxus ATAC-seq experiments and the evolution of synteny analyses. J.L.G.-S., H.E., I.M. and D.P.D. conceived, designed and coordinated the project. J.L.G.-S. and I.M. wrote the manuscript with the help of P.W.H.H. All authors revised and contributed to the final version of the text.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Lonfat, N. & Duboule, D. Structure, function and evolution of topologically associating domains (TADs) at *HOX* loci. *FEBS Lett.* **589**, 2869–2876 (2015).
2. Andrey, G. *et al.* A switch between topological domains underlies *HoxD* genes collinearity in mouse limbs. *Science* **340**, 1234167 (2013).
3. de Laat, W. & Duboule, D. Topology of mammalian developmental enhancers and their regulatory landscapes. *Nature* **502**, 499–506 (2013).
4. Gómez-Díaz, E. & Corces, V.G. Architectural proteins: regulators of 3D genome organization in cell fate. *Trends Cell Biol.* **24**, 703–711 (2014).
5. Ciabrelli, F. & Cavalli, G. Chromatin-driven behavior of topologically associating domains. *J. Mol. Biol.* **427**, 608–625 (2015).
6. Mallo, M. & Alonso, C.R. The regulation of *Hox* gene expression during animal development. *Development* **140**, 3951–3963 (2013).
7. Montavon, T. *et al.* A regulatory archipelago controls *Hox* genes transcription in digits. *Cell* **147**, 1132–1145 (2011).
8. Woltering, J.M., Noordermeer, D., Leleu, M. & Duboule, D. Conservation and divergence of regulatory strategies at *Hox* loci and the origin of tetrapod digits. *PLoS Biol.* **12**, e1001773 (2014).
9. Berlivet, S. *et al.* Clustering of tissue-specific sub-TADs accompanies the regulation of *HoxA* genes in developing limbs. *PLoS Genet.* **9**, e1004018 (2013).
10. Lonfat, N., Montavon, T., Darbellay, F., Gitto, S. & Duboule, D. Convergent evolution of complex regulatory landscapes and pleiotropy at *Hox* loci. *Science* **346**, 1004–1006 (2014).

© 2016 Nature America, Inc. All rights reserved.

11. Dehal, P. & Boore, J.L. Two rounds of whole genome duplication in the ancestral vertebrate. *PLoS Biol.* **3**, e314 (2005).
12. Lehoczky, J.A., Williams, M.E. & Innis, J.W. Conserved expression domains for genes upstream and within the *HoxA* and *HoxD* clusters suggests a long-range enhancer existed before cluster duplication. *Evol. Dev.* **6**, 423–430 (2004).
13. Maeso, I. *et al.* An ancient genomic regulatory block conserved across bilaterians and its dismantling in tetrapods by retrogene replacement. *Genome Res.* **22**, 642–655 (2012).
14. Paps, J., Holland, P.W. & Shimeld, S.M. A genome-wide view of transcription factor gene diversity in chordate evolution: less gene loss in amphioxus? *Brief. Funct. Genomics* **11**, 177–186 (2012).
15. Bertrand, S. & Escriva, H. Evolutionary crossroads in developmental biology: amphioxus. *Development* **138**, 4819–4830 (2011).
16. Holland, L.Z. & Onai, T. Early development of cephalochordates (amphioxus). *Wiley Interdiscip. Rev. Dev. Biol.* **1**, 167–183 (2012).
17. Noordermeer, D. *et al.* Temporal dynamics and developmental memory of 3D chromatin architecture at *Hox* gene loci. *eLife* **3**, e02557 (2014).
18. Noordermeer, D. *et al.* The dynamic architecture of *Hox* gene clusters. *Science* **334**, 222–225 (2011).
19. Pascual-Anaya, J. *et al.* Broken colinearity of the amphioxus *Hox* cluster. *Evodevo* **3**, 28 (2012).
20. Dixon, J.R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 2012).
21. Ghavi-Helm, Y. *et al.* Enhancer loops appear stable during development and are associated with paused polymerase. *Nature* **512**, 96–100 (2014).
22. Gonzalez, F., Duboule, D. & Spitz, F. Transgenic analysis of *Hoxd* gene regulation during digit development. *Dev. Biol.* **306**, 847–859 (2007).
23. Gehrke, A.R. *et al.* Deep conservation of wrist and digit enhancers in fish. *Proc. Natl. Acad. Sci. USA* **112**, 803–808 (2015).
24. Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y. & Greenleaf, W.J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).
25. Ferrier, D.E., Minguillón, C., Cebrián, C. & Garcia-Fernàndez, J. Amphioxus *Evx* genes: implications for the evolution of the midbrain-hindbrain boundary and the chordate tailbud. *Dev. Biol.* **237**, 270–281 (2001).

© 2016 Nature America, Inc. All rights reserved.

ONLINE METHODS

Genome sequencing and assembly. DNA was prepared from a single European amphioxus (*Branchiostoma lanceolatum*) mature male and sequenced using Illumina technology at Genoscope (Centre National de Séquençage, Evry, France). Briefly, two paired-end (180-bp and 700-bp) and six mate-pair (3-, 5- and 8-kb) libraries were generated and sequenced at >200× total coverage.

Reads were quality-trimmed using sickle (v1.290), and errors were corrected using Musket (v1.0.6)[26]; overlapping libraries were merged using Flash[27]. Assembly was carried out using SOAPdenovo (v2.04)[28] with a *k*-mer of 71 for contig generation and of 35 for mapping and scaffolding. Gaps were subsequently filled using GapCloser (v1.2)[28] with an overlap parameter set to 31. The resulting assembly (N50, 649 kb; size, 948.5 Mb) contains allelic copies for most scaffolds (expected genome size of ~500 Mb) that we reconciled using the Haplomerger[29] pipeline, relying on best reciprocal LASTZ alignment after masking repeats using a custom library built with RepeatModeler. The Hox locus was extracted from the final assembly (N50, 1.132 Mb; size, 526.8 Mb) and submitted together with the 4C-seq and ATAC-seq data (GSE68737).

Gene models were built using Evidence Modeler (EVM)[30] on the basis of (i) *de novo* gene prediction obtained using Augustus[31] with custom training based on CEGMA[32] report and (ii) split-aware alignment of human proteins using Exonerate[33] and transcriptome alignment. Models for known genes in the Hox region that were not present in these annotations were added manually. More details regarding *B. lanceolatum* genome assembly and annotation will be provided in a separate publication.

Synteny analyses and genome browsing. Hox-neighboring genes were searched across the different studied species using TBLASTN and BLASTP. We compared the relative orientations and positions of these genes by browsing the genomes of the studied species through the NCBI, UCSC Genome Browser and Ensembl Metazoa webpages, using the following genome versions: elephant shark (*Callorhinchus milii*) 6.1.3, *Lottia gigantea* v1.0, *Mus musculus* Build 38, *Saccoglossus kowalevskii* Build1.1, *Strigamia maritima* Smar1.0 and *Trichoplax adhaerens* v1.0. In the case of the starfish *Acanthaster planci*, no gene annotation or genome browser was available for the published *A. planci* Hox genome scaffold (DF933567.1)[34]. Therefore, we used TBLASTN to search for conserved neighboring genes and Genscan to predict genes *de novo*.

Mouse *Jazf2* pseudogenized exons were detected with VISTA[35] using elephant shark as a reference sequence, LAGAN as the alignment program and the following parameters: 100-bp window and 65% identity in 70 bp.

Amphioxus procurement and culture. *B. lanceolatum* mature adults were collected at the Racou beach in Argelès-sur-Mer (France). Gametes were collected by heat stimulation as previously described[36,37]. Fertilization was undertaken in Petri dishes filled with filtered seawater, and embryos were cultured at 19 °C.

Whole-mount *in situ* hybridization. Partial cDNA for *Gpatch8*, *Nfe2*, *Lnp*, *Slc20*, *Mtx2*, *Hnrnpa* and *Cbx* from *B. lanceolatum* was amplified by RT-PCR and cloned into the pGEM-T Easy vector. DIG-labeled RNA probes were synthesized by *in vitro* translation after plasmid linearization using the appropriate enzymes. Fixation and whole-mount *in situ* hybridization were performed as described in ref. 38. No expression could be detected using whole-mount *in situ* hybridization for *Gpatch8*, *Lnp*, *Slc20* or *Mtx2*; the expression patterns for the rest of the genes are included in **Supplementary Figure 12**.

4C-seq. 4C-seq assays were performed as previously reported[18,39–41]. For each zebrafish biological replicate, 500 embryos at 24 h.p.f. of the Tübingen strain were dechorionated using pronase and deyolked in 1 ml of Ginzburg Fish Ringers (55 mM NaCl, 1.8 mM KCl and 1.25 mM $NaHCO_3$). They were then fixed in 2% formaldehyde in 1× PBS for 15 min at room temperature. For amphioxus biological replicates, embryos (~8,000 at 8 h.p.f. and ~4,000 at 15 and 36 h.p.f.) were concentrated by centrifugation at low speed in 2-ml microtubes. They were fixed for 15 min at room temperature in 1.5 ml of MOPS buffer (0.1 M MOPS pH 7.5, 2 mM $MgSO_2$, 1 mM EGTA and 0.5 M NaCl) containing 1.85% formaldehyde. 155 µl of 10% glycine was added to both species samples to stop fixation, followed by five washes with PBS (NaPBS in the case of amphioxus) at 4 °C. Pellets were frozen in liquid nitrogen and kept at −80 °C. Isolated cells were lysed (lysis buffer: 10 mM Tris-HCl pH 8, 10 mM NaCl, 0.3% Igepal CA-630 (Sigma-Aldrich, I8896) and 1× protease inhibitor cocktail (Complete, Roche, 11697498001)), and the DNA was digested with DpnII (New England BioLabs, R0543M) and Csp6I (Fermentas, Thermo Scientific, FD0214) as primary and secondary enzymes, respectively. T4 DNA ligase (Promega, M1804) was used for both ligation steps. Specific primers were designed around the putative transcriptional start sites of the genes with Primer3 v. 0.4.0 (ref. 42). Illumina adaptors were included in the primer sequences, and eight PCRs were performed with the Expand Long Template PCR System (Roche, 11759060001) and pooled. Two libraries from different biological replicates were generated for each 4C-seq experiment (for each viewpoint and for each developmental stage). These libraries were purified with a High Pure PCR Product Purification kit (Roche, 11732668001), their concentrations were measured using the Quanti-iT PicoGreen dsDNA Assay kit (Invitrogen, P11496) and they were sent for deep sequencing. 4C-seq data were analyzed as previously described[17]. Briefly, raw sequencing data were demultiplexed and aligned using the zebrafish July 2010 assembly (danRer7) and the *B. lanceolatum* reference genomes. Reads located in fragments flanked by two restriction sites of the same enzyme, in fragments smaller than 40 bp or within a window of 10 kb around the viewpoint (indicated by dashed lines in the different figures) were filtered out. Mapped reads were then converted to reads per first enzyme fragment ends and smoothened using a 30-fragment mean running window algorithm. 4C-seq data were normalized by the total weight of reads within the window displayed in the figures.

To calculate statistically significant contacting regions for each viewpoint, an average background level was estimated as previously described[43]. Briefly, fragment distribution in a window of 2 Mb around each viewpoint was randomized, excluding an internal window of 100 kb around the viewpoint to avoid biases due to close contacts. Then, this randomized fragment distribution was smoothened as described above. This randomized profile was then used to calculate the *P* value for each potential target in the observed 4C-seq distribution by means of Poisson probability function. Regions with *P* values below 1×10^{-5} were considered as statistically significant interacting targets.

To calculate the distribution of contacts at each side of the viewpoints, we took into account only those reads overlapping the interacting targets, discarding also those mapped within the 100-kb viewpoint window, as previously reported[8]. The same approach was used to quantify the distribution of contacts in the three windows defined as follows: cluster (from the 5′ UTR of the most 5′ Hox genes (zebrafish, *hoxd13a*; amphioxus, *Hox15*) to the 3′ UTR of the most 3′ Hox genes (zebrafish, *hoxd3a*; amphioxus, *Hox1*)); anterior (downstream of the zebrafish *hoxd3a* and amphioxus *Hox1* genes); and posterior (upstream of the zebrafish *hoxd13a* and amphioxus *Hox15* genes).

Three-dimensional computational modeling and virtual Hi-C. *4C data normalization.* To equalize the amount of reads in all experiments, we normalized the reads for the 4C-seq data sets. We then extracted the data relevant for modeling by calculating the *Z* score (see below on *Z*-score threshold optimization) of those reads as in ref. 44.

Structure determination. The overall approach for determination of genome structure was adapted from a previous work[44] with some variations, using the Integrative Modeling Platform (IMP)[45]. The procedure was divided into three stages:

(1) Representation of the genome locus and translation of the data into spatial restraints. We represented the chromosomal fragment as a flexible string of beads where each bead corresponded to a number of consecutive fragments between ten and 45, depending on the total size of the locus (**Supplementary Fig. 7c**). The size of the beads representing those 20 fragments was proportional to the sum of the sizes of these fragments.

To impose connection between the beads, harmonic upper-bound distance restraints were used between consecutive beads. This distance was the sum of the radii of both beads. Excluded volume restraints were imposed over all the beads so that these would not overlap each other. The reach window of a viewpoint was defined as the area between the furthest upstream and downstream fragments with a *Z* score above the upper *Z* score (u*Z*) (**Supplementary Fig. 13**). Harmonic distance restraints were applied between beads corresponding to the viewpoints and the rest of the beads, as long as the *Z* scores for these beads were above the u*Z* or below the lower *Z* score (l*Z*). We used the absolute

© 2016 Nature America, Inc. All rights reserved.

Z score of the reads to give more weight to the most meaningful reads. Beads outside the reach window were restrained with harmonic lower-bound distances, with a weight equal to the absolute *Z* score. With the harmonic lower-bound restraint, we only imposed the criterion that the beads not be closer than their computed distance (**Supplementary Fig. 7**).

(2) Optimization and sampling of the space of solutions. We combined a Monte Carlo exploration with a local optimization of conjugate gradients and simulated annealing. We started with an individual optimization of five steps of conjugate gradients from an entirely random configuration of beads followed by simulated annealing until the score difference between rounds was below 0.00001 or reached 0 (**Supplementary Fig. 7d**). To sample the space of solutions exhaustively, we computed 50,000 independent optimizations for each genome (**Supplementary Fig. 7e**).

(3) Analysis and assessment of the ensemble of models. We gathered the 200 models with the best score. These solutions were then clustered according to their similarity as measured by their root mean square deviation (r.m.s. deviation). We used the Multiexperiment Viewer, MeV[46], with hierarchical clustering and *k*-means clustering. All models were grouped in two clusters that were the mirror image of each other (**Supplementary Fig. 14**). The most representative models (the closest ones to the mean of all solutions within the most populated cluster) are displayed in **Figure 3**. Results were indistinguishable when we used the solutions for the other mirror-image cluster.

Reconstruction of virtual Hi-C data. We used the models from the most populated cluster to generate the heat map plots that were equivalent to Hi-C data. First, we superimposed all the models (**Supplementary Fig. 7f**). To generate virtual Hi-C heat map plots, we measured the distances between all beads in each model and calculated the mean of these distances (**Supplementary Fig. 7g**).

Empirical calculation of the maximum distance, the lZ and the uZ. The calculation of these parameters was carried out as described previously[44] with small variations. The u*Z* score varied between 0.2 and 1.4 in bins of 0.2. The l*Z* score varied in bins of 0.2 between −1.4 and −0.2. The maximum distance varied from 3,000 to 7,000 in bins of 1,000. Because of the heavy computational load, we did not consider narrower bins or higher or lower values.

For each set of parameters, we generated 500 models, calculated the mean distances between the viewpoints and the rest of the fragments, and compared them to the distances that represented each set of 20 fragments of the normalized 4C data (**Supplementary Fig. 15b**,**d**,**f**).

The set of parameters that best fitted the 4C data included 0.2 for u*Z* and −0.2 for l*Z* in amphioxus, zebrafish and mouse. The best maximum distances were different for each species. To allow comparison, we needed to set the same maximum distance for all three. Taking this into account and for the sake of ease of visualization, we settled on the maximum distance of 7,000, whose score was also among the best (**Supplementary Fig. 15a**,**c**,**e**).

Validation of the virtual Hi-C approach. To validate the virtual Hi-C method, we followed two strategies.

(1) Jackknife resampling. We tested the reproducibility and robustness of the virtual Hi-C results by taking advantage of the extensive number of viewpoints available in our amphioxus and zebrafish Hox 4C-seq data. We performed additional modeling experiments by resampling our original data sets using different subsets of 4C data both in zebrafish and amphioxus (**Supplementary Table 2**). We generated 500 models with the same parameters that we used for our initial modeling and reconstructed virtual Hi-C data for each subset. Subsequently, we calculated Spearman's coefficients between the different subsets. This demonstrated that virtual Hi-C results are very reproducible and robust to perturbations, with high correlations even when 60% of the viewpoints were eliminated (**Supplementary Fig. 10** and **Supplementary Table 2**).

(2) Modeling of other loci and shifted calculation of correlations. To validate the models and the virtual Hi-C results derived from them, we generated models for diverse mouse genomic regions using previously published 4C-seq data (from the HoxD locus and two additional loci: *Wnt6-Ihh-Epha4-Pax3* and *Med13l-Tbx3-Tbx5-Rbm19*; refs. 17,47,48). Using these models, we generated the virtual Hi-C results and compared them with previously published experimental Hi-C data[20] (**Supplementary Figs. 8** and **9**). These comparisons were performed shifting the window used for the modeling by 25% of its size in each direction, in steps of 20 kb (**Supplementary Fig. 8**). For each comparison, Spearman's and Pearson's correlations were calculated. Because of the dominance of read counts corresponding to short distances, we calculated these correlations using bins separated by at least 240 kb (HoxD and *Med13l-Tbx3-Tbx5-Rbm19*) or 480 kb (*Wnt6-Ihh-Epha4-Pax3*), to account for the different size of these three loci (~2.12, ~2.48 and ~4.88 Mb, respectively). In all cases, our 4C-seq–derived virtual Hi-C contact matrices accurately recapitulate the TAD organization and borders present in the experimental Hi-C maps, with Spearman's and Pearson's coefficients within the same range (from 0.63 to 0.88) of those typically obtained between different Hi-C experimental conditions (from 0.4 to 0.99; refs. 20,49–51) (**Supplementary Fig. 9** and **Supplementary Table 3**).

ATAC-seq. ATAC-seq experiments in amphioxus embryos were performed as previously described[23,24]. Approximately 80,000 cells (corresponding to 13 embryos at 36 h.p.f.) were directly lysed in cold lysis buffer (10 mM Tris pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$ and 0.1% Igepal) after removing the seawater by centrifuging briefly. The sample was then incubated for 30 min at 37 °C with TDE1 enzyme and purified with the Qiagen MinElute kit. A PCR reaction was performed with 13 cycles using Ad1F and Ad2.3R primers and KAPA HiFi Hot-Start enzyme (Kapa Biosystems). The resulting library was multiplexed and sequenced in a HiSeq 2000 lane. Reads were aligned using the mentioned *B. lanceolatum* assembly. Duplicated pairs or those separated by more than 2 kb were removed. The enzyme cleavage site was determined as the position −4 (minus strand) or +5 (plus strand) relative to each read start, and this position was extended by 5 bp in both directions for signal visualization. For the zebrafish reporter assays of anterior elements, we selected four regions including ATAC-seq peaks with no overlap with coding exons, transcriptional start sites and repetitive elements. We applied the same criteria to the posterior region, also excluding ATAC-seq peaks tightly associated with amphioxus Evx genes (those located in Evx introns and within 5 kb of Evx transcribed regions). This rendered a single candidate element between *Evxb* and *Lnp* (**Supplementary Fig. 11**).

Transgenesis in zebrafish. Transgenesis assays were performed as previously reported[52]. Putative enhancers were amplified by PCR from amphioxus genomic DNA using the primers listed in **Supplementary Table 4**. The PCR fragments were subcloned into PCR8/GW/TOPO vector and, using Gateway technology (Life Technologies), were shuttled into an enhancer detection vector composed of a *gata2* minimal promoter, an enhanced *GFP* reporter gene and a strong midbrain enhancer (z48) that works as an internal control for transgenesis in zebrafish[23]. Zebrafish transgenic embryos were generated using the Tol2 transposon/transposase method[53], with minor modifications. One-cell embryos were injected with a 2-nl volume containing 25 ng/µl of transposase mRNA, 20 ng/µl of purified constructs and 0.05% phenol red. To ensure the reproducibility of the expression patterns observed in the reporter assays, three or more stable transgenic lines derived from different founders were generated for each construct. All experimental procedures using vertebrates were ethically approved by the Andalusian government.

26. Liu, Y., Schröder, J. & Schmidt, B. Musket: a multistage *k*-mer spectrum–based error corrector for Illumina sequence data. *Bioinformatics* **29**, 308–315 (2013).
27. Magoč , T. & Salzberg, S.L. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* **27**, 2957–2963 (2011).
28. Luo, R. *et al.* SOAPdenovo2: an empirically improved memory-efficient short-read *de novo* assembler. *Gigascience* **1**, 18 (2012).
29. Huang, S. *et al.* HaploMerger: reconstructing allelic relationships for polymorphic diploid genome assemblies. *Genome Res.* **22**, 1581–1588 (2012).
30. Haas, B.J. *et al.* Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biol.* **9**, R7 (2008).
31. Stanke, M. *et al.* AUGUSTUS: *ab initio* prediction of alternative transcripts. *Nucleic Acids Res.* **34**, W435–W439 (2006).
32. Parra, G., Bradnam, K. & Korf, I. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* **23**, 1061–1067 (2007).
33. Slater, G.S. & Birney, E. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* **6**, 31 (2005).
34. Baughman, K.W. *et al.* Genomic organization of *Hox* and *ParaHox* clusters in the echinoderm, *Acanthaster planci*. *Genesis* **52**, 952–958 (2014).
35. Frazer, K.A., Pachter, L., Poliakov, A., Rubin, E.M. & Dubchak, I. VISTA: computational tools for comparative genomics. *Nucleic Acids Res.* **32**, W273–W279 (2004).

© 2016 Nature America, Inc. All rights reserved.

doi:10.1038/ng.3497

36. Fuentes, M. *et al*. Insights into spawning behavior and development of the European amphioxus (*Branchiostoma lanceolatum*). *J. Exp. Zool. B Mol. Dev. Evol.* **308**, 484–493 (2007).
37. Fuentes, M. *et al*. Preliminary observations on the spawning conditions of the European amphioxus (*Branchiostoma lanceolatum*) in captivity. *J. Exp. Zool. B Mol. Dev. Evol.* **302**, 384–391 (2004).
38. Somorjai, I., Bertrand, S., Camasses, A., Haguenauer, A. & Escriva, H. Evidence for stasis and not genetic piracy in developmental expression patterns of *Branchiostoma lanceolatum* and *Branchiostoma floridae*, two amphioxus species that have evolved independently over the course of 200 Myr. *Dev. Genes Evol.* **218**, 703–713 (2008).
39. Dekker, J., Rippe, K., Dekker, M. & Kleckner, N. Capturing chromosome conformation. *Science* **295**, 1306–1311 (2002).
40. Hagège, H. *et al*. Quantitative analysis of chromosome conformation capture assays (3C-qPCR). *Nat. Protoc.* **2**, 1722–1733 (2007).
41. Splinter, E., de Wit, E., van de Werken, H.J., Klous, P. & de Laat, W. Determining long-range chromatin interactions for selected genomic sites using 4C-seq technology: from fixation to computation. *Methods* **58**, 221–230 (2012).
42. Rozen, S. & Skaletsky, H. Primer3 on the WWW for general users and for biologist programmers. *Methods Mol. Biol.* **132**, 365–386 (2000).
43. Smemo, S. *et al*. Obesity-associated variants within *FTO* form long-range functional connections with *IRX3*. *Nature* **507**, 371–375 (2014).
44. Baù, D. *et al*. The three-dimensional folding of the α-globin gene domain reveals formation of chromatin globules. *Nat. Struct. Mol. Biol.* **18**, 107–114 (2011).
45. Russel, D. *et al*. Putting the pieces together: integrative modeling platform software for structure determination of macromolecular assemblies. *PLoS Biol.* **10**, e1001244 (2012).
46. Saeed, A.I. *et al*. TM4: a free, open-source system for microarray data management and analysis. *Biotechniques* **34**, 374–378 (2003).
47. Lupiáñez, D.G. *et al*. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* **161**, 1012–1025 (2015).
48. van Weerd, J.H. *et al*. A large permissive regulatory domain exclusively controls *Tbx3* expression in the cardiac conduction system. *Circ. Res.* **115**, 432–441 (2014).
49. Vietri Rudan, M. *et al*. Comparative Hi-C reveals that CTCF underlies evolution of chromosomal domain architecture. *Cell Rep.* **10**, 1297–1309 (2015).
50. Hou, C., Li, L., Qin, Z.S. & Corces, V.G. Gene density, transcription, and insulators contribute to the partition of the *Drosophila* genome into physical domains. *Mol. Cell* **48**, 471–484 (2012).
51. Zhang, Y. *et al*. Spatial organization of the mouse genome and its role in recurrent chromosomal translocations. *Cell* **148**, 908–921 (2012).
52. Bessa, J. *et al*. Zebrafish enhancer detection (ZED) vector: a new tool to facilitate transgenesis and the functional analysis of *cis*-regulatory regions in zebrafish. *Dev. Dyn.* **238**, 2409–2417 (2009).
53. Kawakami, K. Transgenesis and gene trap methods in zebrafish by using the *Tol2* transposable element. *Methods Cell Biol.* **77**, 201–222 (2004).

© 2016 Nature America, Inc. All rights reserved.

6. Bibliography

Abu-Issa, R., Smyth, G., Smoak, I., Yamamura, K., Meyers, E.N., 2002. Fgf8 is required for pharyngeal arch and cardiovascular development in the mouse. Development 129, 4613-4625.
Acemel, R.D., Tena, J.J., Irastorza-Azcarate, I., Marletaz, F., Gomez-Marin, C., de la Calle-Mustienes, E., Bertrand, S., Diaz, S.G., Aldea, D., Aury, J.M., Mangenot, S., Holland, P.W., Devos, D.P., Maeso, I., Escriva, H., Gomez-Skarmeta, J.L., 2016. A single three-dimensional chromatin compartment in amphioxus indicates a stepwise evolution of vertebrate Hox bimodal regulation. Nature genetics 48, 336-341.
Aguinaldo, A.M., Turbeville, J.M., Linford, L.S., Rivera, M.C., Garey, J.R., Raff, R.A., Lake, J.A., 1997. Evidence for a clade of nematodes, arthropods and other moulting animals. Nature 387, 489-493.
Akiyama, R., Masuda, M., Tsuge, S., Bessho, Y., Matsui, T., 2014. An anterior limit of FGF/Erk signal activity marks the earliest future somite boundary in zebrafish. Development 141, 1104-1109.
Aldea, D., Leon, A., Bertrand, S., Escriva, H., 2015. Expression of Fox genes in the cephalochordate Branchiostoma lanceolatum. Front. Ecol. Evol. 3:80.
Amin, N.M., Shi, H., Liu, J., 2010. The FoxF/FoxC factor LET-381 directly regulates both cell fate specification and cell differentiation in C. elegans mesoderm development. Development 137, 1451-1460.
Anders, S., Huber, W., 2010. Differential expression analysis for sequence count data. Genome biology 11, R106.
Anderson, C.B., Meier, S., 1981. The influence of the metameric pattern in the mesoderm on migration of cranial neural crest cells in the chick embryo. Dev Biol 85, 385-402.
Ando, Z., Sato, S., Ikeda, K., Kawakami, K., 2005. Slc12a2 is a direct target of two closely related homeobox proteins, Six1 and Six4. The FEBS journal 272, 3026-3041.
Andrikou, C., Iovene, E., Rizzo, F., Oliveri, P., Arnone, M.I., 2013. Myogenesis in the sea urchin embryo: the molecular fingerprint of the myoblast precursors. EvoDevo 4, 33.
Aulehla, A., Pourquie, O., 2010. Signaling gradients during paraxial mesoderm development. Cold Spring Harbor perspectives in biology 2, a000869.
Aulehla, A., Wehrle, C., Brand-Saberi, B., Kemler, R., Gossler, A., Kanzler, B., Herrmann, B.G., 2003. Wnt3a plays a major role in the segmentation clock controlling somitogenesis. Developmental cell 4, 395-406.
Aulehla, A., Wiegraebe, W., Baubet, V., Wahl, M.B., Deng, C., Taketo, M., Lewandoski, M., Pourquie, O., 2008. A beta-catenin gradient links the clock and wavefront systems in mouse embryo segmentation. Nature cell biology 10, 186-193.
Balfour, E., 1874. A preliminary account of the development of the elasmobranch fishes. Q. J. Microsc. Sci. 14:323–64.
Balfour, F.M., 1876. The Development of Elasmobranch Fishes. Journal of anatomy and physiology 10, 376 372-411.
Beaster-Jones, L., Kaltenbach, S.L., Koop, D., Yuan, S., Chastain, R., Holland, L.Z., 2008. Expression of somite segmentation genes in amphioxus: a clock without a wavefront? Dev Genes Evol 218, 599-611.
Begbie, J., Brunet, J.F., Rubenstein, J.L., Graham, A., 1999. Induction of the epibranchial placodes. Development 126, 895-902.
Begbie, J., Graham, A., 2001. Integration between the epibranchial placodes and the hindbrain. Science 294, 595-598.
Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. . Journal of the Royal Statistical Society 57(B) :289-300.
Bertrand, S., Aldea, D., Oulion, S., Subirana, L., de Lera, A.R., Somorjai, I., Escriva, H., 2015. Evolution of the Role of RA and FGF Signals in the Control of Somitogenesis in Chordates. PloS one 10, e0136587.

Bertrand, S., Camasses, A., Escriva, H., 2007. [Amphioxus: how to become a vertebrate]. Journal de la Societe de biologie 201, 51-57.
Bertrand, S., Camasses, A., Somorjai, I., Belgacem, M.R., Chabrol, O., Escande, M.L., Pontarotti, P., Escriva, H., 2011. Amphioxus FGF signaling predicts the acquisition of vertebrate morphological traits. Proceedings of the National Academy of Sciences of the United States of America 108, 9160-9165.
Bertrand, S., Escriva, H., 2011. Evolutionary crossroads in developmental biology: amphioxus. Development 138, 4819-4830.
Birchmeier, C., Brohmann, H., 2000. Genes that control the development of migrating muscle precursor cells. Current opinion in cell biology 12, 725-730.
Boorman, C.J., Shimeld, S.M., 2002. Pitx homeobox genes in Ciona and amphioxus show left-right asymmetry is a conserved chordate character and define the ascidian adenohypophysis. Evol Dev 4, 354-365.
Bothe, I., Ahmed, M.U., Winterbottom, F.L., von Scheven, G., Dietrich, S., 2007. Extrinsic versus intrinsic cues in avian paraxial mesoderm patterning and differentiation. Developmental dynamics : an official publication of the American Association of Anatomists 236, 2397-2409.
Bothe, I., Dietrich, S., 2006. The molecular setup of the avian head mesoderm and its implication for craniofacial myogenesis. Dev Dyn 235, 2845-2860.
Bothe, I., Tenin, G., Oseni, A., Dietrich, S., 2011. Dynamic control of head mesoderm patterning. Development 138, 2807-2821.
Boulet, A.M., Capecchi, M.R., 2012. Signaling by FGF4 and FGF8 is required for axial elongation of the mouse embryo. Developmental biology 371, 235-245.
Brent, A.E., Tabin, C.J., 2002. Developmental regulation of somite derivatives: muscle, cartilage and tendon. Current opinion in genetics & development 12, 548-557.
Brent, A.E., Tabin, C.J., 2004. FGF acts directly on the somitic tendon progenitors through the Ets transcription factors Pea3 and Erm to regulate scleraxis expression. Development 131, 3885-3896.
Bronchain, O.J., Pollet, N., Ymlahi-Ouazzani, Q., Dhorne-Pollet, S., Helbling, J.C., Lecarpentier, J.E., Percheron, K., Wegnez, M., 2007. The olig family: phylogenetic analysis and early gene expression in Xenopus tropicalis. Development genes and evolution 217, 485-497.
Bronner, M.E., 2015. Evolution: On the crest of becoming vertebrate. Nature 527, 311-312.
Bryson-Richardson, R.J., Currie, P.D., 2008. The genetics of vertebrate myogenesis. Nat Rev Genet 9, 632-646.
Buckingham, M., 2006. Myogenic progenitor cells and skeletal myogenesis in vertebrates. Current opinion in genetics & development 16, 525-532.
Buckingham, M., Relaix, F., 2007. The role of Pax genes in the development of tissues and organs: Pax3 and Pax7 regulate muscle progenitor cell functions. Annual review of cell and developmental biology 23, 645-673.
Buckingham, M., Rigby, P.W., 2014. Gene regulatory networks and transcriptional mechanisms that control myogenesis. Developmental cell 28, 225-238.
Buenrostro, J.D., Wu, B., Chang, H.Y., Greenleaf, W.J., 2015. ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. Current protocols in molecular biology / edited by Frederick M. Ausubel ... [et al.] 109, 21 29 21-29.
Candiani, S., Kreslova, J., Benes, V., Oliveri, D., Castagnola, P., Pestarino, M., Kozmik, Z., 2003. Cloning and developmental expression of amphioxus Dachschund. Gene Expr Patterns 3, 65-69.
Castro, L.F., Rasmussen, S.L., Holland, P.W., Holland, N.D., Holland, L.Z., 2006. A Gbx homeobox gene in amphioxus: insights into ancestry of the ANTP class and evolution of the midbrain/hindbrain boundary. Dev Biol 295, 40-51.
Cerfontaine, P., 1906. Recherches sur le development de l'amphioxus. Arch. Biol. 22.
Chen, R., Amoui, M., Zhang, Z., Mardon, G., 1997. Dachshund and eyes absent proteins form a complex and function synergistically to induce ectopic eye development in Drosophila. Cell 91, 893-903.

Clegg, C.H., Linkhart, T.A., Olwin, B.B., Hauschka, S.D., 1987. Growth factor control of skeletal muscle differentiation: commitment to terminal differentiation occurs in G1 phase and is repressed by fibroblast growth factor. The Journal of cell biology 105, 949-956.
Conklin, E.G., 1932. The embryology of amphioxus. J. Morphol. 45, 69 151.
Cooke, J., Zeeman, E.C., 1976. A clock and wavefront model for control of the number of repeated structures during animal morphogenesis. Journal of theoretical biology 58, 455-476.
Costa, O.G., 1834. Cenni zoologici, ossia descrizione somaria delle specie nuove di animali discoperti in diverse contrade del regno nell'anno 1834. In Annuario Zoologico Vol. 12, pp. 49-50.
Couly, G.F., Coltey, P.M., Le Douarin, N.M., 1992. The developmental fate of the cephalic mesoderm in quail-chick chimeras. Development 114, 1-15.
Couly, G.F., Coltey, P.M., Le Douarin, N.M., 1993. The triple origin of skull in higher vertebrates: a study in quail-chick chimeras. Development 117, 409-429.
Czajkowski, M.T., Rassek, C., Lenhard, D.C., Brohl, D., Birchmeier, C., 2014. Divergent and conserved roles of Dll1 signaling in development of craniofacial and trunk muscle. Developmental biology 395, 307-316.
da Fonseca, R.R., Albrechtsen, A., Themudo, G.E., Ramos-Madrigal, J., Sibbesen, J.A., Maretty, L., Zepeda-Mendoza, M.L., Campos, P.F., Heller, R., Pereira, R.J., 2016. Next-generation biology: Sequencing and data analysis approaches for non-model organisms. Marine genomics.
Dam, T.M., Kim, H.T., Moon, H.Y., Hwang, K.S., Jeong, Y.M., You, K.H., Lee, J.S., Kim, C.H., 2011. Neuron-specific expression of scratch genes during early zebrafish development. Molecules and cells 31, 471-475.
Damas, H., 1944. Recherches sur le développement de Lampetra fluviatilis L.: contribution à l'étude de la céphalogenèse des vertébrés. Arch. Biol. 55:1–248.
Dastjerdi, A., Robson, L., Walker, R., Hadley, J., Zhang, Z., Rodriguez-Niedenfuhr, M., Ataliotis, P., Baldini, A., Scambler, P., Francis-West, P., 2007. Tbx1 regulation of myogenic differentiation in the limb and cranial mesoderm. Developmental dynamics : an official publication of the American Association of Anatomists 236, 353-363.
Daubas, P., Buckingham, M.E., 2013. Direct molecular regulation of the myogenic determination gene Myf5 by Pax3, with modulation by Six1/4 factors, is exemplified by the -111 kb-Myf5 enhancer. Developmental biology 376, 236-244.
David, B., Mooi, R., 2014. How Hox genes can shed light on the place of echinoderms among the deuterostomes. Evodevo 5, 22.
de Rosa, R., Grenier, J.K., Andreeva, T., Cook, C.E., Adoutte, A., Akam, M., Carroll, S.B., Balavoine, G., 1999. Hox genes in brachiopods and priapulids and protostome evolution. Nature 399, 772-776.
Dehal, P., Boore, J.L., 2005. Two rounds of whole genome duplication in the ancestral vertebrate. PLoS biology 3, e314.
Delfini, M.C., De La Celle, M., Gros, J., Serralbo, O., Marics, I., Seux, M., Scaal, M., Marcelle, C., 2009. The timing of emergence of muscle progenitors is controlled by an FGF/ERK/SNAIL1 pathway. Developmental biology 333, 229-237.
Delsuc, F., Brinkmann, H., Chourrout, D., Philippe, H., 2006. Tunicates and not cephalochordates are the closest living relatives of vertebrates. Nature 439, 965-968.
Delsuc, F., Tsagkogeorga, G., Lartillot, N., Philippe, H., 2008. Additional molecular support for the new chordate phylogeny. Genesis 46, 592-604.
Dequeant, M.L., Glynn, E., Gaudenz, K., Wahl, M., Chen, J., Mushegian, A., Pourquie, O., 2006. A complex oscillating network of signaling genes underlies the mouse segmentation clock. Science 314, 1595-1598.
Desdevises, Y., Maillet, V., Fuentes, M., Escriva, H., 2011. A snapshot of the population structure of Branchiostoma lanceolatum in the Racou beach, France, during its spawning season. PloS one 6, e18520.

Diez del Corral, R., Olivera-Martinez, I., Goriely, A., Gale, E., Maden, M., Storey, K., 2003. Opposing FGF and retinoid pathways control ventral neural pattern, neuronal differentiation, and segmentation during body axis extension. Neuron 40, 65-79.
Diogo, R., Kelly, R.G., Christiaen, L., Levine, M., Ziermann, J.M., Molnar, J.L., Noden, D.M., Tzahor, E., 2015. A new heart for a new head in vertebrate cardiopharyngeal evolution. Nature 520, 466-473.
Dobzhansky, T., 1973. Nothing in Biology Makes Sense Except in the Light of Evolution. American Biology Teacher 35 (3): 125–129.
Dong, F., Sun, X., Liu, W., Ai, D., Klysik, E., Lu, M.F., Hadley, J., Antoni, L., Chen, L., Baldini, A., Francis-West, P., Martin, J.F., 2006. Pitx2 promotes development of splanchnic mesoderm-derived branchiomeric muscle. Development 133, 4891-4899.
Dubrulle, J., Pourquie, O., 2004. fgf8 mRNA decay establishes a gradient that couples axial elongation to patterning in the vertebrate embryo. Nature 427, 419-422.
Dunty, W.C., Jr., Biris, K.K., Chalamalasetty, R.B., Taketo, M.M., Lewandoski, M., Yamaguchi, T.P., 2008. Wnt3a/beta-catenin signaling controls posterior body development by coordinating mesoderm formation and segmentation. Development 135, 85-94.
Dupin, E., Creuzet, S., Le Douarin, N.M., 2006. The contribution of the neural crest to the vertebrate body. Advances in experimental medicine and biology 589, 96-119.
Ernst, J., Bar-Joseph, Z., 2006. STEM: a tool for the analysis of short time series gene expression data. BMC bioinformatics 7, 191.
Escriva, H., Holland, N.D., Gronemeyer, H., Laudet, V., Holland, L.Z., 2002. The retinoic acid signaling pathway regulates anterior/posterior patterning in the nerve cord and pharynx of amphioxus, a chordate lacking neural crest. Development 129, 2905-2916.
Evans, D.J., Noden, D.M., 2006. Spatial relations between avian craniofacial neural crest and paraxial mesoderm cells. Developmental dynamics : an official publication of the American Association of Anatomists 235, 1310-1325.
Feng, J., Li, G., Liu, X., Wang, J., Wang, Y.Q., 2014. Functional analysis of the promoter region of amphioxus beta-actin gene: a useful tool for driving gene expression in vivo. Molecular biology reports 41, 6817-6826.
Ferjentsik, Z., Hayashi, S., Dale, J.K., Bessho, Y., Herreman, A., De Strooper, B., del Monte, G., de la Pompa, J.L., Maroto, M., 2009. Notch is a critical component of the mouse somitogenesis oscillator and is essential for the formation of the somites. PLoS genetics 5, e1000662.
Ferronha, T., Rabadan, M.A., Gil-Guinon, E., Le Dreau, G., de Torres, C., Marti, E., 2013. LMO4 is an essential cofactor in the Snail2-mediated epithelial-to-mesenchymal transition of neuroblastoma and neural crest cells. The Journal of neuroscience : the official journal of the Society for Neuroscience 33, 2773-2783.
Fletcher, R.B., Harland, R.M., 2008. The role of FGF signaling in the establishment and maintenance of mesodermal gene expression in Xenopus. Dev Dyn 237, 1243-1254.
Freund, R., Dorfler, D., Popp, W., Wachtler, F., 1996. The metameric pattern of the head mesoderm--does it exist? Anatomy and embryology 193, 73-80.
Friedman, A.D., Triezenberg, S.J., McKnight, S.L., 1988. Expression of a truncated viral trans-activator selectively impedes lytic infection by its cognate virus. Nature 335, 452-454.
Froriep, A., 1892. Entwicklungsgeschichte des Kopfes [part 1]. Anat. Hefte Abt. 2 Erg. Anat. Entwicklungsgesch 1:521–605.
Froriep, A., 1894. Entwicklungsgeschichte des Kopfes [part 2]. Anat. Hefte Abt. 2 Erg. Anat. Entwicklungsgesch 3:396–459.
Fuentes, M., Benito, E., Bertrand, S., Paris, M., Mignardot, A., Godoy, L., Jimenez-Delgado, S., Oliveri, D., Candiani, S., Hirsinger, E., D'Aniello, S., Pascual-Anaya, J., Maeso, I., Pestarino, M., Vernier, P., Nicolas, J.F., Schubert, M., Laudet, V., Geneviere, A.M., Albalat, R., Garcia Fernandez, J., Holland, N.D., Escriva, H., 2007. Insights into spawning behavior and development of the European amphioxus (Branchiostoma lanceolatum). J Exp Zool B Mol Dev Evol 308, 484-493.

Fuentes, M., Schubert, M., Dalfo, D., Candiani, S., Benito, E., Gardenyes, J., Godoy, L., Moret, F., Illas, M., Patten, I., Permanyer, J., Oliveri, D., Boeuf, G., Falcon, J., Pestarino, M., Fernandez, J.G., Albalat, R., Laudet, V., Vernier, P., Escriva, H., 2004. Preliminary observations on the spawning conditions of the European amphioxus (Branchiostoma lanceolatum) in captivity. J Exp Zool B Mol Dev Evol 302, 384-391.
Fujiwara, S., Corbo, J.C., Levine, M., 1998. The snail repressor establishes a muscle/notochord boundary in the Ciona embryo. Development 125, 2511-2520.
Futch, C.R., Dwinell, S.E., 1977. Nearshore marine ecology at Hutchinson Island, Florida: 1971-1974: IV. lancelets and fishes. Fla. Mer. Res. Publ. 24, 1-23.
Gans, C., Northcutt, R.G., 1983. Neural crest and the origin of vertebrates: a new head. Science 220, 268-273.
Garcia-Fernandez, J., Holland, P.W., 1994. Archetypal organization of the amphioxus Hox gene cluster. Nature 370, 563-566.
Gibb, S., Maroto, M., Dale, J.K., 2010. The segmentation clock mechanism moves up a notch. Trends in cell biology 20, 593-600.
Giordani, J., Bajard, L., Demignon, J., Daubas, P., Buckingham, M., Maire, P., 2007. Six proteins regulate the activation of Myf5 expression in embryonic mouse limbs. Proc Natl Acad Sci U S A 104, 11310-11315.
Glenn Northcutt, R., 2005. The new head hypothesis revisited. J Exp Zool B Mol Dev Evol 304, 274-297.
Goethe, J., 1790. Das Schädelgrüt aus sechs Wirbelknochen aufgebaut. In Zur Naturwissenschaft überhaupt, besonders zur Morphologie Vol. 2.
Goodrich, E., 1918. On the development of the segments of the head in Scyllium. . Q. J. Microsc. Sci. 63:1–30.
Goodrich, E., 1930. Studies on the Structure and Development of Vertebrates.
Gosselck, F., Spittler, P., 1979. Age structure, growth and weight of Branchiostoma senegalense (Acrania, Branchiostomidae) off North-West Africa. Int. Revue Ges. Hydrobiol. 64, 541-550.
Gostling, N.J., Shimeld, S.M., 2003. Protochordate Zic genes define primitive somite compartments and highlight molecular changes underlying neural crest evolution. Evol Dev 5, 136-144.
Green, S.A., Norris, R.P., Terasaki, M., Lowe, C.J., 2013. FGF signaling induces mesoderm in the hemichordate Saccoglossus kowalevskii. Development 140, 1024-1033.
Green, S.A., Simoes-Costa, M., Bronner, M.E., 2015. Evolution of vertebrates as viewed from the crest. Nature 520, 474-482.
Grenier, J., Teillet, M.A., Grifone, R., Kelly, R.G., Duprez, D., 2009. Relationship between neural crest cells and cranial mesoderm during head muscle development. PLoS One 4, e4381.
Grifone, R., Demignon, J., Giordani, J., Niro, C., Souil, E., Bertin, F., Laclef, C., Xu, P.X., Maire, P., 2007. Eya1 and Eya2 proteins are required for hypaxial somitic myogenesis in the mouse embryo. Developmental biology 302, 602-616.
Grifone, R., Demignon, J., Houbron, C., Souil, E., Niro, C., Seller, M.J., Hamard, G., Maire, P., 2005. Six1 and Six4 homeoproteins are required for Pax3 and Mrf expression during myogenesis in the mouse embryo. Development 132, 2235-2249.
Grifone, R., Jarry, T., Dandonneau, M., Grenier, J., Duprez, D., Kelly, R.G., 2008. Properties of branchiomeric and somite-derived muscle development in Tbx1 mutant embryos. Developmental dynamics : an official publication of the American Association of Anatomists 237, 3071-3078.
Grifone, R., Kelly, R.G., 2007. Heartening news for head muscle development. Trends in genetics : TIG 23, 365-369.
Groves, J.A., Hammond, C.L., Hughes, S.M., 2005. Fgf8 drives myogenic progression of a novel lateral fast muscle fibre population in zebrafish. Development 132, 4211-4222.

Guo, C., Sun, Y., Zhou, B., Adam, R.M., Li, X., Pu, W.T., Morrow, B.E., Moon, A., Li, X., 2011. A Tbx1-Six1/Eya1-Fgf8 genetic pathway controls mammalian cardiovascular and craniofacial morphogenesis. The Journal of clinical investigation 121, 1585-1595.
Hacker, A., Guthrie, S., 1998. A distinct developmental programme for the cranial paraxial mesoderm in the chick embryo. Development 125, 3461-3472.
Hall, B.K., 2008. The neural crest and neural crest cells: discovery and significance for theories of embryonic organization. Journal of biosciences 33, 781-793.
Harel, I., Maezawa, Y., Avraham, R., Rinon, A., Ma, H.Y., Cross, J.W., Leviatan, N., Hegesh, J., Roy, A., Jacob-Hirsch, J., Rechavi, G., Carvajal, J., Tole, S., Kioussi, C., Quaggin, S., Tzahor, E., 2012. Pharyngeal mesoderm regulatory network controls cardiac and head muscle morphogenesis. Proceedings of the National Academy of Sciences of the United States of America 109, 18839-18844.
Harel, I., Nathan, E., Tirosh-Finkel, L., Zigdon, H., Guimaraes-Camboa, N., Evans, S.M., Tzahor, E., 2009. Distinct origins and genetic programs of head muscle satellite cells. Developmental cell 16, 822-832.
Hatschek, B., 1893. Amphioxus and its Development (translated by J. Tuckey).
Heanue, T.A., Reshef, R., Davis, R.J., Mardon, G., Oliver, G., Tomarev, S., Lassar, A.B., Tabin, C.J., 1999. Synergistic regulation of vertebrate muscle development by Dach2, Eya2, and Six1, homologs of genes required for Drosophila eye formation. Genes & development 13, 3231-3243.
Himeda, C.L., Barro, M.V., Emerson, C.P., Jr., 2013. Pax3 synergizes with Gli2 and Zic1 in transactivating the Myf5 epaxial somite enhancer. Dev Biol 383, 7-14.
Hirakow, R., Kajita, N., 1994. Electron microscopic study of the development of amphioxus, Branchiostoma belcheri tsingtauense: the neurula and larva. Kaibogaku zasshi. Journal of anatomy 69, 1-13.
Holder, N., Hill, J., 1991. Retinoic acid modifies development of the midbrain-hindbrain border and affects cranial ganglion formation in zebrafish embryos. Development 113, 1159-1170.
Holland, L.Z., 2015. The origin and evolution of chordate nervous systems. Philosophical transactions of the Royal Society of London. Series B, Biological sciences 370.
Holland, L.Z., Albalat, R., Azumi, K., Benito-Gutierrez, E., Blow, M.J., Bronner-Fraser, M., Brunet, F., Butts, T., Candiani, S., Dishaw, L.J., Ferrier, D.E., Garcia-Fernandez, J., Gibson-Brown, J.J., Gissi, C., Godzik, A., Hallbook, F., Hirose, D., Hosomichi, K., Ikuta, T., Inoko, H., Kasahara, M., Kasamatsu, J., Kawashima, T., Kimura, A., Kobayashi, M., Kozmik, Z., Kubokawa, K., Laudet, V., Litman, G.W., McHardy, A.C., Meulemans, D., Nonaka, M., Olinski, R.P., Pancer, Z., Pennacchio, L.A., Pestarino, M., Rast, J.P., Rigoutsos, I., Robinson-Rechavi, M., Roch, G., Saiga, H., Sasakura, Y., Satake, M., Satou, Y., Schubert, M., Sherwood, N., Shiina, T., Takatori, N., Tello, J., Vopalensky, P., Wada, S., Xu, A., Ye, Y., Yoshida, K., Yoshizaki, F., Yu, J.K., Zhang, Q., Zmasek, C.M., de Jong, P.J., Osoegawa, K., Putnam, N.H., Rokhsar, D.S., Satoh, N., Holland, P.W., 2008a. The amphioxus genome illuminates vertebrate origins and cephalochordate biology. Genome research 18, 1100-1111.
Holland, L.Z., Carvalho, J.E., Escriva, H., Laudet, V., Schubert, M., Shimeld, S.M., Yu, J.K., 2013. Evolution of bilaterian central nervous systems: a single origin? EvoDevo 4, 27.
Holland, L.Z., Holland, N.D., Gilland, E., 2008b. Amphioxus and the evolution of head segmentation. Integr Comp Biol 48, 630-646.
Holland, L.Z., Holland, N.N., Schubert, M., 2000. Developmental expression of AmphiWnt1, an amphioxus gene in the Wnt1/wingless subfamily. Development genes and evolution 210, 522-524.
Holland, L.Z., Holland, P.W.H., Holland, N.D., 1996. Revealing homologies between body parts of distantly related animals by in situ hybridization to developmental genes: amphioxus versus vertebrates. In: Ferraris JD, Palumbi SR (eds) Molecular zoology: advances, strategies, and protocols. Wiley-Liss, New York, pp 267–282. (473–483).
Holland, L.Z., Kene, M., Williams, N.A., Holland, N.D., 1997. Sequence and embryonic expression of the amphioxus engrailed gene (AmphiEn): the metameric pattern of transcription resembles that of its segment-polarity homolog in Drosophila. Development 124, 1723-1732.

Holland, L.Z., Onai, T., 2011. Analyses of gene function in amphioxus embryos by microinjection of mRNAs and morpholino oligonucleotides. Methods in molecular biology 770, 423-438.
Holland, L.Z., Onai, T., 2012. Early development of cephalochordates (amphioxus). Wiley interdisciplinary reviews. Developmental biology 1, 167-183.
Holland, L.Z., Panfilio, K.A., Chastain, R., Schubert, M., Holland, N.D., 2005. Nuclear beta-catenin promotes non-neural ectoderm and posterior cell fates in amphioxus embryos. Developmental dynamics : an official publication of the American Association of Anatomists 233, 1430-1443.
Holland, L.Z., Rached, L.A., Tamme, R., Holland, N.D., Kortschak, D., Inoko, H., Shiina, T., Burgtorf, C., Lardelli, M., 2001. Characterization and developmental expression of the amphioxus homolog of Notch (AmphiNotch): evolutionary conservation of multiple expression domains in amphioxus and vertebrates. Developmental biology 232, 493-507.
Holland, L.Z., Schubert, M., Kozmik, Z., Holland, N.D., 1999. AmphiPax3/7, an amphioxus paired box gene: insights into chordate myogenesis, neurogenesis, and the possible evolutionary precursor of definitive vertebrate neural crest. Evolution & development 1, 153-165.
Holland, N.D., Holland, L.Z., 2010. Laboratory spawning and development of the Bahama lancelet, Asymmetron lucayanum (cephalochordata): fertilization through feeding larvae. The Biological bulletin 219, 132-141.
Holland, N.D., Holland, L.Z., Heimberg, A., 2015. Hybrids between the Florida amphioxus (Branchiostoma floridae) and the Bahamas lancelet (Asymmetron lucayanum): developmental morphology and chromosome counts. The Biological bulletin 228, 13-24.
Holland, P.W., 2000. Embryonic development of heads, skeletons and amphioxus: Edwin S. Goodrich revisited. Int J Dev Biol 44, 29-34.
Holland, P.W., Garcia-Fernandez, J., 1996. Hox genes and chordate evolution. Developmental biology 173, 382-395.
Holland, P.W., Koschorz, B., Holland, L.Z., Herrmann, B.G., 1995. Conservation of Brachyury (T) genes in amphioxus and vertebrates: developmental and evolutionary implications. Development 121, 4283-4291.
Horigome, N., Myojin, M., Ueki, T., Hirano, S., Aizawa, S., Kuratani, S., 1999. Development of cephalic neural crest cells in embryos of Lampetra japonica, with special reference to the evolution of the jaw. Dev Biol 207, 287-308.
Horstadius, S., 1950. The neural crest. Oxford University Press, London.
Houtmeyers, R., Souopgui, J., Tejpar, S., Arkell, R., 2013. The ZIC gene family encodes multi-functional proteins essential for patterning and morphogenesis. Cellular and molecular life sciences : CMLS 70, 3791-3811.
Huang, R., Lang, E.R., Otto, W.R., Christ, B., Patel, K., 2001. Molecular and cellular analysis of embryonic avian tongue development. Anatomy and embryology 204, 179-187.
Huang, S., Chen, Z., Yan, X., Yu, T., Huang, G., Yan, Q., Pontarotti, P.A., Zhao, H., Li, J., Yang, P., Wang, R., Li, R., Tao, X., Deng, T., Wang, Y., Li, G., Zhang, Q., Zhou, S., You, L., Yuan, S., Fu, Y., Wu, F., Dong, M., Chen, S., Xu, A., 2014. Decelerated genome evolution in modern vertebrates revealed by analysis of multiple lancelet genomes. Nature communications 5, 5896.
Hubaud, A., Pourquie, O., 2014. Signalling dynamics in vertebrate segmentation. Nature reviews. Molecular cell biology 15, 709-721.
Imai, K.S., Hino, K., Yagi, K., Satoh, N., Satou, Y., 2004. Gene expression profiles of transcription factors and signaling molecules in the ascidian embryo: towards a comprehensive understanding of gene networks. Development 131, 4047-4058.
Inoue, T., Ota, M., Mikoshiba, K., Aruga, J., 2007. Zic2 and Zic3 synergistically control neurulation and segmentation of paraxial mesoderm in mouse embryo. Developmental biology 306, 669-684.
Itoh, N., Ornitz, D.M., 2011. Fibroblast growth factors: from molecular evolution to roles in development, metabolism and disease. J Biochem 149, 121-130.

Jackman, W.R., Langeland, J.A., Kimmel, C.B., 2000. islet reveals segmentation in the Amphioxus hindbrain homolog. Dev Biol 220, 16-26.
Jacobson, A.G., 1988. Somitomeres: mesodermal segments of vertebrate embryos. Development 104 Suppl, 209-220.
Jacobson, A.G., Meier, S., 1984. Morphogenesis of the head of a newt: mesodermal segments, neuromeres, and distribution of neural crest. Dev Biol 106, 181-193.
Jaillon, O., Aury, J.M., Brunet, F., Petit, J.L., Stange-Thomann, N., Mauceli, E., Bouneau, L., Fischer, C., Ozouf-Costaz, C., Bernot, A., Nicaud, S., Jaffe, D., Fisher, S., Lutfalla, G., Dossat, C., Segurens, B., Dasilva, C., Salanoubat, M., Levy, M., Boudet, N., Castellano, S., Anthouard, V., Jubin, C., Castelli, V., Katinka, M., Vacherie, B., Biemont, C., Skalli, Z., Cattolico, L., Poulain, J., De Berardinis, V., Cruaud, C., Duprat, S., Brottier, P., Coutanceau, J.P., Gouzy, J., Parra, G., Lardier, G., Chapple, C., McKernan, K.J., McEwan, P., Bosak, S., Kellis, M., Volff, J.N., Guigo, R., Zody, M.C., Mesirov, J., Lindblad-Toh, K., Birren, B., Nusbaum, C., Kahn, D., Robinson-Rechavi, M., Laudet, V., Schachter, V., Quetier, F., Saurin, W., Scarpelli, C., Wincker, P., Lander, E.S., Weissenbach, J., Roest Crollius, H., 2004. Genome duplication in the teleost fish Tetraodon nigroviridis reveals the early vertebrate proto-karyotype. Nature 431, 946-957.
Jaynes, J.B., O'Farrell, P.H., 1991. Active repression of transcription by the engrailed homeodomain protein. The EMBO journal 10, 1427-1433.
Jeffery, W.R., Strickler, A.G., Yamamoto, Y., 2004. Migratory neural crest-like cells form body pigmentation in a urochordate embryo. Nature 431, 696-699.
Jollie, M., 1977. Segmentation of the vertebrate head. Am. Zool. 17:323–33 17:323–33.
Jouve, C., Iimura, T., Pourquie, O., 2002. Onset of the segmentation clock in the chick embryo: evidence for oscillations in the somite precursors in the primitive streak. Development 129, 1107-1117.
Kageyama, R., Niwa, Y., Isomura, A., Gonzalez, A., Harima, Y., 2012. Oscillatory gene expression and somitogenesis. Wiley interdisciplinary reviews. Developmental biology 1, 629-641.
Kaltenbach, S.L., Holland, L.Z., Holland, N.D., Koop, D., 2009. Developmental expression of the three iroquois genes of amphioxus (BfIrxA, BfIrxB, and BfIrxC) with special attention to the gastrula organizer and anteroposterior boundaries in the central nervous system. Gene Expr Patterns 9, 329-334.
Kelly, R.G., Brown, N.A., Buckingham, M.E., 2001. The arterial pole of the mouse heart forms from Fgf10-expressing cells in pharyngeal mesoderm. Developmental cell 1, 435-440.
Kelly, R.G., Jerome-Majewska, L.A., Papaioannou, V.E., 2004. The del22q11.2 candidate gene Tbx1 regulates branchiomeric myogenesis. Human molecular genetics 13, 2829-2840.
Kenny, D.A., Jurata, L.W., Saga, Y., Gill, G.N., 1998. Identification and characterization of LMO4, an LMO gene with a novel pattern of expression during embryogenesis. Proceedings of the National Academy of Sciences of the United States of America 95, 11257-11262.
Kessel, M., 1992. Respecification of vertebral identities by retinoic acid. Development 115, 487-501.
Kim, G.J., Nishida, H., 2001. Role of the FGF and MEK signaling pathway in the ascidian embryo. Development, growth & differentiation 43, 521-533.
Kim, G.J., Yamada, A., Nishida, H., 2000. An FGF signal from endoderm and localized factors in the posterior-vegetal egg cytoplasm pattern the mesodermal tissues in the ascidian embryo. Development 127, 2853-2862.
Kingsbury, B., 1920. The developmental origin of the notochord. Science 51:190–93.
Kingsbury, B., 1926. Branchiomerism and the theory of head segmentation. J. Morphol. 42:83–109.
Kingsbury, B., Adelmann, H., 1926. The morphological plan of the head. Q. J. Microsc. Sci. 68:239–86.
Kitamura, K., Miura, H., Miyagawa-Tomita, S., Yanazawa, M., Katoh-Fukui, Y., Suzuki, R., Ohuchi, H., Suehiro, A., Motegi, Y., Nakahara, Y., Kondo, S., Yokoyama, M., 1999. Mouse Pitx2 deficiency leads to anomalies of the ventral body wall, heart, extra- and periocular mesoderm and right pulmonary isomerism. Development 126, 5749-5758.

Kok, F.O., Shepherd, I.T., Sirotkin, H.I., 2010. Churchill and Sip1a repress fibroblast growth factor signaling during zebrafish somitogenesis. Developmental dynamics : an official publication of the American Association of Anatomists 239, 548-558.
Koltzoff, N., 1902. Entwicklungsgeschichte des Kopfes von Petromyzon Planeri. Ein Beitrag zur Lehreüber Metamerie desWirbelthierkopfes. . Bull. Soc. Imp. Nat. Moscou 16:259–589.
Kon, T., Nohara, M., Yamanoue, Y., Fujiwara, Y., Nishida, M., Nishikawa, T., 2007. Phylogenetic position of a whale-fall lancelet (Cephalochordata) inferred from whole mitochondrial genome sequences. BMC evolutionary biology 7, 127.
Koop, D., Chen, J., Theodosiou, M., Carvalho, J.E., Alvarez, S., de Lera, A.R., Holland, L.Z., Schubert, M., 2014. Roles of retinoic acid and Tbx1/10 in pharyngeal segmentation: amphioxus and the ancestral chordate condition. Evodevo 5, 36.
Kowalevsky, A., 1867. Entwickelungsgeschichte des Amphioxus lanceolatus. Mem. Acad. Imp. Sci. St-Petersb 11, 1-17.
Kowalevsky, A., 1876. Weitere studienuber die entwicklungsgechichte des Amphioxus lanceolatus, etc. Arch. F. Mik. Anat., 13.
Kozmik, Z., Holland, N.D., Kreslova, J., Oliveri, D., Schubert, M., Jonasova, K., Holland, L.Z., Pestarino, M., Benes, V., Candiani, S., 2007. Pax-Six-Eya-Dach network during amphioxus development: conservation in vitro but context specificity in vivo. Dev Biol 306, 143-159.
Krol, A.J., Roellig, D., Dequeant, M.L., Tassy, O., Glynn, E., Hattem, G., Mushegian, A., Oates, A.C., Pourquie, O., 2011. Evolutionary plasticity of segmentation clock networks. Development 138, 2783-2792.
Kumar, S., Duester, G., 2014. Retinoic acid controls body axis extension by directly repressing Fgf8 transcription. Development 141, 2972-2977.
Kuratani, S., 2003. Evolutionary developmental biology and vertebrate head segmentation: A perspective from developmental constraint. Theory Biosci. 122:230–251.
Kuratani, S., 2005. Craniofacial development and the evolution of the vertebrates: the old problems on a new background. Zoological science 22, 1-19.
Kuratani, S., 2008. Is the vertebrate head segmented?-evolutionary and developmental considerations. Integr Comp Biol 48, 647-657.
Kuratani, S., Horigome, N., Hirano, S., 1999. Developmental morphology of the head mesoderm and reevaluation of segmental theories of the vertebrate head: evidence from embryos of an agnathan vertebrate, Lampetra japonica. Developmental biology 210, 381-400.
Kuratani, S., Ueki, T., Hirano, S., Aizawa, S., 1998. Rostral truncation of a cyclostome, Lampetra japonica, induced by all-trans retinoic acid defines the head/trunk interface of the vertebrate body. Developmental dynamics : an official publication of the American Association of Anatomists 211, 35-51.
Kuratani, S.C., Eichele, G., 1993. Rhombomere transplantation repatterns the segmental organization of cranial nerves and reveals cell-autonomous expression of a homeodomain protein. Development 117, 105-117.
Kusakabe, R., Kuraku, S., Kuratani, S., 2011. Expression and interaction of muscle-related genes in the lamprey imply the evolutionary scenario for vertebrate skeletal muscle, in association with the acquisition of the neck and fins. Dev Biol 350, 217-227.
Lacalli, T., Kelly, S., 2000. The infundibular balance organ in amphioxus larvae and related aspects of cerebral vesicle organization. Acta Zool. Stockholm 81, 37–47.
Lacalli, T.C., 2008. Basic features of the ancestral chordate brain: a protochordate perspective. Brain research bulletin 75, 319-323.
Lane, M.E., Runko, A.P., Roy, N.M., Sagerstrom, C.G., 2002. Dynamic expression and regulation by Fgf8 and Pou2 of the zebrafish LIM-only gene, lmo4. Mechanisms of development 119 Suppl 1, S185-189.

Langeland, J.A., Holland, L.Z., Chastain, R.A., Holland, N.D., 2006. An amphioxus LIM-homeobox gene, AmphiLim1/5, expressed early in the invaginating organizer region and later in differentiating cells of the kidney and central nervous system. Int J Biol Sci 2, 110-116.
Langeland, J.A., Tomsa, J.M., Jackman, W.R., Jr., Kimmel, C.B., 1998. An amphioxus snail gene: expression in paraxial mesoderm and neural plate suggests a conserved role in patterning the chordate embryo. Dev Genes Evol 208, 569-577.
Laurent, A., Bihan, R., Omilli, F., Deschamps, S., Pellerin, I., 2008. PBX proteins: much more than Hox cofactors. Int J Dev Biol 52, 9-20.
Le Douarin, N., 1982. The neural crest. Cambridge: Cambridge University Press.
Le Douarin, N.M., Dupin, E., 2012. The neural crest in vertebrate evolution. Current opinion in genetics & development 22, 381-389.
Le Lievre, C.S., 1978. Participation of neural crest-derived cells in the genesis of the skull in birds. Journal of embryology and experimental morphology 47, 17-37.
Le Lievre, C.S., Le Douarin, N.M., 1975. Mesenchymal derivatives of the neural crest: analysis of chimaeric quail and chick embryos. Journal of embryology and experimental morphology 34, 125-154.
Lewis, E.B., 1978. A gene complex controlling segmentation in Drosophila. Nature 276, 565-570.
Li, G., Feng, J., Lei, Y., Wang, J., Wang, H., Shang, L.K., Liu, D.T., Zhao, H., Zhu, Y., Wang, Y.Q., 2014. Mutagenesis at specific genomic loci of amphioxus Branchiostoma belcheri using TALEN method. Journal of genetics and genomics = Yi chuan xue bao 41, 215-219.
Li, H., Durbin, R., 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 26, 589-595.
Li, X., Zhang, W., Chen, D., Lin, Y., Huang, X., Shi, D., Zhang, H., 2006. Expression of a novel somite-formation-related gene, AmphiSom, during amphioxus development. Dev Genes Evol 216, 52-55.
Lin, C.Y., Chen, W.T., Lee, H.C., Yang, P.H., Yang, H.J., Tsai, H.J., 2009. The transcription factor Six1a plays an essential role in the craniofacial myogenesis of zebrafish. Developmental biology 331, 152-166.
Long, J.A., 1995. The rise of fishes: 500 million years of evolution. Johns Hopkins University Press.
Lu, J.R., Bassel-Duby, R., Hawkins, A., Chang, P., Valdez, R., Wu, H., Gan, L., Shelton, J.M., Richardson, J.A., Olson, E.N., 2002. Control of facial muscle development by MyoR and capsulin. Science 298, 2378-2381.
Maden, M., 2002. Retinoid signalling in the development of the central nervous system. Nature reviews. Neuroscience 3, 843-853.
Maden, M., Graham, A., Zile, M., Gale, E., 2000. Abnormalities of somite development in the absence of retinoic acid. The International journal of developmental biology 44, 151-159.
Mahadevan, N.R., Horton, A.C., Gibson-Brown, J.J., 2004. Developmental expression of the amphioxus Tbx1/ 10 gene illuminates the evolution of vertebrate branchial arches and sclerotome. Development genes and evolution 214, 559-566.
Mansfield, J.H., Haller, E., Holland, N.D., Brent, A.E., 2015. Development of somites and their derivatives in amphioxus, and implications for the evolution of vertebrate somites. EvoDevo 6, 21.
Maroto, M., Reshef, R., Munsterberg, A.E., Koester, S., Goulding, M., Lassar, A.B., 1997. Ectopic Pax-3 activates MyoD and Myf-5 expression in embryonic mesoderm and neural tissue. Cell 89, 139-148.
Maruhashi, M., Van De Putte, T., Huylebroeck, D., Kondoh, H., Higashi, Y., 2005. Involvement of SIP1 in positioning of somite boundaries in the mouse embryo. Developmental dynamics : an official publication of the American Association of Anatomists 234, 332-338.
Mason, I., 2007. Initiation to end point: the multiple roles of fibroblast growth factors in neural development. Nature reviews. Neuroscience 8, 583-596.
Mayor, R., Guerrero, N., Young, R.M., Gomez-Skarmeta, J.L., Cuellar, C., 2000. A novel function for the Xslug gene: control of dorsal mesendoderm development by repressing BMP-4. Mech Dev 97, 47-56.

Mazet, F., 2002. The Fox and the thyroid: the amphioxus perspective. Bioessays 24, 696-699.
Meier, S., 1979. Development of the chick embryo mesoblast. Formation of the embryonic axis and establishment of the metameric pattern. Dev Biol 73, 24-45.
Meier, S., Packard, D.S., Jr., 1984. Morphogenesis of the cranial segments and distribution of neural crest in the embryos of the snapping turtle, Chelydra serpentina. Dev Biol 102, 309-323.
Meier, S., Tam, P.P., 1982. Metameric pattern development in the embryonic axis of the mouse. I. Differentiation of the cranial segments. Differentiation; research in biological diversity 21, 95-108.
Meyer, A., Schartl, M., 1999. Gene and genome duplications in vertebrates: the one-to-four (-to-eight in fish) rule and the evolution of novel gene functions. Current opinion in cell biology 11, 699-704.
Minguillon, C., Garcia-Fernandez, J., 2002. The single amphioxus Mox gene: insights into the functional evolution of Mox genes, somites, and the asymmetry of amphioxus somitogenesis. Dev Biol 246, 455-465.
Minguillon, C., Jimenez-Delgado, S., Panopoulou, G., Garcia-Fernandez, J., 2003. The amphioxus Hairy family: differential fate after duplication. Development 130, 5903-5914.
Mjaatvedt, C.H., Nakaoka, T., Moreno-Rodriguez, R., Norris, R.A., Kern, M.J., Eisenberg, C.A., Turner, D., Markwald, R.R., 2001. The outflow tract of the heart is recruited from a novel heart-forming field. Developmental biology 238, 97-109.
Molkentin, J.D., Olson, E.N., 1996. Defining the regulatory networks for muscle development. Current opinion in genetics & development 6, 445-453.
Moncaut, N., Cross, J.W., Siligan, C., Keith, A., Taylor, K., Rigby, P.W., Carvajal, J.J., 2012. Musculin and TCF21 coordinate the maintenance of myogenic regulatory factor expression levels during mouse craniofacial development. Development 139, 958-967.
Morriss-Kay, G.M., 2001. Derivation of the mammalian skull vault. Journal of anatomy 199, 143-151.
Morriss-Kay, G.M., Murphy, P., Hill, R.E., Davidson, D.R., 1991. Effects of retinoic acid excess on expression of Hox-2.9 and Krox-20 and on morphological segmentation in the hindbrain of mouse embryos. The EMBO journal 10, 2985-2995.
Mou, C.Y., Zhang, S.C., Lin, J.H., Yang, W.L., Wu, W.Y., Wei, J.W., Wu, X.K., Du, J.C., Fu, Z.Y., Ye, L.T., Lu, Y., Xie, X.J., Wang, Y.L., Xu, A.L., 2002. EST analysis of mRNAs expressed in neurula of Chinese amphioxus. Biochemical and biophysical research communications 299, 74-84.
Munchberg, S.R., Steinbeisser, H., 1999. The Xenopus Ets transcription factor XER81 is a target of the FGF signaling pathway. Mech Dev 80, 53-65.
Munsterberg, A.E., Kitajewski, J., Bumcrot, D.A., McMahon, A.P., Lassar, A.B., 1995. Combinatorial signaling by Sonic hedgehog and Wnt family members induces myogenic bHLH gene expression in the somite. Genes & development 9, 2911-2922.
Naiche, L.A., Holder, N., Lewandoski, M., 2011. FGF4 and FGF8 comprise the wavefront activity that controls somitogenesis. Proc Natl Acad Sci U S A 108, 4018-4023.
Nakakura, E.K., Watkins, D.N., Schuebel, K.E., Sriuranpong, V., Borges, M.W., Nelkin, B.D., Ball, D.W., 2001. Mammalian Scratch: a neural-specific Snail family transcriptional repressor. Proceedings of the National Academy of Sciences of the United States of America 98, 4010-4015.
Niwa, Y., Masamizu, Y., Liu, T., Nakayama, R., Deng, C.X., Kageyama, R., 2007. The initiation and propagation of Hes7 oscillation are cooperatively regulated by Fgf and notch signaling in the somite segmentation clock. Developmental cell 13, 298-304.
Niwa, Y., Shimojo, H., Isomura, A., Gonzalez, A., Miyachi, H., Kageyama, R., 2011. Different types of oscillations in Notch and Fgf signaling regulate the spatiotemporal periodicity of somitogenesis. Genes & development 25, 1115-1120.
Noden, D.M., 1983. The embryonic origins of avian cephalic and cervical muscles and associated connective tissues. The American journal of anatomy 168, 257-276.
Noden, D.M., 1984. Craniofacial development: new views on old problems. The Anatomical record 208, 1-13.

Noden, D.M., 1988. Interactions and fates of avian craniofacial mesenchyme. Development 103 Suppl, 121-140.
Northcutt, R.G., Gans, C., 1983. The genesis of neural crest and epidermal placodes: a reinterpretation of vertebrate origins. The Quarterly review of biology 58, 1-28.
Nowotschin, S., Liao, J., Gage, P.J., Epstein, J.A., Campione, M., Morrow, B.E., 2006. Tbx1 affects asymmetric cardiac morphogenesis by regulating Pitx2 in the secondary heart field. Development 133, 1565-1573.
Oginuma, M., Hirata, T., Saga, Y., 2008. Identification of presomitic mesoderm (PSM)-specific Mesp1 enhancer and generation of a PSM-specific Mesp1/Mesp2-null mouse using BAC-based rescue technology. Mechanisms of development 125, 432-440.
Ohno, S., 1970. Evolution by Gene Duplication. Springer Berlin Heidelberg.
Olivera-Martinez, I., Storey, K.G., 2007. Wnt signals provide a timing mechanism for the FGF-retinoid differentiation switch during vertebrate body axis extension. Development 134, 2125-2135.
Olwin, B.B., Rapraeger, A., 1992. Repression of myogenic differentiation by aFGF, bFGF, and K-FGF is dependent on cellular heparan sulfate. The Journal of cell biology 118, 631-639.
Oulion, S., Bertrand, S., Belgacem, M.R., Le Petillon, Y., Escriva, H., 2012a. Sequencing and analysis of the Mediterranean amphioxus (Branchiostoma lanceolatum) transcriptome. PLoS One 7, e36554.
Oulion, S., Bertrand, S., Escriva, H., 2012b. Evolution of the FGF Gene Family. Int J Evol Biol 2012, 298147.
Owen, R., 1854. The Principal Forms of the Skeleton and of the Teeth.
Pallas, P.S., 1774. Limax lanceolatus. p. 19 + Plate I. . In Spicilegia zoologica Vol. I (Quadrupedium, avium, amphibiorum, piscium, insectorum, molluscorum, aliorumque marinorum).
Palmeirim, I., Henrique, D., Ish-Horowicz, D., Pourquie, O., 1997. Avian hairy gene expression identifies a molecular clock linked to vertebrate segmentation and somitogenesis. Cell 91, 639-648.
Pan, H., Gustafsson, M.K., Aruga, J., Tiedken, J.J., Chen, J.C., Emerson, C.P., Jr., 2011. A role for Zic1 and Zic2 in Myf5 regulation and somite myogenesis. Developmental biology 351, 120-127.
Pani, A.M., Mullarkey, E.E., Aronowicz, J., Assimacopoulos, S., Grove, E.A., Lowe, C.J., 2012. Ancient deuterostome origins of vertebrate brain signalling centres. Nature 483, 289-294.
Papalopulu, N., Clarke, J.D., Bradley, L., Wilkinson, D., Krumlauf, R., Holder, N., 1991. Retinoic acid causes abnormal development and segmental patterning of the anterior hindbrain in Xenopus embryos. Development 113, 1145-1158.
Pascual-Anaya, J., Adachi, N., Alvarez, S., Kuratani, S., D'Aniello, S., Garcia-Fernandez, J., 2012. Broken colinearity of the amphioxus Hox cluster. Evodevo 3, 28.
Piekarski, N., Olsson, L., 2007. Muscular derivatives of the cranialmost somites revealed by long-term fate mapping in the Mexican axolotl (Ambystoma mexicanum). Evolution & development 9, 566-578.
Pires-daSilva, A., Sommer, R.J., 2003. The evolution of signalling pathways in animal development. Nature reviews. Genetics 4, 39-49.
Platt, J.B., 1897. The development of the cartilaginous skull and of the branchial and hypoglossal musculature in Necturus. . Morphol Jb. 25:377-464.
Portman, A., 1969. Ernfiihrüng in die vergleichende Morphologie der Wirbeltier.
Poss, S.G., Boschung, H.T., 1996. Lancelets (Cephalochordata: Branchiostomatidae): how many species are valid? Israel J. Zool. 42 Suppl., 13-66.
Putnam, N.H., Butts, T., Ferrier, D.E., Furlong, R.F., Hellsten, U., Kawashima, T., Robinson-Rechavi, M., Shoguchi, E., Terry, A., Yu, J.K., Benito-Gutierrez, E.L., Dubchak, I., Garcia-Fernandez, J., Gibson-Brown, J.J., Grigoriev, I.V., Horton, A.C., de Jong, P.J., Jurka, J., Kapitonov, V.V., Kohara, Y., Kuroki, Y., Lindquist, E., Lucas, S., Osoegawa, K., Pennacchio, L.A., Salamov, A.A., Satou, Y., Sauka-Spengler, T., Schmutz, J., Shin, I.T., Toyoda, A., Bronner-Fraser, M., Fujiyama, A., Holland, L.Z., Holland, P.W., Satoh, N., Rokhsar, D.S., 2008. The amphioxus genome and the evolution of the chordate karyotype. Nature 453, 1064-1071.

Raible, F., Brand, M., 2001. Tight transcriptional control of the ETS domain factors Erm and Pea3 by Fgf signaling during early zebrafish development. Mechanisms of development 107, 105-117.
Rasmussen, S.L., Holland, L.Z., Schubert, M., Beaster-Jones, L., Holland, N.D., 2007. Amphioxus AmphiDelta: evolution of Delta protein structure, segmentation, and neurogenesis. Genesis 45, 113-122.
Reifers, F., Bohli, H., Walsh, E.C., Crossley, P.H., Stainier, D.Y., Brand, M., 1998. Fgf8 is mutated in zebrafish acerebellar (ace) mutants and is required for maintenance of midbrain-hindbrain boundary development and somitogenesis. Development 125, 2381-2395.
Relaix, F., Demignon, J., Laclef, C., Pujol, J., Santolini, M., Niro, C., Lagha, M., Rocancourt, D., Buckingham, M., Maire, P., 2013. Six homeoproteins directly activate Myod expression in the gene regulatory networks that control early myogenesis. PLoS genetics 9, e1003425.
Rhinn, M., Dolle, P., 2012. Retinoic acid signalling during development. Development 139, 843-858.
Robinson-Rechavi, M., Boussau, B., Laudet, V., 2004. Phylogenetic dating and characterization of gene duplications in vertebrates: the cartilaginous fish reference. Molecular biology and evolution 21, 580-586.
Romer, A., 1972. The vertebrate as a dual animal—somatic and visceral. Evol. Biol. 6:121–56.
Roussigne, M., Blader, P., 2006. Divergence in regulation of the PEA3 family of ETS transcription factors. Gene Expr Patterns 6, 777-782.
Sambasivan, R., Gayraud-Morel, B., Dumas, G., Cimper, C., Paisant, S., Kelly, R.G., Tajbakhsh, S., 2009. Distinct regulatory cascades govern extraocular and pharyngeal arch muscle progenitor cell fates. Developmental cell 16, 810-821.
Sambasivan, R., Kuratani, S., Tajbakhsh, S., 2011. An eye on the head: the development and evolution of craniofacial muscles. Development 138, 2401-2415.
Sambasivan, R., Tajbakhsh, S., 2007. Skeletal muscle stem cell birth and properties. Seminars in cell & developmental biology 18, 870-882.
Santolini, M., Sakakibara, I., Gauthier, M., Ribas-Aulinas, F., Takahashi, H., Sawasaki, T., Mouly, V., Concordet, J.P., Defossez, P.A., Hakim, V., Maire, P., 2016. MyoD reprogramming requires Six1 and Six4 homeoproteins: genome-wide cis-regulatory module analysis. Nucleic acids research.
Sato, S., Ikeda, K., Shioi, G., Nakao, K., Yajima, H., Kawakami, K., 2012. Regulation of Six1 expression by evolutionarily conserved enhancers in tetrapods. Dev Biol 368, 95-108.
Sawada, K., Fukushima, Y., Nishida, H., 2005. Macho-1 functions as transcriptional activator for muscle formation in embryos of the ascidian Halocynthia roretzi. Gene expression patterns : GEP 5, 429-437.
Scaal, M., Wiegreffe, C., 2006. Somite compartments in anamniotes. Anatomy and embryology 211 Suppl 1, 9-19.
Schlosser, G., 2006. Induction and specification of cranial placodes. Developmental biology 294, 303-351.
Schlosser, G., 2008. Do vertebrate neural crest and cranial placodes have a common evolutionary origin? BioEssays : news and reviews in molecular, cellular and developmental biology 30, 659-672.
Schlosser, G., 2015. Vertebrate cranial placodes as evolutionary innovations--the ancestor's tale. Curr Top Dev Biol 111, 235-300.
Schlosser, G., Northcutt, R.G., 2000. Development of neurogenic placodes in Xenopus laevis. The Journal of comparative neurology 418, 121-146.
Schubert, F.R., Tremblay, P., Mansouri, A., Faisst, A.M., Kammandel, B., Lumsden, A., Gruss, P., Dietrich, S., 2001a. Early mesodermal phenotypes in splotch suggest a role for Pax3 in the formation of epithelial somites. Developmental dynamics : an official publication of the American Association of Anatomists 222, 506-521.
Schubert, M., Holland, L.Z., Panopoulou, G.D., Lehrach, H., Holland, N.D., 2000. Characterization of amphioxus AmphiWnt8: insights into the evolution of patterning of the embryonic dorsoventral axis. Evolution & development 2, 85-92.

Schubert, M., Holland, L.Z., Stokes, M.D., Holland, N.D., 2001b. Three amphioxus Wnt genes (AmphiWnt3, AmphiWnt5, and AmphiWnt6) associated with the tail bud: the evolution of somitogenesis in chordates. Developmental biology 240, 262-273.
Schubert, M., Holland, N.D., Escriva, H., Holland, L.Z., Laudet, V., 2004. Retinoic acid influences anteroposterior positioning of epidermal sensory neurons and their gene expression in a developing chordate (amphioxus). Proceedings of the National Academy of Sciences of the United States of America 101, 10320-10325.
Schubert, M., Holland, N.D., Laudet, V., Holland, L.Z., 2006. A retinoic acid-Hox hierarchy controls both anterior/posterior patterning and neuronal specification in the developing central nervous system of the cephalochordate amphioxus. Developmental biology 296, 190-202.
Schubert, M., Meulemans, D., Bronner-Fraser, M., Holland, L.Z., Holland, N.D., 2003. Differential mesodermal expression of two amphioxus MyoD family members (AmphiMRF1 and AmphiMRF2). Gene expression patterns : GEP 3, 199-202.
Schubert, M., Yu, J.K., Holland, N.D., Escriva, H., Laudet, V., Holland, L.Z., 2005. Retinoic acid signaling acts via Hox1 to establish the posterior limit of the pharynx in the chordate amphioxus. Development 132, 61-73.
Seale, P., Sabourin, L.A., Girgis-Gabardo, A., Mansouri, A., Gruss, P., Rudnicki, M.A., 2000. Pax7 is required for the specification of myogenic satellite cells. Cell 102, 777-786.
Seed, J., Hauschka, S.D., 1988. Clonal analysis of vertebrate myogenesis. VIII. Fibroblasts growth factor (FGF)-dependent and FGF-independent muscle colony types during chick wing development. Developmental biology 128, 40-49.
Shih, H.P., Gross, M.K., Kioussi, C., 2007. Cranial muscle defects of Pitx2 mutants result from specification defects in the first branchial arch. Proceedings of the National Academy of Sciences of the United States of America 104, 5907-5912.
Smith, J.J., Kuraku, S., Holt, C., Sauka-Spengler, T., Jiang, N., Campbell, M.S., Yandell, M.D., Manousaki, T., Meyer, A., Bloom, O.E., Morgan, J.R., Buxbaum, J.D., Sachidanandam, R., Sims, C., Garruss, A.S., Cook, M., Krumlauf, R., Wiedemann, L.M., Sower, S.A., Decatur, W.A., Hall, J.A., Amemiya, C.T., Saha, N.R., Buckley, K.M., Rast, J.P., Das, S., Hirano, M., McCurley, N., Guo, P., Rohner, N., Tabin, C.J., Piccinelli, P., Elgar, G., Ruffier, M., Aken, B.L., Searle, S.M., Muffato, M., Pignatelli, M., Herrero, J., Jones, M., Brown, C.T., Chung-Davidson, Y.W., Nanlohy, K.G., Libants, S.V., Yeh, C.Y., McCauley, D.W., Langeland, J.A., Pancer, Z., Fritzsch, B., de Jong, P.J., Zhu, B., Fulton, L.L., Theising, B., Flicek, P., Bronner, M.E., Warren, W.C., Clifton, S.W., Wilson, R.K., Li, W., 2013. Sequencing of the sea lamprey (Petromyzon marinus) genome provides insights into vertebrate evolution. Nature genetics 45, 415-421, 421e411-412.
Somorjai, I., Bertrand, S., Camasses, A., Haguenauer, A., Escriva, H., 2008. Evidence for stasis and not genetic piracy in developmental expression patterns of Branchiostoma lanceolatum and Branchiostoma floridae, two amphioxus species that have evolved independently over the course of 200 Myr. Development genes and evolution 218, 703-713.
Srivastava, M., Larroux, C., Lu, D.R., Mohanty, K., Chapman, J., Degnan, B.M., Rokhsar, D.S., 2010. Early evolution of the LIM homeobox gene family. BMC biology 8, 4.
Stern, H.M., Brown, A.M., Hauschka, S.D., 1995. Myogenesis in paraxial mesoderm: preferential induction by dorsal neural tube and by cells expressing Wnt-1. Development 121, 3675-3686.
Streit, A., 2004. Early development of the cranial sensory nervous system: from a common field to individual placodes. Developmental biology 276, 1-15.
Tajbakhsh, S., Borello, U., Vivarelli, E., Kelly, R., Papkoff, J., Duprez, D., Buckingham, M., Cossu, G., 1998. Differential activation of Myf5 and MyoD by different Wnts in explants of mouse paraxial mesoderm and the later activation of myogenesis in the absence of Myf5. Development 125, 4155-4162.
Tajbakhsh, S., Rocancourt, D., Cossu, G., Buckingham, M., 1997. Redefining the genetic hierarchies controlling skeletal myogenesis: Pax-3 and Myf-5 act upstream of MyoD. Cell 89, 127-138.

Taylor, J.S., Braasch, I., Frickey, T., Meyer, A., Van de Peer, Y., 2003. Genome duplication, a trait shared by 22000 species of ray-finned fish. Genome research 13, 382-390.
Theis, S., Patel, K., Valasek, P., Otto, A., Pu, Q., Harel, I., Tzahor, E., Tajbakhsh, S., Christ, B., Huang, R., 2010. The occipital lateral plate mesoderm is a novel source for vertebrate neck musculature. Development 137, 2961-2971.
Trainor, P.A., Melton, K.R., Manzanares, M., 2003. Origins and plasticity of neural crest cells and their roles in jaw and craniofacial evolution. Int J Dev Biol 47, 541-553.
Tremblay, P., Dietrich, S., Mericskay, M., Schubert, F.R., Li, Z., Paulin, D., 1998. A crucial role for Pax3 in the development of the hypaxial musculature and the long-range migration of muscle precursors. Developmental biology 203, 49-61.
Tzahor, E., 2009. Heart and craniofacial muscle development: a new developmental theme of distinct myogenic fields. Developmental biology 327, 273-279.
Tzahor, E., 2015. Head muscle development. Results and problems in cell differentiation 56, 123-142.
Tzahor, E., Evans, S.M., 2011. Pharyngeal mesoderm development during embryogenesis: implications for both heart and head myogenesis. Cardiovascular research 91, 196-202.
Tzahor, E., Kempf, H., Mootoosamy, R.C., Poon, A.C., Abzhanov, A., Tabin, C.J., Dietrich, S., Lassar, A.B., 2003. Antagonists of Wnt and BMP signaling promote the formation of vertebrate head muscle. Genes & development 17, 3087-3099.
Vermot, J., Pourquie, O., 2005. Retinoic acid coordinates somitogenesis and left-right patterning in vertebrate embryos. Nature 435, 215-220.
von Scheven, G., Alvares, L.E., Mootoosamy, R.C., Dietrich, S., 2006. Neural tube derived signals and Fgf8 act antagonistically to specify eye versus mandibular arch muscles. Development 133, 2731-2745.
Wada, H., Garcia-Fernandez, J., Holland, P.W., 1999. Colinear and segmental expression of amphioxus Hox genes. Developmental biology 213, 131-141.
Wagner, G.P., Lynch, V.J., 2010. Evolutionary novelties. Current biology : CB 20, R48-52.
Waldo, K.L., Kumiski, D.H., Wallis, K.T., Stadt, H.A., Hutson, M.R., Platt, D.H., Kirby, M.L., 2001. Conotruncal myocardium arises from a secondary heart field. Development 128, 3179-3188.
Wang, Y., Zhang, P.J., Yasui, K., Saiga, H., 2002. Expression of Bblhx3, a LIM-homeobox gene, in the development of amphioxus Branchiostoma belcheri tsingtauense. Mechanisms of development 117, 315-319.
Wasylyk, B., Hagman, J., Gutierrez-Hartmann, A., 1998. Ets transcription factors: nuclear effectors of the Ras-MAP-kinase signaling pathway. Trends in biochemical sciences 23, 213-216.
Webb, J.E., Hill, M.B., 1958. The ecology of Lagos lagoon. IV. On the reactions of Branchiostoma nigeriense Webb to its environment. Philos. Trans. Royal Soc. Lond. Ser. B Biol. Sci. 241, 355-391.
Wellik, D.M., 2007. Hox patterning of the vertebrate axial skeleton. Developmental dynamics : an official publication of the American Association of Anatomists 236, 2454-2463.
Wilson, E.B., 1892. On multiple and partial development in amphioxus. Anat. Anz. 7.
Wilson, E.B., 1893. Amphioxus and the mosaic theory of development. J. Morphol. 8, 379 638.
Winchell, C.J., Sullivan, J., Cameron, C.B., Swalla, B.J., Mallatt, J., 2002. Evaluating hypotheses of deuterostome phylogeny and chordate evolution with new LSU and SSU ribosomal DNA data. Molecular biology and evolution 19, 762-776.
Wray, G.A., Hahn, M.W., Abouheif, E., Balhoff, J.P., Pizer, M., Rockman, M.V., Romano, L.A., 2003. The evolution of transcriptional regulation in eukaryotes. Molecular biology and evolution 20, 1377-1419.
Wu, W., Huang, R., Wu, Q., Li, P., Chen, J., Li, B., Liu, H., 2014. The role of Six1 in the genesis of muscle cell and skeletal muscle development. International journal of biological sciences 10, 983-989.
Yabe, T., Takada, S., 2016. Molecular mechanism for cyclic generation of somites: Lessons from mice and zebrafish. Development, growth & differentiation 58, 31-42.

Yarrell, W., 1836. In A History of British Fishes. Vol. II, 1st edn, pp. i-iv + 1-472.
Yu, J.K., Holland, L.Z., Holland, N.D., 2002. An amphioxus nodal gene (AmphiNodal) with early symmetrical expression in the organizer and mesoderm and later asymmetrical expression associated with left-right axis formation. Evolution & development 4, 418-425.
Yu, J.K., Holland, N.D., Holland, L.Z., 2004. Tissue-specific expression of FoxD reporter constructs in amphioxus embryos. Dev Biol 274, 452-461.
Zhang, Q.J., Zhong, J., Fang, S.H., Wang, Y.Q., 2006. Branchiostoma japonicum and B. belcheri are distinct lancelets (Cephalochordata) in Xiamen waters in China. Zoological science 23, 573-579.
Zhao, X., Duester, G., 2009. Effect of retinoic acid signaling on Wnt/beta-catenin and FGF signaling during body axis extension. Gene expression patterns : GEP 9, 430-435.
Zou, D., Silvius, D., Davenport, J., Grifone, R., Maire, P., Xu, P.X., 2006. Patterning of the third pharyngeal pouch into thymus/parathyroid by Six and Eya1. Dev Biol 293, 499-512.